OBSERVATIONS ON ANALYSIS OF VARIANCE THEORY

By Hilda Geiringer[1]

Bryn Mawr College

One of the important problems of theoretical statistics is the following. Let $x_1, x_2, \cdots x_N$ be the results of N observations; by means of these results we want to test the hypothesis that $V_i(x)$ is the distribution of the ith chance variable x_i. In that situation we often decide to choose a test function $F(x_1, x_2, \cdots x_N)$ and to determine the distribution of F under the above assumption. By means of this distribution we compute the probability of $\xi_1 \leqq F \leqq \xi_2$ and compare this result with the observed value of F.

Suppose there are m groups, each of n observations on $m \cdot n$ chance variables $x_{\mu\nu}$. We may test hypotheses regarding the mn distributions of the $x_{\mu\nu}$ in the way just mentioned. In analysis of variance theory we often use as test functions certain quadratic forms s_w^2 and s_a^2 ("variance within" and "among classes") and their quotient (multiplied by $m(n-1)/(m-1)$), usually denoted by z. Its distribution has been investigated by R. A. Fisher [2] under the assumption that the chance variables are mutually independent and subject to the same normal law. "The five per cent and one per cent points of this distribution have been tabulated by R. A. Fisher and are used to test, whether these two estimates of the same magnitude are significantly different. One gets thus a test of significance *to test whether our sample is a random sample from a homogeneous normal population or not.*[2] If the probability of a certain z-value is too small we shall reject the hypothesis that the sample is a random sample from a homogeneous normal population" [5].

The use of Fisher's z-test is also recommended if we may reasonably assume that the theoretical distributions are approximately normal. "Unless some rather startling lack of normality is known or suspected analysis of variance may be used with confidence." This last remark can be understood by considering that, as we will see in detail, some of the basic results of our theory are independent of the normality of the populations. It is however this assumption of normality which makes possible the complete and elegant solution of the problem of distribution obtained by R. A. Fisher.

If it is *not* possible to determine the exact distribution of a test function under sufficiently general assumptions we may:

a) make simple and particular assumptions concerning the populations

b) investigate an asymptotic solution of the problem, i.e. determine the distributions of the test functions for large samples,[3] or

c) study the mathematical expectations and the variances of the test functions

---

[1] Research under a grant in aid of the American Philosophical Society.

[2] My italics.

[3] cf. statement (a) page 355.