

6-1950

## Sample Size Required For Estimating The Standard Deviation as a Percent of Its True Value

Joseph A. Greenwood

*Navy Department*

Marion M. Sandomire

*Navy Department*

Follow this and additional works at: <http://digitalcommons.unl.edu/usnavyresearch>

---

Greenwood, Joseph A. and Sandomire, Marion M., "Sample Size Required For Estimating The Standard Deviation as a Percent of Its True Value" (1950). *U.S. Navy Research*. 34.

<http://digitalcommons.unl.edu/usnavyresearch/34>

This Article is brought to you for free and open access by the U.S. Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. Navy Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## SAMPLE SIZE REQUIRED FOR ESTIMATING THE STANDARD DEVIATION AS A PER CENT OF ITS TRUE VALUE

JOSEPH A. GREENWOOD

AND

MARION M. SANDOMIRE

*Navy Department, Washington, D. C.\**

One frequently encounters the need of a rational selection of sample size when it is desired to estimate the standard deviation. Changing the permissible error of the estimate from an absolute to a relative one is acceptable in many cases and permits an exact, *a priori* solution to the problem of sample size without involving any previous estimates.

**T**HERE ARE many practical engineering situations in which an estimate of the linear dispersion of a characteristic is desired. Examples of such are longitudinal or range error of a low-altitude bomb sight, or errors of a measuring device. Experience in working with the engineer has shown that 1) he grasps the idea of a standard deviation as a measure of dispersion; 2) he accepts the concept of expressing a sample standard deviation as being within a previously agreed-upon percentage of its true value for the normal population; 3) he has enough intuitive understanding of a confidence coefficient to be willing to take action accordingly. Thus, a quick means of telling him the required sample size for obtaining estimates within various confidence intervals for different confidence coefficients has been found useful. These estimates are those in which the half-length of the confidence interval about the true standard deviation is expressed as a percentage of the true standard deviation. The proposed solution for the required sample size for estimating this does not utilize any previous estimate of dispersion or its true value, and thus is widely applicable.

Let it be assumed that a normal population is being sampled. Let  $s^2$  be an estimate of the squared standard deviation of the population,  $\sigma^2$ , such that  $ns^2/\sigma^2$  is distributed as  $\chi^2$  with  $n$  degrees of freedom. Then the average value of  $s^2$  will be  $E(s^2) = \sigma^2$ , where  $\sigma^2$  is the true, but unknown, population variance. Let  $0 < u < 1$ . The proportion of the time that  $s$  lies within a given fraction  $u$  of  $\sigma$  will be known if we know the prob  $\{s > (1+u)\sigma\} = p_1$ , and the prob  $\{s < (1-u)\sigma\} = p_2$ .

The values of  $p_1$  and  $p_2$  were obtained through the following relation.

---

\* The opinions herein are not necessarily those of the Navy Department or the Naval Service at large.

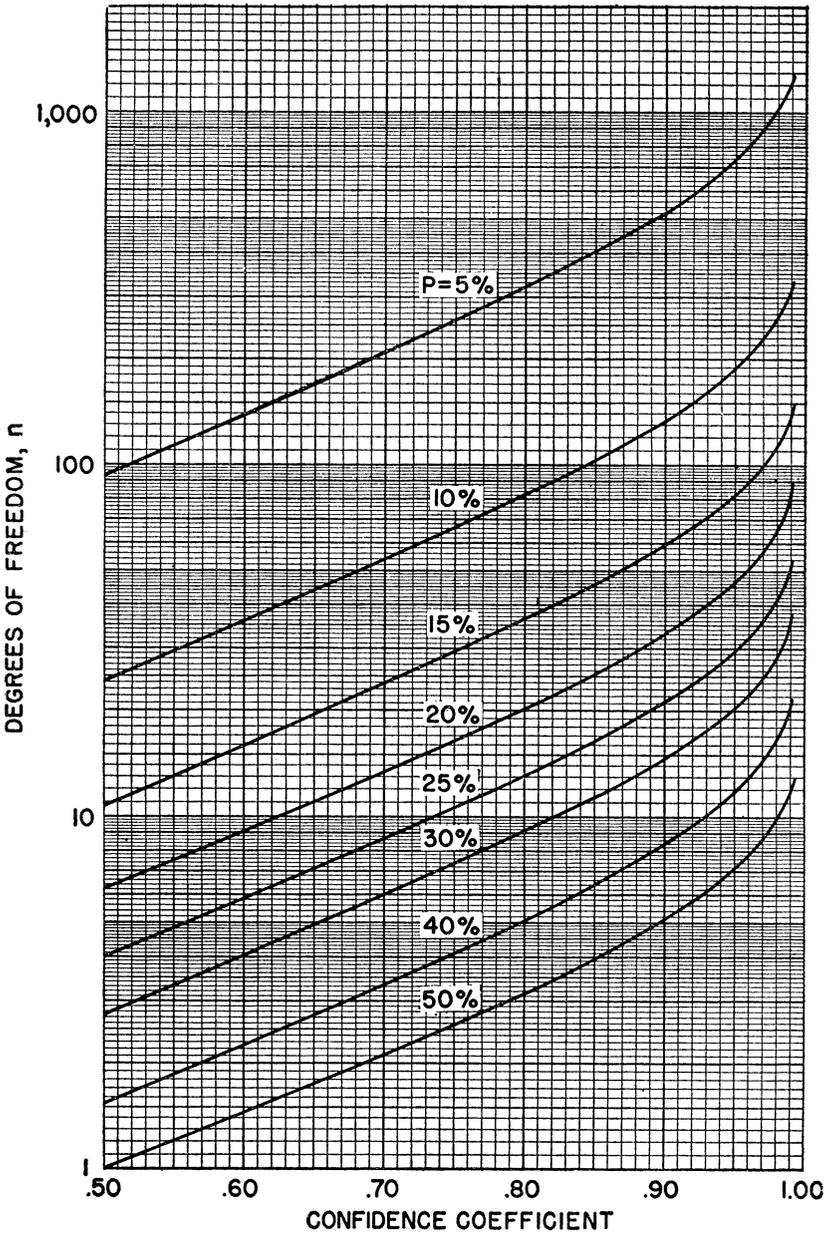


FIGURE 1

DEGREES OF FREEDOM REQUIRED TO ESTIMATE THE STANDARD DEVIATION WITHIN STATED PER CENT P OF ITS TRUE VALUE, WITH PRESCRIBED CONFIDENCE

When  $s > (1+u)\sigma$ , the quantity  $ns^2/\sigma^2 > n(1+u)^2$ . Thus,  $n(1+u)^2$  is that value of  $\chi_n^2$  such that the prob  $\{\chi_n^2 > n(1+u)^2\} = p_1$ . Similarly,  $n(1-u)^2$  is that value of  $\chi_n^2$  such that the prob  $\{\chi_n^2 < n(1-u)^2\} = p_2$ . Convenient values of  $n$  and  $u$  were selected and these probabilities were computed. Then points forming the curve for a constant value of  $u$  were plotted with coordinates  $(1-p_1-p_2)$  on the confidence-coefficient axis, and  $n$  on the degrees-of-freedom axis. In Figure 1,  $u$  is changed to per cent  $P$ .

When  $n \leq 100$ , linear interpolation between the logarithms of probabilities in the Thompson tables [1] was used. When  $n > 100$ , the Wilson-Hilferty formula [2]

$$\chi_n^2 = n \left( 1 - \frac{2}{9n} + y_p \sqrt{\frac{2}{9n}} \right)^3$$

was used to find the corresponding normal deviate  $y_p$ . It is clear from the graph that when any two of the three variables,  $P$ , degrees of freedom, and confidence coefficient, are given, the third variable is readily determined.

Sample size is determined from the number of degrees of freedom read and the considerations entering into the original design of the experiment. In a simple replicated test, sample size is greater by one than the number of degrees of freedom. In a more complex, multi-factor arrangement, the degrees of freedom obtained from the graph are considered as those corresponding to the residual error, and a sufficient number of measurements plus one must be added thereto to allow for the estimates associated with each of the main factors and their interactions as prescribed by the original design. In planning an experiment for estimating the standard deviation, the loss in this way of degrees of freedom, leading to an increase in total sample size, should be balanced against the need for the additional information to be gained from the multiple classification. In such cases the assumption must be satisfied concerning the homogeneity of the set of population variances pertaining to the different groupings.

*Example 1:* A series of radar pulses is to be sent out to a target and the strength of the return signal measured. How many readings under identical conditions shall be taken so that the standard deviation of the return signal strengths shall, with 0.80 confidence, be within 10 per cent of the true value? Solution—From the graph it is seen that  $n$  is 83. Sample size, in this case one greater than the degrees of freedom, is 84.

*Example 2:* How many bombs shall each of 6 bombardiers drop at a target in order that an average bombardier's standard deviation for

horizontal range shall, with 0.90 confidence, be within 15 per cent of the true standard deviation? Solution—From the graph it is seen that  $n=60$ . Then each of 6 bombardiers would contribute 10 degrees of freedom, or would drop 11 bombs. This results in a total of 66 bombs to be dropped.

It was considered of interest to compare the graphical results with a distribution of estimates that could be made from a set of random numbers. Dodd's table [3] contains 800 (in 16 columns of 50) "almost normally-distributed deviates" in random order, with an average close to zero ( $-.016$ ) and a standard deviation near 1 ( $.993$ ). If a proposed problem be one of determining the standard deviation to within 10 per cent of its true value, half of the time, Figure 1 indicates that  $n=24$ , or samples of 25 are required. Dodd's table furnished 32 samples of 25 values, for which standard deviations were calculated, and then converted to percentage errors of the true value of 1. It was found that 16, or just one-half, of these values were not over  $\pm 10$  per cent. In other words, estimates within 10 per cent of the true value were obtained half of the time.

A horizontal line of Figure 1 provides a series of distribution points for the expected frequency of errors obtained from samples of a given size. For samples of 25, the following are the expected frequencies for various errors, and the proportion calculated from the set of 32:

Maximum Error in Estimate of $\sigma$ , Per Cent	Expected Frequency, Per Cent	Observed Frequency Per Cent, in 32 Samples
10	.50	.50
15	.70	.69
20	.83	.81
25	.92	.91
30	.96	.97
40	.99	1.00

#### REFERENCES

- [1] Thompson, Catherine M., "Table of percentage points of the  $\chi^2$  distribution," *Biometrika*, 32 (1941), 188-189.
- [2] Wilson, E. B. and Hilferty, M. M., *Proceedings of the National Academy of Science* (1931), Wash., 17, 684.
- [3] Dodd, Edward L., "A transformation of Tippett random sampling numbers into numbers normally distributed," *Boletín Matemático* (Buenos Aires), 15 (1942), 73-77.