ON A CRITERION FOR THE REJECTION OF OBSERVATIONS AND THE DISTRIBUTION OF THE RATIO OF DEVIATION TO SAMPLE STANDARD DEVIATION

By William R. Thompson

Criteria for the rejection of outlying observations may be designed to reject a given fraction of all observations, or a proportion varying with the size of the sample. Irwin[1] has discussed several criteria based on sampling from a normal population which had been used previously, as well as one which he proposed. This is based on the principal of fixing the expectation of rejecting an observation from a sample independently of the aggregate number, N, of the sample. The criterion, λ, is $1/\sigma$ times the interval between successive observations in ascending order of magnitude, where σ is the standard deviation of the sampled population. In the same paper he gave, for different values of N, a table of $P_1(\lambda)$ and $P_2(\lambda)$, respectively probabilities of exceeding given values of λ for the first or second such interval from either end. In actual use, however, σ is estimated from the sample standard deviation, and we are left to decide whether observations in question are to be included or not in estimating the standard deviation as also whether or not to modify this by addition or subtraction of an estimate of its probable error. The object of the present communication is to develop a criterion free from defects of this nature, depending only on the assumption of random sampling from a normal universe. For this purpose we develop the distribution of τ defined by

$$\tau \equiv \frac{\delta}{s}, \tag{1}$$

where s is the sample standard deviation and δ is the deviation of an arbitrary observation of the sample from the sample mean. This leads to definite criteria, which are simple in application.

Accordingly, consider a sample $\{x_i\}$, $i = 1, \cdots, N$, to be drawn at random from a normal population of unknown mean and standard deviation, and that the order of enumeration is arbitrary. Then x_N is an arbitrary one of the elements or *observations*. Now, let

$$\bar{x} = \frac{1}{N}\cdot\sum_{i=1}^{N} x_i, \qquad s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}, \qquad \text{and} \tag{2}$$

$$\delta \equiv x_N - \bar{x}. \tag{3}$$