A GENERALIZED ANALYSIS OF VARIANCE

By Franklin E. Satterthwaite

University of Iowa and Aetna Life Insurance Company

The analysis of variance is a statistical technique whose fields of application are only beginning to be explored. A few simple standard designs appear in the literature and a great deal has been done with them. However, if the applied statistician limits himself to such standard designs, he soon finds that many of his problems are receiving inadequate or inappropriate treatment. The writer has found this particularly true in his own field where most of the raw data are in the nature of frequencies or averages which lack homogeneity of variance. Also the nature of the problem usually indicates the use of weighted averages rather than simple averages and sometimes part of the data are missing.

The purpose of this study is to examine the fundamental principles underlying analysis of variance designs and to show how designs may be constructed and applied to practically any data which can be assumed to be normally distributed.

1. **Test of independence.** In the analysis of variance we calculate two or more statistics of the types,

$$\chi^2 = \Sigma(x_i - m_i)^2,$$
$$\chi^2 = \Sigma\theta_i^2 .$$

The x_i's are considered to be independent variables from a normal population. The m_i's and the θ_i's are homogeneous linear functions of the x_i's. Heretofore the demonstration of the independence of the χ^2's used has only been made for certain special θ_i's and m_i's. To make our analysis general we shall let our θ_i's be general homogeneous linear functions of the x_i's and we shall define our m_i's through certain linear homogeneous restrictions.

Let us define Chi-square as

$$\chi^2 = \Sigma(x_i - m_i)$$

where the x_i's are independent normally distributed variables with mean zero and unit variance. We also define certain linear functions of the x_i's,[1]

$$\theta_j = a_{ji}x_i , \qquad j = 1, 2, \cdots s, \tag{1}$$

which we shall assume to have been orthogonalized.[2] To define the m_i's we make use of the linear restrictions

[1] A repeated lower case subscript will always indicate summation with respect to that subscript. All subscripts range from 1 to n unless otherwise specified. The Kronecker Delta, δ_{ij}, equals one or zero depending on whether i equals or does not equal j.

[2] The θ_j's are orthogonal if $a_{ik}a_{jk} = \delta_{ij}$. Any algebraically independent set may be