A New Test for 2 × 2 Tables

UNDER this heading, G. A. Barnard[1] puts forward a test which, in language adopted from Neyman and Pearson, "is more powerful than Fisher's". This means in practice that the test advocated passes as significant certain classes of experimental result which, by the test I had put forward[2], would have been judged insignificant; and that, as judged by Barnard's method, my test is thought to be too stringent. However one may choose to express it, the cause of the difference in these calculations is worth elucidating, and, in taking the view he does, Barnard is following the very distinguished precedent of Prof. E. B. Wilson, whose similar proposal a few years ago[3] led to some clarification of the issue[4,5,6].

In the treatment of the problem for which I am responsible, all possible fourfold tables are classified according to the marginal distributions they exhibit. Thus in the case considered by Barnard in which, using three experimental and three control animals, the experimental animals all die and the control animals all survive, the two marginal distributions are both specified by the partition (3^2). Subject to this restriction, only four different experimental results are possible, symbolized by

3	0	2	1	1	2	0	3
0	3	1	2	2	1	3	0

;

and I have demonstrated that, *whatever may be the probability of survival*, if it is the same for both lots, the probabilities of these four possible outcomes are in the ratio 1 : 9 : 9 : 1; or, in other words, the probability of obtaining the most successful outcome by chance is always 1 in 20. For any other marginal distributions a similar series can be obtained, but with these even the most favourable outcome has so high a chance probability that in no case could it be judged significant. It is my view that the existence of these less informative possibilities should not affect our judgment of significance based on the series actually observed.

It may, however, be demonstrated that with repeated sampling, using always three experimental and three control animals having the same probability of death, such outcomes will often occur. If it were legitimate to judge the level of significance

from the proportion of significant judgments in the whole series of 'repeated sampling from the same population', these cases would be brought in to inflate the denominator of the fraction. The least possible frequency of these other series occurs when the chance of death is $\frac{1}{2}$ for both groups, and, in the aggregate, they will then occur 44 times to 20 occurrences of the series observed. Thus Barnard's argument leads to the conclusion that the acceptance of the result as significant is at the significance level 1/64 rather than 1/20; or, more properly, that it has some unknown value not greater than 1 in 64.

In my view the notion of defining the level of significance by 'repeated sampling of the same population' is misleading in the theory of small samples just because it allows of the uncritical inclusion in the denominator of material irrelevant to a critical judgment of what has been observed. In 2 of the 64 cases enumerated above, all animals die or all survive. The fact that such an unhelpful outcome as these might occur, or must occur with a certain probability, is surely no reason for enhancing our judgment of significance in cases where it has not occurred; any more than the possibility that a breeding experiment might have yielded too few offspring to allow one to draw any significant conclusion should not enhance our judgment of significance whenever there are enough offspring for the significance of any supposed effect to be worth discussing.

Of course, the notion of repeated sampling from the same population is usually taken to imply that the total size of the sample is fixed. The total size does not, however, always suffice to specify the type of sample which has been obtained, and it is only the sampling distribution of samples of the same type that can supply a rational test of significance.

R. A. FISHER.

Department of Genetics,
University of Cambridge.
Aug. 13.

[1] Barnard, G. A., *Nature*, **156**, 177 (1945).
[2] Fisher, R. A., "Statistical Methods for Research Workers" (Edinburgh: Oliver and Boyd, 1944), Section 21.02.
[3] Wilson, E. B., *Science*, **93**, 557 (1941).
[4] Fisher, R. A., *Science*, **94**, 210 (1941).
[5] Wilson, E. B., *Proc. U.S. Nat. Acad. Sci.*, **28**, 94 (1942).
[6] Wilson, E. B., and Worcester, Jane, *Proc. U.S. Nat. Acad. Sci.*, **28**, 378 (1942).