A RECONSIDERATION OF SHEPPARD'S CORRECTIONS

By W. T. Lewis[1]

In computing the moments of a frequency distribution it is customary to find first what are known as the raw moments. These are obtained on the assumption that all the material of each class interval is concentrated at the middle point of the interval. It introduces what is called a grouping error because in fact the material does not all lie at the middle point. To compensate for this error W. F. Sheppard[2] derived a set of corrections. The hypothesis underlying his method is that the distribution may be regarded as similar to one to which the Euler-MacLaurin summation formula without its end terms may be applied. He presupposed such a curve, found its true moments, and then the raw moments that would be obtained if its area were concentrated at several equidistant abscissae. The relationship between these raw moments and the true moments of the curve furnished him with the corrections required for that distribution. If now our observed distribution may be supposed to be sufficiently like that one, we may use his corrections also on the observed data. One may note four points of criticism.

(1) The given distribution may not be similar to the one suggested, in the sense that it would be close to such a curve if the intervals of grouping were made very small; or at all events the purpose of finding the moments may be in part to decide whether or not it would become such a curve, and so one would not like to assume that to be true at the outset. A special case of importance in which this last is true occurs when one is finding the moments of a sample in order to determine whether it may have been drawn from a presupposed universe. It is inexact to use raw moments but it is illogical to use corrections that have been proved only for the universe being tested.

(2) Sheppard's argument does not make use of the one certain fact that is given in the hypothesis, viz: that the partial area of the given distribution over each class interval is exactly as stated. In fact, if, following the argument of some authors, the given curve be assumed to be exponential, it obviously cannot have partial areas everywhere exactly equal to the several given frequencies, for in particular its partial area is not zero beyond the given range.

(3) It is common to find distributions which do not have high contact at the ends of the range and for them Sheppard's corrections certainly fail. To obviate this criticism new corrections have been derived by Pairman and Pear-

[1] With the assistance of Burton H. Camp.

[2] The true values are given on page 220 of "Mathematical Part of Elementary Statistics, by Camp, D. C. Heath and Company, 1931.