THE PROBABILITY THAT THE MEAN OF A SECOND SAMPLE WILL DIFFER FROM THE MEAN OF A FIRST SAMPLE BY LESS THAN A CERTAIN MULTIPLE OF THE STANDARD DEVIATION OF THE FIRST SAMPLE

By G. A. Baker, Ph.D.

The following statement of the significance of a probable error is often made: "The probable error of the mean is a value above and below the mean such that if the test were repeated under the same conditions there would be, on the average, equal chances that the mean would fall within or without this range." The probable error is attached to the mean of the sample and it is assumed that the standard deviation of the sample is that of the sampled normal population. This was formerly a very usual explanation of the meaning of probable error by research workers, but it is inaccurate and misleading, especially for samples of 20 or less such as are dealt with in agricultural experiments. The inaccuracy of this explanation of the meaning of probable error has been realized for many years by competent statisticians, but no satisfactory treatment has heretofore been devised.[1]

The attempted explanation of the probable error in terms of the expected frequency of the occurrence of different size deviations of the means of future samples from the sample mean does raise a very interesting, important, and legitimate question, namely, what is the probability of a second mean lying within a certain multiple of the standard deviation of a first sample of the mean of a first sample? This question is of fundamental concern to those engaged in experimental work. Its answer will indicate to investigators reasonable deviations from the results of their first experiments, will form a valid basis for the rejection of doubtful observations or groups of such observations, and will form a basis for a test of the significance of the divergence of results in different experiments. It is found that the usual method of treating the probable error gives an overly optimistic idea of the smallness of the deviations that may be expected in future samples.

The distribution function of the variable

$$v = \frac{x - z}{y}$$

where x is the mean of the first sample, z is the mean of the second sample, and y is the standard deviation of the first sample, is obtained in this paper. The sampled population is assumed to be normal.

[1] Camp, Burton H. "Suggested Problems for Mathematical Research," *Journal American Statistical Association*, Supplement Vol. 30, No. 189A, Mar. 1935, p. 259, No. 5.