

TEST OF HOMOGENEITY FOR NORMAL POPULATIONS

BY G. A. BAKER

University of California

1. Introduction. In biological experiments it is often of interest to test whether or not all the subjects can be regarded as coming from the same normal population. If they have not come from the same normal population, usually the most plausible alternative is that the subjects have come from a population which is the combination of two or more normal populations combined in some proportions. The combination of normal populations is a "smooth" alternative to the hypothesis of a single normal population. Such non-homogeneous populations are not the only "smooth" alternatives, of course, but are included among the "smooth" alternatives. If there is reason to believe that the only deviation from a normal population is due to non-homogeneity, then the results of Professor Neyman in his paper [1] are available in studying this problem.

It is desirable not to make any hypotheses about the mean and standard deviation of the sampled population, but to base all computations and tests on the data contained in the sample. Such a viewpoint has been stressed in a previous paper [2] where it was shown that if the sampling is from a normal population, the probability of a deviation from the mean of a first sample of n measured in terms of the standard deviation of the sample is proportional to

$$(1.1) \quad \frac{dv}{\left(1 + \frac{v^2}{n+1}\right)^{n/2}}.$$

The result (1.1) and Neyman's results give rise to a test of homogeneity which is valid for "large" samples. Empirical results show that fairly conclusive evidence of non-homogeneity may be obtained with samples of 100. Samples of 50 or less may be suggestive but rarely decisive.

2. Development of Test. Suppose that a sample of $n + 1$ is drawn from a normal population. It can be regarded as being made up of a first sample of n and a second sample of one. The value of v corresponding to (1.1) can then be computed and its distribution function is (1.1). This partition, of course, can be made in $n + 1$ ways. That is, $n + 1$ values of v are determined from a random sample of $n + 1$ from the original parent. It is true that these values of v are not independent among themselves. The correlation between the values of v , to a first approximation at least, is of the order of $1/n$ and can be neglected if n is "large."

A suitable transformation as discussed in [3], [1] and elsewhere, transforms (1.1) into a rectangular distribution.

If the same computations are made when the sampled population is not