~~Designation:E122–07~~ Designation: E 122 – 09

An American National Standard

Standard Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process[1]

This standard is issued under the fixed designation E 122; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice covers simple methods for calculating how many units to include in a random sample in order to estimate with a specified precision, a measure of quality for all the units of a lot of material, or produced by a process. This practice will clearly indicate the sample size required to estimate the average value of some property or the fraction of nonconforming items produced by a production process during the time interval covered by the random sample. If the process is not in a state of statistical control, the result will not have predictive value for immediate (future) production. The practice treats the common situation where the sampling units can be considered to exhibit a single (overall) source of variability; it does not treat multi-level sources of variability.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E 456 Terminology Relating to Quality and Statistics

3. Terminology

3.1 *Definitions*: Unless otherwise noted, all statistical terms are defined in Terminology E 456.

3.2 *Symbols:* Symbols used in all equations are defined as follows:

E = the maximum acceptable difference between the true average and the sample average.
e = E/μ, maximum acceptable difference expressed as a fraction of μ.
f = degrees of freedom for a standard deviation estimate (7.5).
k = the total number of samples available from the same or similar lots.
μ = lot or process mean or expected value of X, the result of measuring all the units in the lot or process.
μ_0 = an advance estimate of μ.
N = size of the lot.
n = size of the sample taken from a lot or process.
n_j = size of sample j.
n_L = size of the sample from a finite lot (7.4).
p' = fraction of a lot or process whose units have the nonconforming characteristic under investigation.
p_0 = an advance estimate of p'.
p = fraction nonconforming in the sample.
R = range of a set of sampling values. The largest minus the smallest observation.
R_j = range of sample j.
$\bar{R}$ = $\sum_{j=1}^{k} R_j/k$, average of the range of k ~~samples, all of the same size (8.2.2).~~ samples, all of the same size (8.2.2).
σ = lot or process standard deviation of X, the result of measuring all of the units of a finite lot or process.
σ_0 = an advance estimate of σ.

[1] This practice is under the jurisdiction of ASTM Committee E11 on Quality and Statistics and is the direct responsibility of Subcommittee E11.10 ~~on Sampling.~~ ~~Current edition approved Oct. 1, 2007. Published November 2007 . Originally published as E122–89. Last previous edition approved in 2000 as E122–00.~~ on Sampling/Statistics.

Current edition approved Aug. 1, 2009. Published September 2009. Originally approved in 1958. Last previous edition approved in 2007 as E 122–07.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

s = $[\sum_{k=1}^{n} (X_i - \bar{X})^2/(n-1)]^{1/2}$, an estimate of the standard deviation σ from n observation, X_i, $i = 1$ to n.
$\bar{s}$ = $\sum_{j=1} S_j/k$, average s from k ~~samples all of the same size (8.2.1).~~ samples all of the same size (8.2.1).
s_p = pooled (weighted average) s from k samples, not all of the same size (8.2).
s_j = standard deviation of sample j.
~~t = a factor (the 99.865th percentile of the Student's distribution) corresponding to the degrees of freedom f_o of an advance estimate σ_o (5.1).~~
V_o = an advance estimate of V, equal to δ_o/μ_o.
v = $s/\bar{X}$, the coefficient of variation estimated from the sample.
v_p = pooled (weighted average) of v from k samples (8.3).
v_j = coefficient of variation from sample j.
X = numerical value of the characteristic of an individual unit being measured.
$\bar{X}$ = $\sum_{i=1}^{n} X_i/n$ average of n observations, X_i, $i = 1$ to n.

4. Significance and Use

4.1 This practice is intended for use in determining the sample size required to estimate, with specified precision, a measure of quality of a lot or process. The practice applies when quality is expressed as either the lot average for a given property, or as the lot fraction not conforming to prescribed standards. The level of a characteristic may often be taken as an indication of the quality of a material. If so, an estimate of the average value of that characteristic or of the fraction of the observed values that do not conform to a specification for that characteristic becomes a measure of quality with respect to that characteristic. This practice is intended for use in determining the sample size required to estimate, with specified precision, such a measure of the quality of a lot or process either as an average value or as a fraction not conforming to a specified value.

5. Empirical Knowledge Needed

5.1 Some empirical knowledge of the problem is desirable in advance.

5.1.1 We may have some idea about the standard deviation of the characteristic.

5.1.2 If we have not had enough experience to give a precise estimate for the standard deviation, we may be able to state our belief about the range or spread of the characteristic from its lowest to its highest value and possibly about the shape of the distribution of the characteristic; for instance, we might be able to say whether most of the values lie at one end of the range, or are mostly in the middle, or run rather uniformly from one end to the other (Section 9).

5.2 If the aim is to estimate the fraction nonconforming, then each unit can be assigned a value of 0 or 1 (conforming or nonconforming), and the standard deviation as well as the shape of the distribution depends only on p', the fraction nonconforming in the lot or process. Some rough idea concerning the size of p' is therefore needed, which may be derived from preliminary sampling or from previous experience.

5.3 More knowledge permits a smaller sample size. Seldom will there be difficulty in acquiring enough information to compute the required size of sample. A sample that is larger than the equations indicate is used in actual practice when the empirical knowledge is sketchy to start with and when the desired precision is critical.

5.4 The precision of the estimate made from a random sample may itself be estimated from the sample. This estimation of the precision from one sample makes it possible to fix more economically the sample size for the next sample of a similar material. In other words, information concerning the process, and the material produced thereby, accumulates and should be used.

6. Precision Desired

6.1 The approximate precision desired for the estimate must be prescribed. That is, it must be decided what maximum deviation, E, can be tolerated between the estimate to be made from the sample and the result that would be obtained by measuring every unit in the lot or process.

6.2 In some cases, the maximum allowable sampling error is expressed as a proportion, e, or a percentage, 100 e. For example, one may wish to make an estimate of the sulfur content of coal within 1 %, or $e = 0.01$.

7. Equations for Calculating Sample Size

7.1 Based on a normal distribution for the characteristic, the equation for the size, n, of the sample is as follows:

$$n = (3\sigma_o/E)^2 \tag{1}$$

The multiplier 3 is a factor corresponding to a low probability that the difference between the sample estimate and the result of measuring (by the same methods) all the units in the lot or process is greater than E. The value 3 is recommended for general use. With the multiplier 3, and with a lot or process standard deviation equal to the advance estimate, it is practically certain that the sampling error will not exceed E. Where a lesser degree of certainty is desired a smaller multiplier may be used (Note 1).

NOTE 1—For example, multiplying by 2 in place of 3 gives a probability of about 45 parts in 1000 that the sampling error will exceed E. Although distributions met in practice may not be normal, the following text table (based on the normal distribution) indicates approximate probabilities:

Factor	Approximate Probability of Exceeding *E*
3	0.003 or 3 in 1000 (practical certainty)
2.56	0.010 or 10 in 1000
2	0.045 or 45 in 1000
1.96	0.050 or 50 in 1000 (1 in 20)
1.64	0.100 or 100 in 1000 (1 in 10)

7.1.1 If a lot of material has a highly asymmetric distribution in the characteristic measured, the sample size as calculated in Eq 1 may not be adequate. There are two things to do when asymmetry is suspected.

7.1.1.1 Probe the material with a view to discovering, for example, extra-high values, or possibly spotty runs of abnormal character, in order to approximate roughly the amount of the asymmetry for use with statistical theory and adjustment of the sample size if necessary.

7.1.1.2 Search the lot for abnormal material and segregate it for separate treatment.

7.2 There are some materials for which σ varies approximately with μ, in which case $V\ (=\sigma/\mu)$ remains approximately constant from large to small values of μ.

7.2.1 For the situation of 7.2, the equation for the sample size, n, is as follows:

$$n = (3\ V_o/e)^2 \tag{2}$$

If the relative error, e, is to be the same for all values of μ, then everything on the right-hand side of Eq 2 is a constant; hence n is also a constant, which means that the same sample size n would be required for all values of μ.

7.3 If the problem is to estimate the lot fraction nonconforming, then σ_o^2 is replaced by $p_o(1 - p_o)$ so that Eq 1 becomes:

$$n = (3/E)^2\, p_o(1 - p_o) \tag{3}$$

7.4 When the average for the production process is not needed, but rather the average of a particular lot is needed, then the required sample size is less than Eq 1, Eq 2, and Eq 3 indicate. The sample size for estimating the average of the finite lot will be:

$$n_L = n/[1 + (n/N)] \tag{4}$$

where n is the value computed from Eq 1, Eq 2, or Eq 3. This reduction in sample size is usually of little importance unless n is 10 % or more of N

7.5 When the information on the standard deviation is limited, a sample size larger than indicated in Eq 1, Eq 2, and Eq 3 may be appropriate. When the advance estimate σ_0 is based on f degrees of freedom, the sample size in Eq 1 may be replaced by

$$n = (3\sigma_0/E)^2\ (1 + \sqrt{2/f}) \tag{6}$$

NOTE 2—The standard error of a sample variance with f degrees of freedom, based on the normal distribution, is $\sqrt{2\sigma^4/f}$. The factor $(1 + \sqrt{2/f})$ has the effect of increasing the preliminary estimate σ_0^2 by one times its standard error.

8. Reduction of Empirical Knowledge to a Numerical Value of σ_o (Data for Previous Samples Available)

8.1 This section illustrates the use of the equations in Section 7 when there are data for previous samples.

8.2 *For Eq 1*—An estimate of σ_o can be obtained from previous sets of data. The standard deviation, s, from any given sample is computed as: ~~s=[i=1~~

$$s = [\sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)]^{1/2} \tag{6}$$

The s value is a sample estimate of σ_o. A better, more stable value for σ_o may be computed by pooling the s values obtained from several samples from similar lots. The pooled s value s_p for k samples is obtained by a weighted averaging of the k results from use of Eq ~~5.~~6.

$$s_p = [\sum_{j=1}^{k} (n_j - 1)s_j^2 / \sum_{j=1}^{k} (n_j - 1)]^{1/2} \tag{7}$$

8.2.1 If each of the previous data sets contains the same number of measurements, n_j, then a simpler, but slightly less efficient estimate for σ_o may be made by using an average ($\bar{s}$) of the s values obtained from the several previous samples. The calculated $\bar{s}$ value will in general be a slightly biased estimate of σ_o. An unbiased estimate of σ_o is computed as follows:

$$\sigma_o = \frac{\bar{s}}{c_4} \tag{8}$$

where the value of the correction factor, c_4, depends on the size of the individual data sets (n_j) (Table 1[3]).

8.2.2 An even simpler, and slightly less efficient estimate for σ_o may be computed by using the average range ($\bar{R}$) taken from the several previous data sets that have the same group size. ~~σ~~

[3] ~~*ASTM Manual on Presentation of Data and Control Chart Analysis, ASTM STP 15D*, 1976, Part 3, Table 27.~~ASTM Manual on Presentation of Data and Control Chart Analysis, ASTM MNL 7A, 2002, Part 3