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 meant a procedure or technique. Fisher (1924) states, "From the definition of the z distribution it will obviously be reproduced by the errors in the ratio of two independent estimates of the variance,.... The practical working of cases involving the z distribution can usually be shown most simply in the form of an analysis of variance." Again Fisher (1928) states, "...the analysis of variance, in which the variance is analyzed into two parts, that within and that between the classes or 'fraternities' of which the data are composed." In that paper he also states, "...they are all amenable to the same technical procedure known as the analysis of variance; and all may be reduced to an equivalent problem of the distribution of the difference of the logarithm of two independent estimates of variance... ." Since Fisher's paper appeared, ANOVA has been "generalized" and includes techniques used to define and analyze very complicated models that have numerous variances, covariances and means. In the traditional linear model, ANOVA generally (but not always) implies a partitioning of a (total) sum of squares into component *nonnegative* quadratic forms. However, sometimes a procedure is called ANOVA when the partition of a (total) sum of squares into component quadratic forms results in some of the quadratic forms being *negative*; this is the case in unbalanced component of variance models when at least one sum of squares is obtained by subtraction. Also a procedure has been called ANOVA when the component parts of a partition of a (total) sum of squares can be quite general functions that are not necessarily quadratic forms; this is the case in generalizations of Tukey's test for additivity.

With so many uses of ANOVA, it is perhaps inevitable that someone asks, as Dr. Speed does, "What is an analysis of variance?" He states that he expressed the definition "solely in terms of the class V of dispersion matrices." He gives a very detailed, clear and comprehensive discussion and generalization of ANOVA, and states that his formulation is the one deserving the title of ANOVA because almost all examples and the natural generalizations and variants all derive from the present approach and no other. In some ways it seems that Dr. Speed is defining a *model* not a statistical analysis. But ANOVA has the word "analysis" in it and perhaps should be associated with an "analysis" since a given model could have several analyses.

In the paper by Graybill and Hultquist (1961) we were principally interested in *statistically analyzing variances* in a certain model, so we gave a definition of ANOVA (with the assumption of normality) for three reasons: (1) to find minimal sufficient statistics under normality; (2) to determine when the minimal sufficient statistics are complete; and (3) to determine minimum variance quadratic unbiased estimators when the normality assumptions are relaxed. We attempted to distinguish between the model and the analysis. We stated, "For the model II case in which there appear $k + 2$ unknown parameters σ_i^2, $i = 0, 1, \ldots, k + 1$, we make the definition: An analysis of variance will be said to exist... ." We were interested in Eisenhart's model II. Another definition might be appropriate for a different model.

Since Fisher used ANOVA for a statistical analysis, it seems to me that any definition must include a statistical analysis, but in the paper by Speed it is not clear whether a statistical analysis or a model is the dominating feature. If we follow Fisher, ANOVA should be mainly concerned with a statistical analysis. The definition of models for which variances are to be analyzed should at least include what is often referred to as Eisenhart's models I, II and III, both "balanced" and "unbalanced," since they fit into the framework of analyzing variances, and are useful in modelling real world situations.

REFERENCES

FISHER, R. A. (1924). On a distribution yielding the error functions of several well known statistics. *Proc. Internat. Math. Congr., Toronto* 805–813.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc. London Ser. A* **121** 654–673.

GRAYBILL, F. A. and HULTQUIST, R. A. (1961). Theorems concerning Eisenhart's model II. *Ann. Math. Statist.* **32** 261–269.

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

E. J. HANNAN AND C. H. HESSE

Australian National University and Harvard University

Dr. Speed is to be congratulated on the work that is surveyed here. He has sought to describe the basic framework for balanced designs and, moreover, has, with collaborators, elucidated the complex structure consequent on nesting and the proper analysis following from that structure.

What one can argue about, however, is the appropriateneess of the title ANOVA for only this class of structures. Of course, once one begins to abstract it is difficult to know where to stop. Fourier methods provide an example. But in the present situation the position is somewhat reversed and cases are left out in the cold which to the people who regard analysis of variance as rather down to earth (and what could be more down to earth than a field experiment?) are near to the center of the idea. When a subset S of a full replication T is considered in Speed (1985), then it is assumed that the adjacency matrices when restricted to S continue to fulfill the requirements (4.1) in the discussion paper and, in particular, define a commutative algebra. It is interesting to observe that this algebra $\mathbf{A}$ of Section 4, is the same as that introduced in Bose and Mesner (1959) who did not, however, concern themselves with Γ but were instead concerned with the construction of partially balanced designs. (The class of objects for which the A_a are adjacency matrices are not the plots but, for Bose and Mesner, the varieties.) If the subset S is a subset constituting a partially balanced incomplete block and T corresponds to the fully replicated experiment, then the adjacency matrices of Section 4 for S will not produce a commutative algebra so that, in spite of Fisher and Yates (1948, page 19) the resulting analysis is not to be called ANOVA. The algebra $\mathbf{A}$ is often the commuting algebra of the representation of a group G by permutation of the points of T. The analysis of variance will be unique if that algebra is commutative, which is closely connected with the existence of a