ON TRANSFORMATIONS USED IN THE ANALYSIS OF VARIANCE

BY J. H. CURTISS

Cornell University

1. Introduction. Transformations of variates to render their distributions more tractable in various ways have long been used in statistics [12, chapter XVI]. The present extensive use of the analysis of variance, particularly as applied to data derived from designs such as randomized blocks and Latin squares, has placed new emphasis on the usefulness of such transformations. In the more usual significance tests associated with the analysis of variance, it is assumed *a priori* that the plot yields are statistically independent normally distributed variates which all have the same variance, but which have possibly different means. The hypotheses to be tested are then concerned with relations among these means. But in practice, it sometimes seems appropriate to specify for each variate a distribution in which the variance depends functionally upon the mean; moreover, in such cases, the specification is generally not normal. For example, when the data is in the form of a series of counts or percentages, a Poisson exponential or binomial specification may seem in order, and the variance of either of these distributions is functionally related to the mean of the distribution. Before applying the usual normal theory to such data, it is clearly desirable to transform each variate so that normality and a stable variance are achieved as nearly as possible.

Various transformations have been devised to do this, and a number of articles explaining the nature and use of these transformations have recently been published.[1] However, the available literature on the subject appears to be mainly descriptive and non-mathematical. The object of this paper is to provide a general mathematical theory (sections 2 and 3) for certain types of transformations now in use. In the framework of this theory we shall discuss in particular the square root and inverse sine transformations (section 4), and also several logarithmic transformations (section 4 and section 5).

2. General theory. As it arises in the analysis of variance, the problem of stabilizing a variance functionally related to a mean may be stated as follows: Suppose X is a variate whose mean $\mu = E(X)$ is a real variable with a range S of possible values, and whose standard deviation $\sigma = \sigma_X = \sigma(\mu)$ is a function of μ not identically constant. Required, to find a function $T = f(X)$ such that both $f(X)$ and $\sigma_T^2 = E\{[T - E(T)]^2\}$ are functionally independent of μ for μ on S. $\left(\text{By "functionally independent," we mean that } \frac{\partial f}{\partial \mu} \equiv 0, \text{ and } \frac{\partial \sigma_T^2}{\partial \mu} \equiv 0 \text{ for } \mu \text{ on } S.\right)$

[1] See references [1], [2], [3], [4], [5], [6], [13], [16].