OSCAR KEMPTHORNE

Iowa State University

Introduction. I wish first to make some very general comments.

The title is a well-posed question. Analysis of variance is, arguably, the most widely used of the very broad spectrum of processes that are used in applied statistics. If we attempt to use language reasonably, as mathematical statisticians, the phrase "analysis of variance" should mean analysis of $E[X - E(X)]^2$, where X is a random variable. Is this what applied statisticians are doing when they obtain "analysis of variance" or "anovas?" I believe not. I have taught courses on linear models and analysis of variance for many years. I found that in a course of 30 lectures, I do not become involved in random variables until about the 20th lecture. If we accept Speed's picture in which variance of a random variable is the "true" or "real" underlying concept, I have obviously been doing the "wrong thing." But, obviously, I would not have been doing what I regard as the "wrong thing" over all those years.

This leads me to a general plaint about the field of statistics. Over the decades, a language has been developed which is strongly misleading when examined in the light of our general nontechnical language. Let me mention a few of our basic words. One, which is strongly related to the present topic, is the word "regression." In mathematical statistics, this relates to a multivariate random variate, say (X, Y), with regression being a conditional expectation, but in many, perhaps most, uses of the term, X is not a random variable but a controlled variable. Other words are "confidence" and "test," as in "tests of significance" or "tests of hypotheses." And, of course, our most basic word is "probability," in which we, all I think, have problems and our field is, perhaps, schizophrenic.

The Fisher quotation. Speed gives a very challenging beginning with his Fisher quotation. This surely needs examination. Fisher says analysis of variance is not a mathematical theorem, but "rather a convenient method of arranging the arithmetic." The remark by Fisher (made as a contribution to the discussion of a paper) seems reasonable at first sight. But I have to ask if it has any "real" content. What is "the arithmetic?" I would say, rather, that it is a convenient method of arranging a "species" of arithmetic. If we accept my modification, we then have to specify what "species" of arithmetic is involved. I shall discuss this, but at present merely point out that it seems to have no relationship to random variables and variance of random variables.

What then is analysis of variance? I hold the view that analysis of variance is, indeed, a species of "arithmetic" that is related to linear models without any concept of random variables. We have a vector, y, say $n \times 1$, that we are trying to explain by a linear model

$$y = X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_k + \text{disturbance},$$

and we can construct orthogonal projection matrices projecting y onto