ON A METHOD OF TESTING THE HYPOTHESIS THAT AN OBSERVED SAMPLE OF n VARIABLES AND OF SIZE N HAS BEEN DRAWN FROM A SPECIFIED POPULATION OF THE SAME NUMBER OF VARIABLES

By John W. Fertig

With the Technical Assistance of Margaret V. Leary*

The problem of determining whether or not a given observation may be regarded as randomly drawn from a certain population completely specified with respect to its parameters is readily solved if the probability integral of that population be known. In particular if the population specified be a normal population, one may calculate the relative deviate $(x - a)/\sigma$, where a and σ are the population mean and standard deviation respectively, and refer to tables of the normal probability integral. The hypothesis that x was drawn from this population may be rejected if P is less than an arbitrarily fixed value, say $\leq .01$. Generalizations of this problem may be made in two directions: 1) May a single observation simultaneously made on n variables be considered as randomly drawn from a specified population of n variables? 2) May a sample of one variable and of size N be regarded in its entirety as randomly drawn from a specified univariate population?

The solution to the first problem for the case of sampling from a normal population of n variables was given by Karl Pearson in 1908[1] as the "Generalized Probable Error." Let

$$\chi^2 = \frac{1}{P}\left\{ \mathop{S}_{i,j=1}^{n} P_{ij}\left[\frac{(x_i - a_i)(x_j - a_j)}{\sigma_i \sigma_j}\right]\right\}$$

where a_i and σ_i are the population mean and standard deviation respectively of the i^{th} variable, and P_{ij} is the usual cofactor of the element in the i^{th} row and j^{th} column of the determinant P of population correlation coefficients. That is,

$$P = |\rho_{ij}|;\ i, j = 1, 2, 3, \cdots, n.$$

The probability of an observation yielding a smaller discrepancy than that represented by the value of χ^2, i.e., lying between 0 and χ^2, may then be calculated from Tables of the Incomplete Normal Moment Functions[2]. The tables are entered in terms of $(\chi^2)^{\frac{1}{2}}$ and $(n - 1)$, and the tabled value multiplied by $(2\,\pi)^{\frac{1}{2}}$ or 2 depending upon whether n be even or odd respectively.

* From the Memorial Foundation for Neuro-Endocrine Research and the Research Service of the Worcester State Hospital, Worcester, Massachusetts.