THE WEIGHTED COMPOUNDING OF TWO INDEPENDENT SIGNIFICANCE TESTS

By M. Zelen and L. S. Joel

National Bureau of Standards

1. Introduction and outline of the problem. In a recent paper on the analysis of incomplete block designs [9], the situation arose where one had two statistically independent F statistics for testing the same null hypothesis. A test was proposed for combining the two tests themselves into a single test which weighted one test relative to the other. It is the purpose of this paper to investigate numerically the power function of this proposed test as it will shed some light as to when an intra-block analysis is worthwhile.

Other situations where one has more than one independent test for testing the same null hypothesis are not uncommon. The tests may have arisen from several sets of independent data or from independent tests made on the same data. General discussions of combining independent tests can be found in Mosteller and Bush [4], Birnbaum [1], and E. S. Pearson [6]. For example a common situation in clinical experiments is that one desires to investigate the effects of two treatments (say) t_1 and t_2 on $2n + m$ people. It is known in advance that m of these people will be available for receiving only one treatment. The experiment is run by assigning t_1 to $(m + n)$ subjects and t_2 to the remaining n people. At a later time, r new people are available who receive treatment t_2. Also of the $2n$ original remaining people, the n people who first received t_1 receive t_2 and vice-versa. Thus the data consist of a cross-over design making use of $2n$ people, and also data where a person received only a single treatment. Thus it is possible to have two tests of the same null hypothesis that the treatments have no effect.[1]

The problem of combining information can be formulated as a problem in estimation. Generally for applications, this latter formulation is usually preferred as it will lead to confidence statements which are usually preferred to tests of a null hypothesis. However it seems interesting from a theoretical point of view to explore the consequences of combining the significance tests themselves.

Let there be two independent variance ratio statistics given by

$$F_j = s_{tj}^2/s_{ej}^2, \qquad j = 1, 2,$$

with degrees of freedom ν and f_j $(j = 1, 2)$ respectively used to test the same null hypothesis. The numerator and denominator mean squares will be referred to as the "treatment" and "error" mean squares and are such that $f_j s_{ej}^2/\sigma_j^2$

Received February 20, 1958; revised January 8, 1959.

[1] We are indebted to Dr. S. Geisser, National Institute of Mental Health for pointing out this example.