Designation:E141–91(Reapproved 1997) Designation: E 141 – 91 (Reapproved 2003)$^{\epsilon 1}$

An American National Standard

Standard Practice for Acceptance of Evidence Based on the Results of Probability Sampling[1]

This standard is issued under the fixed designation E 141; 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.

ϵ^1 NOTE—Editorial changes were made throughout in November 2003.

1. Scope

1.1 This practice presents rules for accepting or rejecting evidence based on a sample. Statistical evidence for this practice is in the form of an estimate of a proportion, an average, a total, or other numerical characteristic of a finite population or lot. It is an estimate of the result which would have been obtained by investigating the entire lot or population under the same rules and with the same care as was used for the sample.

1.2 One purpose of this practice is to describe straightforward sample selection and data calculation procedures so that courts, commissions, etc. will be able to verify whether such procedures have been applied. The methods may not give least uncertainty at least cost, they should however furnish a reasonable estimate with calculable uncertainty.

1.3 This practice is primarily intended for one-of-a-kind studies. Repetitive surveys allow estimates of sampling uncertainties to be pooled; the emphasis of this practice is on estimation of sampling uncertainty from the sample itself. The parameter of interest for this practice is effectively a constant. Thus, the principal inference is a simple point estimate to be used as if it were the unknown constant, rather than, for example, a forecast or prediction interval or distribution devised to match a random quantity of interest.

1.4 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*

E 105 Practice for Probability Sampling of Materials[2]

E 122 Practice for Choice of Sample Size to Estimate a Measure of Quality for a Lot or Process[2]

E 178 Practice for Dealing with Outlying Observations[2]

E 456 Terminology for Statistical Methods[2]

NOTE 1—Practice E 105 provides a statement of principles for guidance of ASTM technical committees and others in the preparation of a sampling plan for a specific material. Practice E 122 aids in deciding on the required sample size. Practice E 178 helps insure better behaved estimates. Terminology E 456 provides definitions of statistical terms used in this standard.

3. Terminology

3.1 *Definitions:*

3.1.1 *Equal Complete Coverage Result*, *n*— the numerical characteristic (θ) of interest calculated from observations made by drawing randomly from the frame, all of the sampling units covered by the frame.

3.1.1.1 *Discussion*—Locating the units and evaluating them are supposed to be done in exactly the same way and at the same time as was done for the sample. The quantity itself is denoted θ. The equal complete coverage result is never actually calculated. Its purpose is to serve as the objectively defined concrete goal of the investigation. The quantity θ may be the population mean, ($\bar{Y}$), total (Y), median (M), the proportion (P), or any other such quantity.

3.1.2 *frame*, *n*—a list, compiled for sampling purposes, which designates all of the sampling units (items or groups) of a population or universe to be considered in a specific study.

3.1.2.1 *Discussion*—The list may cover a specific shipment or lot, all households in a county, a state, or country; for example,

[1] This practice is under the jurisdiction of ASTM Committee E-11E11 on Quality and Statistical Methods and is the direct responsibility of Subcommittee E11.10 on Sampling and Data Analysis. Sampling.

Current edition approved August 15, 1991. Published November 1991. Originally published as E 141 – 59 T. Last previous edition E 141 – 69 (1975).

[2] *Annual Book of ASTM Standards*, Vol 14.02.

any population of interest. Every sampling unit in the frame (1) has a unique serial number, which may be preassigned or determined by some definite rule, (2) has an address—a complete and clear instruction (or rules for its formulation) as to where and when to make the observation or evaluation, (3) is based on physically concrete clerical materials such as directories, dials of clocks or of meters, ledgers, maps, aerial photographs, etc., referred to in the addresses.

3.1.3 *sample*, *n*—a group of items, observations, test results, or portions of material, taken from a larger collection of such items; it provides information for decisions concerning the larger collection.

3.1.3.1 *Discussion*—A particular sample is identified by the set of serial numbers from the randomization device and by the addresses on the frame generated by those serial numbers.

3.1.4 *sampling unit, n*—an item, test specimen or portion of material that is to be subjected to evaluation as part of the sampling plan.

3.1.4.1 *Discussion*—If it is not feasible to select test specimens or laboratory samples individually, the sampling unit may be a group of items, for example, a row, an entire case of items, or a prescribed area (as in the examination of a finishing process).

3.1.4.2 By a more expensive method of measurement (future time, more elaborate frame) it may be possible to define a quantity, θ', as a target parameter or ideal goal of an investigation. Criticism that holds θ to be an inappropriate goal should demonstrate that the numerical difference between θ and θ' is substantial. Measurements may be imprecise but so long as measurement errors are not too biased, a large size of the lot or population, N, insures that θ and θ' are essentially equal.

4. Significance and Use

4.1 This practice is designed to permit users of sample survey data to judge the trustworthiness of results from such surveys. Section 5 gives extended definitions of the concepts basic to survey sampling and the user should verify that such concepts were indeed used and understood by those who conducted the survey. What was the frame? How large (exactly) was the quantity N? How was the parameter θ estimated and its standard error calculated? If replicate subsamples were not used, why not?

4.2 Adequate answers should be given for all questions. There are many acceptable answers to the last question. If the sample design was relatively simple, such as simple random or stratified, then good estimates of sampling variance are easily available. If a more complex design was used then methods such as discussed in [1] may be acceptable. Replicate subsamples is the most straightforward way to estimate sampling variances when the survey design is complex.

4.3 Once the survey procedures that were used satisfy Section 5, consult Section 4 to see if any increase in sample size is needed. The calculations for making it are objectively described in Section 4.

4.4 Refer to Section 6 to guide in the interpretation of the uncertainty in the reported value of the parameter estimate, θ, i.e. the value of its standard error, se(θ). The quantity se(θ) should be reviewed to verify that the risks it entails are commensurate with the size of the sample.

5. Descriptive Terms and Procedures

5.1 *Probability Sampling Plans*—include instructions for using either:

5.1.1 carefully prepared tables of random number,

5.1.2 computer algorithms, carefully programmed and run on a large computer, to generate pseudo-random numbers or,

5.1.3 certifiably honest physical devices, such as coin flips, to select the sample units so that inferences may be drawn from the test results and decisions may be made with risks correctly calculated by probability theory.

5.1.4 Such plans are defined and their relative advantages discussed in [**1**], [**2**] and [**6**].

5.2 *Replicate Subsamples*—a number of disjoint samples, each one separately drawn from the frame in accord with the same probability sampling plan. When appropriate, separate laboratories should each work on separate replicate subsamples and teams of investigators should be assigned to separate replicate subsamples. This approach insures that the calculated standard error will not be a systematic underestimate. Such subsamples were called interpenetrating in [**5**] where many of their basic properties were described. See [**2**] for further theory and applications.

5.2.1 *Discussion*—For some types of material a sample selected with uniform spacing along the frame (systematic sample) has increased precision over a selection made with randomly varying spacings (simple random sample). Examples include sampling mineral ore or grain from a conveyor belt or sampling from a list of households along a street. If the systematic sample is obtained by a single random start the plan is then a probability sampling plan, but it does not permit calculating the standard error as required by this practice. After dividing the sample size by an integer k (such as $k = 4$ or $k = 10$) and using a random start for each of k replicate subsamples, some of the increased precision of systematic sampling (and a standard error on $k - 1$ degrees of freedom) can be achieved.

5.2.2 *Audit Subsample*—a small subsample of the survey sample (as few as 10 observations may be adequate) for review of all procedures from use of the random numbers through locating and measurement, to editing, coding, data entry and tabulation. Selection of the audit subsample may be done by putting the n sample observations in order as they are collected, calculating the nearest integer to $\sqrt{n}$, or some other convenient integer, and taking this number to be the spacing for systematic selection of the audit subsample. The review should uncover any gross departures from prescribed practices or any conceptual misunderstandings in the definitions. If the audit subsample is large enough (say 30 observations or more) the regression of audited values on initial observations may be used to calibrate the estimate. This technique is the method of two-phase sampling as discussed in [**1**]. Helpful discussion of an audit appears in [**2**].

5.2.3 *Estimate*—a quantity calculated on the n sample observations in the same way as the equal complete coverage result θ would have been calculated from the entire set of N possible observations of the population; the symbol θ denotes the estimate. (In calculating θ, replicate subsample membership is ignored.)

5.2.3.1 *Discussion*—An estimate has a sampling distribution induced from the randomness in sample selection. The equal complete coverage result is effectively a constant while any estimate is only the value from one particular sample. Thus, there is a mean value of the sampling distribution and there is also a standard deviation of the sampling distribution.

5.2.4 *Standard Error*—the quantity computed from the observations as an estimate of the sampling standard deviation of the estimate; *se* (θ) denotes the standard error.

5.2.4.1 *Example 1*—When θ is the population average of the N quantities and a simple random sample of size *n* was drawn, then the sample average *y* becomes the usual estimate θ, where

$$\theta = \bar{y} = \sum_{i=1}^{n} y_i / n. \quad (1)$$

$$\theta = \bar{y} = \sum_{i=1}^{n} y_i / n. \quad (1)$$

The quantities y_1, y_2, ..., y_n denote the observations. The standard error is calculated as:

$$se(\theta) = se(\bar{y}) = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n(n-1)}. \quad (2)$$

$$se(\theta) = se(\bar{y}) = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n(n-1)}. \quad (2)$$

There are $n - 1$ degrees of freedom in this standard error. When the observations are:

81.6, 78.7, 79.7, 78.3, 80.9, 79.5, 79.8, 80.3, 79.5, 80.7

then $y = 79.90$ and $se(\bar{y}) = 0.32$. As this example illustrates, formula (2) is correct when *k* replaces *n* and subsample estimates are used in place of observations.

5.2.4.2 *Example 2 on the Finite Population Correction (fpc)*—Multiplying se $(\bar{y})$ by $\sqrt{1 - n/N}$ is always correct when the goal of the survey is to estimate the finite population mean ($\theta = Y$). Using the previous data and if $N = 50$, then $se(\bar{y})$ becomes $se(\bar{y}) = 0.28$ after applying the fpc. If random measurement error exists in the observations, then θ′ based on a reference measurement method may be a more appropriate survey goal than θ (see section 4.1.4.1). If so, then $se(\bar{y})$ would be further adjusted upward by an amount somewhat less than the downward adjustment of the fpc. Both of these adjustments are often numerically so small that these adjustments may be omitted—leaving $se(\bar{y})$ of (2) as a slight overestimate.

5.2.4.3 *Example 3*—If the quantity of interest is (a) a proportion or (b) a total and the sample is simple random then the above formulas are still applicable. A proportion is the mean of zeroes and ones, while the total is a constant times the mean. Thus: (a) when θ is taken to be the population proportion ($\theta = P$) then;

$$\theta = p = \Sigma y_i / n = a/n \quad (3)$$

where:

a is the number of units in the sample with the attribute, and

$$se(p) = \sqrt{p(1-p)/(n-1)} \quad (4)$$

(b) when θ = the population total (θ = Y) then

$$\theta = N\bar{y} \text{ and } se(\theta) = N se(\bar{y}) \quad (5)$$

If a simple random sample of size $n = 200$ has $a = 25$ items with the attribute then the conclusion is $\theta = 0.125$ and $se(\theta) = 0.023$ on 199 degrees of freedom.

5.2.4.4 *Example 4*. If θ is a parameter other than a mean or if the sample design is complex, then replicate subsamples should be used in the sample design. Denote the *k* separate estimates as θ_i, $i = 1, 2, ..., k$ and denote by θ the estimate based on the whole sample. The average of the θ_i will be close to, but in general not equal to θ. The standard error of θ is calculated as:

$$se(\theta) = \sqrt{\sum_{i=1}^{k} (\theta_i - \theta)^2 / k(k-1)} \quad (6)$$

$$se(\theta) = \sqrt{\sum_{i=1}^{k} (\theta_i - \theta)^2 / k(k-1)} \quad (6)$$

where θ is the average of the θ_i. The standard error is based on $k - 1$ degrees of freedom.

The following estimates of percent "drug-in-suit" sales of prescription drugs were based on 20 replicate subsamples; each followed a stratified cluster sampling design. The separate estimates were: 6.8, 7.1, 8.4, 9.5, 8.6, 4.1, 3.7, 3.2, 3.8, 5.8, 8.8, 5.0,