LETTERS TO THE EDITOR

HARDY-WEINBERG EQUILIBRIUM AND PRIMITIVE POPULATIONS

Dear Sir:

The long paper by Neel *et al.* in the March issue of this Journal (16:81) includes, in a discussion of "Problems in Calculating Gene Frequencies," some remarks about chi square which are based partly on errors and partly on misunderstanding. All the chi squares and, separately, all the significance levels in their Table 17 are incorrect.

The chi squares, said to be "corrected for continuity," are erroneous because the "correction" used reduced the deviation by one half in each of the three classes, but the appropriate procedure (given by W. L. Stevens, "Estimation of blood-group gene frequencies," *Annals of Eugenics* 8:362–375, 1938, who also gives the calculation of the exact probability) reduces the deviation for heterozygotes by unity, not one-half. This can be appreciated by noticing that the heterozygote class consists of two types. If these types could be distinguished, the data would fall into a 2 × 2 table and the calculation of (corrected) chi square would be as usual. When the two cells for heterozygotes are combined, so are the adjustments of half a unit to their departures from expectation. Thus the adjusted deviations sum to zero, as they should. Hence the "corrected" chi squares given by Neel *et al.* are too large. In passing, I note that the last of six chi squares, that for the Gc system, has *not* been "corrected."

Further, in referring to the chi square distribution, the significance levels were sought as for two degrees of freedom, but these chi squares have but one degree of freedom, there being three classes and the expected numbers being required to conform to the total and to the gene frequency observed. So much for the errors.

The misunderstanding concerns the utility of chi square for detecting departure from Hardy-Weinberg equilibrium. Having found significance levels of roughly 0.9 for their six chi squares for data which they are confident do not come from a population in Hardy-Weinberg equilibrium, Neel *et al.* say, "If this is not a statistical accident, then one is forced to attribute a remarkable insensitiveness to this statistic when applied to small populations, an insensitiveness which in some ways impairs its usefulness." After correction of the errors mentioned above, this statement would presumably not be altered, the *total* (uncorrected) chi square, with six degrees of freedom, being only 4.027, giving no sign of inflation by nonrandomness of mating or other factors. This is not in any way characteristic of chi square. It is easy to calculate the exact probability, and the only deficiency of chi square is that it is imperfect in approximating the exact probabilities. Had Neel *et al.* had before them the exact probabilities (0.58, 0.49, 0.82, 0.49, 0.22, and 0.80, respectively), they would have felt the same insensitivity but would have realized that it should be ascribed not to chi square but to sample size.

Apparently, the various deviations from randomness affecting these data are

less important than Neel *et al.* suppose or (more probably, in my opinion) different deviations have tended to balance so that the net departure is not statistically significant.

HORACE W. NORTON
Department of Animal Science
University of Illinois
College of Agriculture
Urbana, Illinois.

Dear Sir:

We stand corrected in our use of chi square. However, while I hesitate to appear to argue the nuances of this statistic with Dr. Norton, there is an aspect of his comments which seems worth pursuing. He is quite correct in that we erred inadvertently in assigning two degrees instead of one degree of freedom to the chi squares under discussion. However, with respect to the correction for continuity, the situation is perhaps not as clear as it would appear from his comments. We followed the procedure described in the fifth edition of Snedecor's *Statistical Methods*, overlooking the presumably more appropriate method of Stevens. These are both *empirical* corrections. Stevens himself, after applying his correction to a numerical example, wrote: "This is an improvement (over the results without *any* correction), although the correction is too large." In this particular situation, the procedure we employed yields a result closer to the exact probability than that of Stevens (although, of course, this may not be generally so). I wonder how far Dr. Norton would go in following the example set by Stevens? How would he advocate that we correct for continuity when testing three phenotypes for adherence to a 9:6:1 ratio based on two pairs of alleles? To approach the general question somewhat differently, if the correction for continuity is, as it seems to be, entirely empirical, how can we defend the logic of a refinement of this correction unless it yields on the average a value closer to the exact probability (which has not been demonstrated for Stevens' correction)? For a good discussion of how diverse the approaches to these empirical corrections may be, the reader might like to consult Cochran (*Ann. Math. Stat.* 23:315 [1952]).

But even after whichever correction one employs, there is still excellent agreement between the observations and the hypothesis of Hardy-Weinberg equilibrium. Accordingly, we reiterate our original point, which we feel Dr. Norton missed. By a variety of criteria, this population falls far short of meeting the conditions of Hardy-Weinberg equilibrium. In many situations, even a small sample can and will reveal gross departures from expectation. Thus, if expectation of an event is 0.5 and it fails to occur on six consecutive trials, one is already beyond the 0.05 probability level. Our statement was that in this population the departures from Hardy-Weinberg conditions were so gross that, if the usual test for agreement with Hardy-Weinberg equilibrium (agreement between observed and expected genotype frequencies) were a sensitive statistic, it should have revealed the fact, *even* with these small numbers. Dr. Norton's statement that the insensitivity should merely be ascribed to sample