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.

Let the sampled population be represented by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty \leqq x \leqq \infty. \tag{1}$$

If a sample of n is drawn from (1) the means, as is well known, will be distributed as proportional to

$$e^{-\frac{1}{2}nx^2}, \qquad -\infty \leqq x \leqq \infty, \tag{2}$$

and the standard deviation will be distributed as proportional to

$$y^{n-2}e^{-\frac{1}{2}ny^2}, \qquad 0 \leqq y \leqq \infty. \tag{3}$$

If a second sample of n is drawn from (1) its mean will be distributed as proportional to

$$e^{-\frac{1}{2}nz^2}, \qquad -\infty \leqq z \leqq \infty. \tag{4}$$

Consider the expression

$$\frac{x - z}{y} \tag{5}$$

and call it v. Then v is the difference between the means of the two samples measured in terms of the standard deviation of the first sample. The distribution function of v is sought.

The three variables x, y, and z are independent. Let y, for the moment, have a constant value and write

$$vy = x - z. \tag{6}$$

The probability of a given value of vy in $d(vy)$ for a given value of y is now being sought, that is, vy is regarded as constant. This probability is proportional to

$$\left[y^{n-2}e^{-\frac{1}{2}ny^2}e^{-\frac{1}{4}nv^2y^2} \int_{-\infty}^{\infty} e^{-(z+\frac{1}{2}vy)^2}\, dz \right] d(vy), \tag{7}$$

from the application of the following

Lemma. If x and y are independent variables, $-\infty \leqq x \leqq \infty$, $-\infty \leqq y \leqq \infty$, and the probability of an x in dx is $f(x)dx$ and the probability of a y in dy is $\varphi(y)dy$, then the probability of $v = y - x$ in dv is proportional to[2]

$$\left[\int_{-\infty}^{\infty} f(x)\varphi(v + x)\, dx \right] dv.$$

Thus the probability of a value of v in dv for a given y is proportional to

$$y^{n-1}e^{-\frac{1}{2}n\left[1+\frac{v^2}{2}\right]y^2}\, dv \tag{8}$$

[2] Baker, G. A. "Random Sampling from Non-Homogeneous Populations," *Metron*, Vol. 8, No. 3, Feb. 1930, p. 68 (slightly modified).

since $d(vy) = ydv$ for y constant. Hence the total probability of a particular value of v in dv will be given as proportional to

$$\text{(9)} \qquad \left[\int_0^\infty y^{n-1} e^{-\frac{1}{2}n\left(1+\frac{v^2}{2}\right)y^2}\, dy\right] dv$$

which is proportional to

$$\text{(10)} \qquad \frac{dv}{\left(1 + \frac{v^2}{2}\right)^{\frac{n}{2}}}$$

If the number in the first sample is n_1 and the number in the second sample is n_2, then (10) becomes

$$\text{(11)} \qquad \frac{dv}{\left(1 + \frac{n_2}{n_1 + n_2} v^2\right)^{\frac{n_1}{2}}}.$$

This distribution, (11), permits an answer to be given to the question, what is the probability that the mean of a sample of a given size n_2 will differ from the mean of a first sample of size n_1 by as much as a constant multiple of the standard deviation of the first sample? Thus, this distribution gives a clear and comprehensible indication of the expected conformity of future experiments and gives a valuable test for the significance of the difference between two means. If it is desired to use this distribution as a rejection criterion, n_1 should be taken so as to include as many items as possible and so as to exclude the doubtful ones. The doubtful items should be included in the second sample. If the original sample is broken up into two or more samples it must be done in such a way as not to destroy the randomness of the resulting parts.

Example. Suppose for the purpose of illustration that a sample of four is to be considered. The proper v-distribution is

$$\frac{\sqrt{2}\, dv}{\pi\left(1 + \frac{v^2}{2}\right)^2}.$$

The value of v which is necessary to give a probability of one-half is a root of

$$\tan^{-1} \frac{\rho}{\sqrt{2}} + \frac{1}{2} \frac{\sqrt{2}\,\rho}{\rho^2 + 2} = \frac{\pi}{4}$$

which is .9. That is, an interval of 1.8 times the standard deviation of the sample of four with center at the mean of the sample is necessary for a probability of one-half that the mean of the next sample of four will lie in this interval. This compares with .75 times the standard deviation of the sample if

$$\frac{\sigma}{\sqrt{n-1}}$$

is used as the probable error of the mean and with .65 times the standard deviation of the sample if

$$\frac{\sigma}{\sqrt{n}}$$

is used as the probable error of the mean. The last two methods of calculating a probable error with the interpretation indicated at the beginning of this paper give the investigator an unwarranted feeling of assurance about the agreement of future samples with a first sample.

If two samples of n_1 and n_2 are drawn from the normal population, (1), then these samples can be combined for the purpose of calculating a standard deviation and the difference between the means of the samples can be measured in terms of the standard deviation of the combined sample. The distribution function of the difference of the means divided by the standard deviation of the combined sample is

$$\text{(11')} \qquad \frac{dv}{\left[1 + \frac{n_1 n_2}{(n_1 + n_2)^2} v^2\right]^{\frac{n_1+n_2}{2}}}.$$

This distribution, (11′), is the basis for a valid test for the significance of the difference between two means. If either this test or the test based on distribution (11) shows a significant difference between the means it can not be ignored.

"Student's" t-distribution is proportional to

$$\text{(12)} \qquad \frac{dt}{\left(1 + \frac{t^2}{N-1}\right)^{\frac{N}{2}}}.$$

The above distributions can be easily transformed into t-distributions so that "Student's" tables can be used. For instance, if we put

$$v = \frac{\sqrt{2}\, t}{\sqrt{n-1}}, \qquad N = n,$$

then (10) becomes proportional to (12). Again, put

$$v = \frac{\sqrt{n_1 + n_2}\, t}{\sqrt{n_2}\sqrt{n_1 - 1}}, \qquad N = n_1,$$

and (11) becomes proportional to (12). Finally, put

$$v = \frac{(n_1 + n_2)\, t}{\sqrt{n_1 n_2}\sqrt{n_1 + n_2 - 1}}, \qquad N = n_1 + n_2,$$

and (11′) becomes proportional to (12).

Summary. The distributions found for the difference of the means of two samples in terms of a standard deviation of one sample or combination of both

samples are similar to and easily transformed into "Student's" t-distribution so that his tables can be used. However, these distributions answer a practical, interesting, and important question that "Student's" t-distribution does not. If in an experimental science a series of observations is made it is desirable to know how much a similar series of observations could be expected to differ from the set of observations now available. This deviation, if it is to mean anything, must be expressed in terms of quantities available from the observations already made. This paper gives the probability function of a deviation in the mean of a future sample measured from the mean of a first sample and measured in terms of the standard deviation of a first sample, that is, in terms of quantities known from the first sample. It is a very definite advantage and a great gain in assurance to know the point from which measurements are being made and the unit in which they are expressed instead of making vague, ill-defined assumptions about the zero point and unit length of the measuring scale. It is true that differences that were formerly considered significant may not be so considered now. But these differences would appear insignificant if experiments were sufficiently repeated, so that the net result is fewer inconsistencies to explain away.