ANALYSIS OF VARIANCE CONSIDERED AS AN APPLICATION OF SIMPLE ERROR THEORY

By Walter A. Hendricks

The need for an elementary presentation of the methods of analysis of variance has been recognized by many investigators in various fields of research. A recent monograph by Snedecor (1934) is undoubtedly the most comprehensive attempt to satisfy this need which has appeared in the literature relating to the subject. Snedecor's treatment of the subject consists largely of the presentation of a number of standard types of problems to which the methods of analysis of variance are applicable, directions for performing the necessary computations, and a discussion of the conclusions which may be drawn from the data on the basis of the analysis.

In the opinion of the author of this paper, an elementary presentation of some of the theoretical considerations upon which the methods of analysis of variance are based would also be of some value. The methods of analysis of variance, as given by Fisher (1932), are presented as a natural consequence of intraclass correlation theory. However, the essential concepts may be presented in a more comprehensible form by the use of simple error theory.

It seems appropriate to begin such a presentation with a definition of variance. If we have an infinite number of measurements of the same quantity, the variance of a single measurement is defined as the arithmetic mean of the squares of the errors of those measurements. In actual practice, an infinite number of measurements can never be obtained. We have instead a sample of n measurements, $x_1, x_2, \cdots x_n$, from which the variance of a single measurement may be estimated. By referring to any text on the method of least squares, it may be verified that the best estimate, S^2, of the variance of a single measurement which can be obtained from a sample of n measurements is given by the equation:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2 \quad \text{.................... (1)}$$

in which m represents the arithmetic mean of the n measurements. The quantity, $n - 1$, in the terminology of analysis of variance, is designated as the number of degrees of freedom available for estimating S^2.

It is often necessary to estimate S^2 from a number of different samples of measurements. In such cases, the best estimate of S^2 is obtained by calculating the weighted mean of the variances estimated from the individual samples, each variance being weighted by the number of degrees of freedom which were avail-