and Masani (1972)], the process has a spectral representation

$$Y(t) = \int_{-\infty}^{\infty} [\exp\{i\lambda t\} - 1]/(i\lambda)Z(d\lambda),$$

with $Z(\cdot)$ a random process satisfying $\operatorname{cov}\{Z(d\lambda), Z(d\mu)\} = \delta(\lambda - \mu)F(d\lambda)\,d\mu$ with $\delta(\cdot)$ the Dirac delta function and with $F(d\lambda)/(1 + \lambda^2)$ a bounded nonnegative measure. The (co)variance function of the process takes the form

$$\operatorname{cov}\{Y(t), Y(u)\} = \int_{-\infty}^{\infty} [\exp\{i\lambda t\} - 1][\exp\{-i\lambda u\} - 1]/\lambda^2 F(d\lambda).$$

I would submit that these results and in particular the representation

$$\operatorname{var} Y(t) = \int_{-\infty}^{\infty} (\sin \lambda t/2)^2/(\lambda/2)^2 F(d\lambda)$$

constitute an analysis of variance. It should be further mentioned that there are accompanying empirical analyses in the case that $F(\cdot)$ is absolutely continuous see, e.g., Bartlett (1963) and Brillinger (1972).

One way to be led to these results, and indeed corresponding results for stationary random generalized processes, is to apply the ordinary process results to a general linear functional, such as $\int a(t - u)\,dY(t)$, of the process of interest. This leads me to propose the following extension of Dr. Speed's definition. An anova is said to exist for some group of variates, if they satisfy the conditions of Dr. Speed's defintion or if some natural class of functionals of them does. This definition would seem to obviate the need for some of the particular considerations in the manova case. I wonder if it does not lead to a general algebraic result on how a manova structure relates to the corresponding anova structure, quite independently of what the anova design was, for example.

I would like to end by thanking the Editor and Dr. Speed for the opportunity to comment on this important paper.

REFERENCES

BARTLETT, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25** 264–296.

BOCHNER, S. (1947). Stochastic processes. *Ann. of Math.* **48** 1014–1061.

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 483–513. Univ. California Press.

DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

ITÔ, K. (1953). Stationary random distributions. *Mem. Coll. Sci. Univ. Kyoto* **28** 209–223.

MASANI, P. (1972). On helixes in Hibert space. I. *Theory Probab. Appl.* **17** 1–19.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

PERSI DIACONIS

Stanford University

The algebraic aspects of the analysis of variance are an intricate, well worked piece of ground. I am grateful to Speed and his coworkers for carefully sifting the work of our elders and presenting a unified piece in modern language. My comments will outline a somewhat different point of view, using group theory, and point to some other problems that must be solved when merging the present account with modern statistical practice.

Data analysis of anova data. The paper is based on a class of patterned covariance matrices given by an association scheme. Section 6 shows how such schemes can be derived from more classical crossing and nesting operations. A natural basis for algebra of the association scheme now yields projection matrices, an orthogonal direct sum decomposition and an analysis of variance.

This view takes the underlying stochastic structure, at least to second order, as basic. This second-order structure is given in terms of equality constraints on the covariances. From this, a naturally associated analysis of means follows "for free."

There is a more primitive base which treats the problem data analytically. Consider a two-way table with one observation per cell. Perhaps the observed array can be well approximated by $X_{ij} \doteq a + b_i + c_j$, where the residuals are small with respect to the main effects. If this happens, we feel we have a simpler description of the basic underlying phenomena. We might make plots (or fit curves) of the main effects versus other variables. We can next examine the residuals, main effects removed, looking for higher-order structure. If an additive decomposition fails we can search for other simple decompositions. Tukey (1977) gives an aggressively illustrated account of data analytic techniques for arrays.

This approach makes sense, without an underlying model, from a data reduction view. Model or not, most modern statisticians now look at their data as a matter of course.

For more complicated designs, the basic simple descriptions need to be thought about more carefully. One idea that captures all the natural examples I know of involves a group.

Spectral analysis. Consider a finite index set T and data X_t indexed by T. For instance, T might be (i, j), $1 \le i \le I$, $1 \le j \le J$, giving a two-way array. Suppose a finite group G can be found that acts transitively on T. In the example, G can be taken as $S_I \times S_J$, where S_n is the symmetric group on n letters.

The observed data X_t can be viewed as a function on T. Let $L(T)$ be the set of all functions on T. Since G acts on T, G acts on $L(T)$. By an elementary argument, $L(T)$ decomposes into subspaces invariant under G and irreducible so no further decomposition is possible,

$$L(T) = V_1 \oplus V_2 \oplus \cdots \oplus V_j. \tag{1}$$

I have been working with the following definition. *Spectral analysis* for this situation consists of the projection of X_t into the various invariant subspaces and the approximation of X_t by as many terms as give a reasonable fit.

In the example of a two-way table,

$$L(T) = V_0 \oplus V_1 \oplus V_2 \oplus V_3,$$

with V_0 the one-dimensional space of constant functions, V_1 the row effects, V_2 the column effects and V_3 the residuals.

The procedure makes no use of probability. If one makes the usual assumptions, one gets an analysis of variance for free from the direct sum decomposition (1).

Fortini (1977) developed a procedure to produce a natural group G for analysis of variance problems. Let me briefly describe his idea. A factor F is a set valued function from T to a range space $\mathscr{F}$. Since everything is finite, a factor can be represented as a $|T|$ by $|\mathscr{F}|$ matrix, with ones in row t indicating the assigned set.

Given a collection of factors $\{F_i\}$, the associated group G consists of all permutations π of T that preserve the factor structure in the sense that for each i, there is π_i of $\mathscr{F}_i$ such that $F_i(\pi t) = \pi_i F(t)$ for all t.

When T indexes a two-way layout, G is $S_I \times S_J$. As Speed emphasizes in his Section 7, there is a close connection between the group and association scheme approaches. With repeated observations, crossing and nesting, Bailey, Praeger, Rowley and Speed (1983) have made an important contribution which gives a neat description of G.

There are association schemes not generated by groups. There are also important applied examples of group decompositions which do not arise from association schemes. It seems worthwhile to give the leading special case, where something new actually happens.

Balanced incomplete blocks. Consider t flavors of ice cream. Subjects taste a subset of $k < t$ of the flavors. Suppose $\binom{t}{k}$ subjects are recruited so the experiment involves $k\binom{t}{k}$ servings. The subjects respond with a rating (say on a continuous scale). Here, the index set T can be thought of as pairs (i, s), where i is the flavor of the individual serving and s is an (unordered) subset of size $|s| = k - 1$ representing the other flavors served to the subject receiving this serving of flavor i.

The symmetric group S_t acts transitively on T. The decomposition of $L(T)$ was one of the problems solved by Young. Now called Young's rule, it is given in modern notation by James (1978). It can be set out as

$$L(T) = S^t \bigoplus_{j=1}^{k-1} 2S^{t-j,j} \oplus S^{t-k,k} \bigoplus_{j=1}^{k-1} S^{t-j-1,j,1},$$

with S^λ representing the irreducible representations of the symmetric group indexed by the partition λ.

This result was first given by Fortini (1977) who derived it by considering the structure with two factors—treatment effect and block effect.

The classical analysis of this problem only involves looking at the projection onto the "grand mean" space S^t, the treatment space $S^{t-1,1}$ and block effect space (adjusted for treatments) $S^{t-2,2}$. Fortini's analysis suggests there is more to think about. The classical analysis depends critically on the assumptions of additive treatment and block effects. The full spectral analysis gives a natural hierarchy of other effects.

The higher-order terms can be interpreted along the following lines: If tasters work partially by comparison, then ratings will not be independent of the other available flavors. Calvin (1954) enlarged the usual model to try to account for this. Fortini gives a systematic extension.

The point for now is that these other terms make good common sense, and that they were delivered by the group theoretic analysis.

It is worth noting that James (1957) treated balanced incomplete blocks via relationship matrices which only give the classical analysis. Later, James (1982) sketched out an example close in spirit to Fortini's treatment—James treated the diallel cross recapturing Yates' (1947) (and Fortini's) earlier analysis.

One benefit of Speed's approach using association schemes: The projection matrices come in a useful form. Decomposing group representations in general involves a sum over the group. This problem frustrated Fortini's treatment. I have solved the computational problem for a wide class of practical problems—representations induced by Young subgroups—by using ideas connected to Radon transforms. There is more work to be done in relating the various approaches.

AN ≠ THE. The present paper focuses on the variance decomposition. For many people this, and the associated F-testing ritual, is all there is. It must be pointed out how much more is possible. Aside from the data analytic aspects emphasized by Tukey, there are a host of other practical problems and available remedies. These include problems of robustness [see, e.g., Hoaglin, Mosteller and Tukey (1985), Chapters 2–5]. There are the benefits of Stein-like shrinking [see, e.g., Stein (1966)] or the closely related possibilities of a Bayesian analysis [see, e.g., Consonni and Dawid (1985)]. There are problems associated to missing data which seem to plague big designed experiments [see, e.g., Dempster, Laird and Rubin (1977)]. Finally, there are the possibilities of linking-in with computer graphics, projection pursuit and other nonlinear analyses.

It will give all of us plenty to do, for years to come, to try to balance the elegant algebraic treatment emphasized by Speed with the practical realities and possibilities of modern practice.

Acknowledgments. I thank Peter Fortini for allowing me to discuss his thesis results. I thank Terry Speed for persistent efforts to link statistics with modern mathematics, and for taking the trouble to explain them in English.

REFERENCES

BAILEY, R. A., PRAEGER, C. E., ROWLEY, C. A. and SPEED, T. P. (1983). Generalised wreath products of permutation groups. *Proc. London Math. Soc* (3) **47** 69–82.

CALVIN, L. (1954). Doubly balanced incomplete block designs for experiments in which the treatment effects are correlated. *Biometrics* **10** 61–88.

CONSONNI, G. and DAWID, A. P. (1985). Invariant normal Bayesian linear models and experimental designs. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) **2** 629–644. North-Holland, Amsterdam.

DEMPSTER, A. P., LAIRD, N. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DIACONIS, P. (1987). *Group Theory in Statistics*. IMS, Hayward, Calif. To appear.
FORTINI, P. (1977). Representations of groups and the analysis of variance. Ph.D. dissertation, Dept. of Statistics, Harvard Univ.
HOAGLIN, D. C., MOSTELLER, F. M. and TUKEY, J. W. (1985). *Exploring Data Tables, Trends, and Shapes*. Wiley, New York.
JAMES, A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Statist.* **28** 993–1002.
JAMES, A. T. (1982). Analysis of variance determined by symmetry and combinatorial properties of zonal polynomials. In *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, eds.) 329–341. North-Holland, Amsterdam.
JAMES, G. D. (1978). *The Representation Theory of the Symmetric Groups. Lecture Notes in Math.* **682**. Springer, New York.
STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics* (F. N. David, ed.) 351–366. Wiley, New York.
TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.
YATES, F. (1947). Analysis of data from all possible reciprocal crosses from a set of parental lines. *Heredity* **1** 287–301.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

FRANKLIN A. GRAYBILL

Colorado State University

"Analysis of Variance (ANOVA)" is undoubtedly one of the most used and most useful techniques in statistics, but it may be one of the least understood procedures that nonstatisticians use. My experience is that when ANOVA is discussed in elementary statistical texts or taught in methods courses, particularly to nonstatistics majors, there is very little attempt to clearly state what *variance* is being analyzed. Students who take these courses often do not even realize they are analyzing a variance in an ANOVA so the words do not imply any special meaning.

It seems to me that there are two aspects to this: (1) a model that contains, means, variances and covariances; and (2) a statistical analysis of this model (this is where ANOVA comes in if appropriate for a statistical analysis of the model under study). It is important to precisely state each. Writers often tend to use the same words to describe the model and the statistical analysis of the model.

My understanding of what Fisher meant when he used ANOVA to analyze means is the following: To test a null hypothesis of equal means there are two models (1) the original model, and (2) the model specified by the null hypothesis. An estimate of the variance is computed for each model and if the estimates are sufficiently different, the null hypothesis of equal means is rejected. If this is what Fisher meant, then he was indeed *analyzing a variance* and by ANOVA he