MINNESOTA FORESTRY NOTES

No. 37
January 15, 1955

STANDARD ERRORS FROM SMALL SAMPLES, THEIR ADJUSTMENT WITH RESPECT TO PROBABILITY

Ronald Beazley (1)

In the formula $\pm s = \sqrt{\frac{Sx^2}{n-1}}$ (where S = "sum of", x is a deviation from the sample mean and n is the number of items in the sample), $\pm s$ provides an estimate of the standard deviation of a normal population for any sample size, assuming the sample was taken in a random manner, i.e., any item in the population has as much likelihood of being included in the sample as any other item. The interest here lies in those cases where n is less than 30 and where s is used in estimating errors of the mean (where s_m = the standard error of the mean), for any particular probability.

In common usage, the formula $s_m = \frac{s}{\sqrt{n}}$, or $s_m^2 = \frac{s^2}{n}$, means that the true population mean will not be within plus or minus one standard error of the sample mean 32 times out of 100, i.e. the probability (p) is .32 (or conversely it means that the true population mean will be within plus or minus one standard error of the sample mean 68 times out of 100). This, however, is true only if the sample size (n) is 30 or larger. For a sample smaller than 30 the probability that the population mean is not within plus or minus one s_m is greater than .32, (or that the population mean is within $\pm s_m$ less than 68 times out of 100). For example, if n = 2, the population mean will fall outside plus or minus one s_m 50 times out of 100, instead of only 32 times out of 100, as is true for a sample size = 30 or greater (2). The following numerical examples illustrate the method of adjusting for sample size and probability.

Suppose a small sample is taken such that n = 4 and $Sx^2 = 1200$, so that $s = \sqrt{\frac{Sx^2}{n-1}} = \sqrt{\frac{1200}{4-1}} = \sqrt{400} = \pm 20$. Then $s_m^2 = \frac{s^2}{n} = \frac{400}{4} = 100$, and $s_m = \sqrt{100} = \pm 10$. As conventionally used, this states that the population mean falls outside the range covered by the sample mean $\pm s_m$ 32 times out of 100; but actually for this small sample it falls outside 39 times out of 100 (2). This can be adjusted by multiplying the s_m (as computed above) by the "t" value for n = 4 and probability (p) = .32. Thus the adjusted s_m (call it "E"), $E = \pm 10 \times t_{.32} = \pm 10 \times 1.2 = \pm 12$. It is now true to say that the population mean will not fall within ± 12 of the sample mean 32 times out of 100 -- i.e. with a p of .32. Some useful values of the "t" distribution appear in the table at the end of this paper.

If a probability other than .32 is desired it can be introduced in the same manner. For example if n = 4, $s = \pm 20$ and $s_m = \pm 10$, for a probability of .05 (five times out of 100 the population mean will be greater or less than $\pm$ one s_m from the sample mean) the adjusted

(1) *Assistant Professor, University of Minnesota School of Forestry.*
(2) *Ezekiel, M. 1941. Methods of Correlation Analysis. Table A, p. 23 and Fig. A, p. 505. John Wiley and Sons, Inc. N. Y.*

Published by the School of Forestry, University of Minnesota, St. Paul 1, Minnesota, cooperating with the Division of Forestry, Minnesota Conservation Department, and Forest Industries of Minnesota

s_m, $E = \pm s_m \times t_{.05} = \pm 10 \times 3.2 = \pm 32$. (Note that s and s_m are in real terms such as feet or inches.)

Frequently it is necessary to find the number of items (n') to be included in a sample so that the population mean may be expected to fall outside a certain error, plus or minus, on either side of the sample mean only a given number of times out of 100. Perhaps it is desired to find the mean diameter at breast height of a group of trees so that the population mean will fall outside the sample mean, $\pm$ 1.25 inches, only once in 100 such samples if they were repeated over and over, i.e., if p = .01. The formula commonly used to find n', the number of items in the large sample, is $n' = \left(\frac{s}{s_m}\right)^2$, which is a rearrangement of $\pm s_m = \frac{s}{\sqrt{n}}$. But to find n', it is necessary to first find s from a small pre-sample. Suppose the pre-sample is taken and say, n = 15, and $Sx^2 = 350$, then $s = \sqrt{\frac{Sx^2}{n-1}} = \sqrt{\frac{350}{15-1}} = \sqrt{25} = \pm 5$. Now assuming the plus or minus error which can be tolerated about the sample mean is $\pm$ 1.25, and assuming that the "t" adjustment is <u>not</u> used, the probability in this case remains at approximately .32, and $n' = \left(\frac{s}{s_m}\right)^2 = \left(\frac{5}{1.25}\right)^2 = 16$. But 32 chances out of 100 of having the population mean fall outside this error about the sample mean may be too risky, hence it is decided to find n' such that there is only one chance in 100 that the true mean will not fall within $\pm$ 1.25 of the sample mean, i.e. p = .01. Now the "t" adjustment must be used and at n = 15 and p = .01, $n' = \left(\frac{s}{s_m} . t\right)^2 = \left(\frac{5}{1.25} x 3\right)^2 = 12^2 = 144$.

The following table of "t" values shows usable probabilities and sample sizes. If other combinations, or more accurate values of "t" are desired, Snedecor (3), may be used by adding 1 to the d.f. listed therein to net n; or Ezekiel (2) can be used, where n is shown directly.

Table of "t" Values*

Probability (p) of a larger value of "t"	Sample Size (n) 2	3	4	5	6	7	8	10	15	20	30	∞
.01	63.7	9.9	5.8	4.6	4.0	3.7	3.5	3.3	3.0	2.9	2.8	2.6
.05	12.7	4.3	3.2	2.8	2.6	2.5	2.4	2.3	2.2	2.1	2.0	2.0
.10	6.3	2.9	2.4	2.1	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.7
.20	3.1	1.9	1.6	1.5	1.5	1.4	1.4	1.4	1.4	1.3	1.3	1.3
.32	1.8	1.3	1.2	1.1	1.1	1.1	1.1	1.1	1.0	1.0	1.0	1.0
.50	1.0	.8	.8	.7	.7	.7	.7	.7	.7	.7	.7	.7

*To be used as a multiple of s_m in the formula $E = \pm s_m t$, where $s_m = \frac{s}{\sqrt{n}}$ and as a multiple of $\frac{s}{s_m}$ in the formula $n' = \left(\frac{s}{s_m} . t\right)^2$.

(3) *Snedecor, G. W. 1946. Statistical Methods, Table 3.8, p. 65. The Iowa State College Press, Ames, Iowa.*

Published as Sci. Jour. Ser. Paper No. 3308 of the Minn. Agr. Expt. Sta.