A SIGNIFICANCE TEST FOR COMPONENT ANALYSIS

By Paul G. Hoel

1. Introduction

During the last few years several papers and books have been written on various aspects of what has been termed component or factor analysis. This analysis has arisen from the psychological problem of describing the results on a series of tests in terms of a few distinct abilities or components. In much of such work it is claimed that there does not exist more than a certain number of components, the material discarded in order to substantiate such a claim being considered as due to random errors of sampling or errors of measurement. However, mere inspection of results or the calculation of standard errors of residual correlations is hardly sufficient to justify such conclusions, and therefore a significance test of some kind is necessary. Hotelling[1] considered such a test but based it upon an uncertain analogy with the analysis of variance and upon the legitimacy of using standard errors. The purpose of this paper is to derive a test which is more general in scope and in which all assumptions are explicitly stated.

If each test score is thought of as being made up of two parts, a true score and an error element, the assumption that there exists fewer components than the number of tests implies that the scatter diagram of the true scores will lie in a space of correspondingly smaller dimensionality. Consequently, an ideal test for the number of components would be one which would test the rank of the true moment matrix. In the case of normally distributed variables, this line of approach leads one to the sampling distribution of the generalized variance. Unfortunately, this distribution appears in unintegrated form; however, by considering its moments it is possible to find a good approximation to this exact distribution for samples which are not too small.

The paper proceeds by first finding two approximation distributions for the generalized variance, one for samples which are not too small and one for large samples. It then considers the type of population from which it will be assumed the sample was drawn, and finally applies the test to two numerical examples from recent literature along such lines.

2. Approximation Distributions

Suppose that N individuals have been drawn at random from an n variate normal population whose distribution is expressed by

$$P(x_1, x_2, \cdots, x_n) = Ke^{-\sum_1^n A_{ij} x_i x_j} \tag{1}$$

[1] Harold Hotelling, Analysis of a Complex of Statistical Variables into Principal Components, The Journal of Educational Psychology, September and October, 1933, pp. 21–25.