INTERNATIONAL STANDARD

ISO 3534-4

First edition
2014-04-15

Statistics — Vocabulary and symbols —

Part 4: Survey sampling

Statistique — Vocabulaire et symboles —

Partie 4: Échantillonnage d'enquête

Reference number
ISO 3534-4:2014(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 1, *Terminology and symbols*.

ISO 3534 consists of the following parts, under the general title *Statistics — Vocabulary and symbols*:

— *Part 1: General statistical terms and terms used in probability*

— *Part 2: Applied statistics*

— *Part 3: Design of experiments*

— *Part 4: Survey sampling*

Introduction

Survey sampling is essentially a strategy of planning for the collection of information on a population. In cases where all entities in the population can be listed, statistical methodologies of sampling without replacement play a key role. The design of a survey and its implementation depends on the type of questions to be addressed, the degree of generality to be attached to the conclusions, and ultimately, the resources available for conducting the survey and analysis of the results.

Political polls, customer satisfaction surveys, and personal interviews are pervasive in modern society as mechanisms to provide decision makers with information to formulate or to adjust their strategies. The news media frequently reports results from sampling efforts that typically address a country's pulse with regard to political leadership. This is by no means a recent phenomenon as sampling (especially census work) has occurred for thousands of years. Survey sampling as a general methodology and finite population sampling as its rigorous theoretical basis are the subject areas of this part of ISO 3534.

The methodology of survey sampling consists of a process of selecting a sample of items from a population, measuring these items, and then estimating population characteristics based on the results from the sample. Reference [4] has defined the concept of a survey with the following description.

1) A survey concerns a set of items comprising the population.

2) A survey involves a population having one or more measurable properties.

3) A survey has an objective to describe the population according to one or more parameters defined in terms of these properties.

4) A survey requires operationally a representation of the population (frame) such as a list of items in order to facilitate the measurements on individual items.

5) A survey is applied to a subset of items from the frame that are selected according to a sampling design consisting of a sample size and a probability mechanism for selection.

6) A survey proceeds via extracting measurements of the items in the sample.

7) A survey needs an associated estimation process to obtain parameter estimates for the population.

This brief introduction by no means captures all of the subtleties and advancements in survey sampling that have evolved over the centuries and especially in the past several decades with improved computational capabilities. Advancements have progressed in tandem with real applications.

Some definitions in this part of ISO 3534 are adopted from ISO 3534-1:2006 or ISO 3534-2:2006. If the adopted definition is identical with the original one, reference in square brackets is added to the definition and if some differences exist, they are noted.

Statistics — Vocabulary and symbols —

Part 4:
Survey sampling

1 Scope

This part of ISO 3534 defines the terms used in the field of survey sampling and can be used in the drafting of other International Standards.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 3534-1:2006, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

3 Terms and Definitions

For the purposes of this document, the terms and definitions given in ISO 3534-1:2006 and ISO 3534-2:2006 and the following apply.

3.1 General terms

3.1.1
population
totality of items under consideration

[SOURCE: ISO 3534-1:2006, 1.1]

Note 1 to entry: A population can be real and finite, real and infinite, or completely hypothetical. Of particular interest in this part of ISO 3534 is a *finite population* (3.1.2). Much of the field of *sample survey* (3.1.20) concerns finite populations. The term population has superceded the term universe in usage. Population should be construed to involve a fixed point in time, as populations can evolve over time.

3.1.2
finite population
population (3.1.1) which consists of a limited number of items

Note 1 to entry: *Survey sampling* (3.1.21) concentrates solely on applications with a finite number of items in the population. The number of items could be very large (for example, hybrid automobiles in Europe, artefacts in a museum, sheep in New Zealand) but their number is finite. The number of items in the population is generally denoted as N. The specific value of N may or may not be known explicitly prior to conducting the survey.

EXAMPLE 1 The registry of citizens of a country is an example of a finite population with a known size.

EXAMPLE 2 Although, generally, the population size *N* is known in advance, this situation need not be the case. For example, the proportion of hybrid cars is of interest and observations could be taken at a checkpoint (e.g. toll booth or toll plaza). The number of cars that pass through the booth on a given day would not be known in advance, although the investigators would likely have a rough idea of the number from previous history. Perhaps a digital photo is taken of a select number of these vehicles to determine if they are hybrid cars.

3.1.3
subpopulation
well-defined subset of the *population* (3.1.1)

Note 1 to entry: *Sample surveys* (3.1.20) often have multiple objectives. Although the primary objective may concern the population as a whole, it is possible that select subsets are also of interest. For the example noted in 3.1.2, hybrid vehicles or, alternatively, sub-compact automobiles, comprise subpopulations that may warrant particular interest. In some situations, the actual size of the subpopulation is unknown (e.g. number of teen-aged children among tourists visiting EuroDisney) and the interest may centre on estimating this value.

Note 2 to entry: In ISO 3534-2:2006, 1.2.3, the definition of subpopulation is “part of a population.” For *survey sampling* (3.1.21), subpopulations that are well defined (specifically identifiable) are of primary interest rather than consideration of arbitrary “parts” of a population.

EXAMPLE Children in school in a province constitute a subpopulation of residents of the province. Working adults in the province is another subpopulation among the residents of the province. Of interest but likely to be more difficult to identify are homeless people in the province. The size of such a subpopulation is usually unknown.

3.1.4
superpopulation
expanded *population* (3.1.1) that includes the population of interest

Note 1 to entry: For inferential or assessment purposes, it can prove useful to imagine that the population of interest is embedded in a larger population having the base population as a special case. Such a theoretical construct facilitates the development of optimal *sampling designs* (3.1.28) and allows the calculation of sampling design properties. The population of values can be treated as a *random sample* (3.1.10) from a hypothetical superpopulation as opposed to a set of fixed values from which random selection is used to constitute a *sample* (3.1.8). According to Reference [2], the superpopulation concept can be given several interpretations. One of the interpretations is that the *finite population* (3.1.2) is actually drawn from a larger universe. This is the superpopulation concept in its purest form. The superpopulation approach can be a useful device for incorporating the treatment of *non-sampling errors* (3.2.10) in *survey sampling* (3.1.21).

EXAMPLE For a stable country (consistent political boundaries without immigration or emigration), a superpopulation could be the citizenry over the centuries. Thus, a decennial *census* (3.1.19) in such a country could reflect an individual observation from its population size at a specific time.

3.1.5
sampling unit
unit
one of the individual parts into which a *population* (3.1.1) is divided

[SOURCE: ISO 3534-2:2006, 1.2.14]

Note 1 to entry: A population consists of a number of sampling units. The population could be divided into groups of units which are distinct, non-overlapping, identifiable, observable, and convenient for sampling. Depending on the circumstances, the smallest part of interest can be an individual, a voucher, a household, a school district, or an administrative unit. This definition allows for the possibility in complex settings to have distinct sampling units comprised of varying number of units. At a high level, the sampling unit could be school districts. Within various school districts, the sampling unit could be individual households. Within a household, the sampling unit could be school-age children.

Note 2 to entry: Every element of the population should belong to exactly one sampling unit. In some cases, the population consists of individual elements, subunits, or items, but owing to the purpose of the sampling study, it may be appropriate to group the individual elements into higher-level entities which then are treated as the sampling unit of interest. For instance, the grouping could constitute *clusters* (3.1.6), each of which consists of a set of elements.

EXAMPLE In a *multi-stage sampling* (3.1.40) project, the first stage could use provinces as the primary sampling units. In the second stage, the sampling units could be counties. In the third stage, the sampling units could be incorporated towns.

3.1.6
cluster
part of a *population* (3.1.1) divided into mutually exclusive groups related in a certain manner

Note 1 to entry: For economies of *sampling* (3.1.16), it may be much more efficient to sample collections of *sampling units* (3.1.5) that constitute clusters. *Cluster sampling* (3.1.38) is useful when the frame of sampling units is not available. Cluster sampling can also be an integral part of *multi-stage sampling* (3.1.40), where a first-level stage is given by towns, followed by a stage with apartment/condominium buildings as the next level cluster, and then finally specific floors/stages/levels of the building. At the lowest level stage, all sampling units are examined.

Note 2 to entry: The definition given here differs from ISO 3534-2:2006, 1.2.28 which states "part of a population divided into mutually exclusive groups *of sampling units* related in a certain manner."

The phrase "of sampling units" is omitted in this standard to reflect sampling practices, such as multi-stage sampling.

EXAMPLE In investigating medical insurance fraud (overpayment to the provider of medical services), it is easier to obtain a *sample* (3.1.8) of patients and then examine all of their submitted claims than to consider the *population* (3.1.1) of claims across many patients. Common examples of clusters include a household or residents in a given building, agricultural fields in villages, patients of medical practitioners, and students in classes in a school.

3.1.7
stratum

subpopulation (3.1.3) considered to be more homogeneous with respect to the characteristics investigated than that within the total *population* (3.1.1)

Note 1 to entry: The plural form of stratum is strata.

Note 2 to entry: Stratification is the division of a population into mutually exclusive and exhaustive strata.

Note 3 to entry: The fundamental aspect of stratification is that the strata should be homogeneous with respect to the characteristic of interest in the population. On the other hand, if the stratification is not related to the characteristic of interest (but was performed for administrative convenience), there may be little or no gain in the precision of estimation of the population characteristic of interest. Further, it is advantageous if the variable or variables that are the basis of the stratification are highly correlated with the characteristic of interest in the population.

Note 4 to entry: Stratification can proceed along a geographical basis with the presumption that contiguous areas may provide more homogeneous groupings of the *sampling units* (3.1.5). Such stratification may also have economic and administrative advantages in the efficiency in conducting the survey.

Note 5 to entry: A fundamental difference between *cluster* (3.1.6) and stratum is that a stratum ought to consist of rather homogeneous items whereas a cluster could consist of heterogeneous items. A common example is the use of a household as a cluster that is generally heterogeneous with respect to ages of the members of the household.

Note 6 to entry: A compatible definition is given in ISO 3534-2:2006, 1.2.29, but it is formulated slightly incorrectly. A more correct definition is given here.

EXAMPLE Two examples are the stratification of a cat or dog population into breeds and a human population stratified by gender and social class.

3.1.8
sample
subset of a *population* (3.1.1) made up of one or more *sampling units* (3.1.5)

[SOURCE: ISO 3534-2:2006, 1.2.17]

Note 1 to entry: The selection of the sample should occur according to some specified procedure so as to obtain information regarding the population. The sampling units chosen in the sample could be items, numerical values, or even abstract entities, depending on the population of interest.

Note 2 to entry: Although the definition suggests that any subset of the population could be a sample, in practice, there is an underlying objective for constituting the sample. In other words, a sample is selected for a specific reason in support of a survey. Even a *census* (3.1.19) that intends to examine every item in the population could end up examining a subset owing to difficulties in contacting every individual in the population.

3.1.9
sample size
n
number of *sampling units* (3.1.5) in a *sample* (3.1.8)

[SOURCE: ISO 3534-2:2006, 1.2.26]

Note 1 to entry: Determination of the sample size occurs in virtually every *sample survey* (3.1.20) application. A typical approach to determining the sample size is to specify a bound on the true but unknown population characteristic to be estimated and to equate a function of the variance of the estimator to this bound. In other words, a sample size is computed such that the estimated population characteristic is within a pre-specified difference from the population characteristic.

Note 2 to entry: In complex surveys, the sample size refers to the ultimate number of items in the final stage in the sampling. A further complication in surveys is that the planned sample size could be the potential sample size, but owing to *non-response* (3.2.11), the actual sample size may be less than that determined by fixing the margin of error and the level of significance. There may be a difference between planned and actual sample size due to many possible unforeseen circumstances.

3.1.10
random sample
sample (3.1.8) constituted by a method of random selection

[SOURCE: ISO 3534-1:2006, 1.6]

Note 1 to entry: The method of random selection can be such that the actual probability of selection of *sampling units* (3.1.5) in the sample cannot be determined in advance nor at the conclusion of the study. If the probabilities of selection of each sampling unit can be determined, then the random sample is referred to more specifically as a *probability sample* (3.1.13).

Note 2 to entry: When the sample of *n* sampling units is selected from a *finite population* (3.1.2), each of the possible combinations of *n* sampling units will have a particular probability of being taken. For survey *sampling plans* (3.1.24), the particular probability for each possible combination can be calculated in advance. The probability of being selected need not be identical for each sampling unit, depending on the *sampling design* (3.1.28) chosen.

Note 3 to entry: For *survey sampling* (3.1.21) from a finite population, a random sample can be selected by different sampling plans such as *stratified sampling* (3.1.32), *systematic sampling* (3.1.29), *cluster sampling* (3.1.38), sampling with probability of *sampling proportional to size* (3.1.44) of an *auxiliary variable* (3.2.15), and many other possibilities.

Note 4 to entry: Of particular interest are the actual observed values associated with the items in the random sample. The values may be quantitative or reflect the presence of a specific characteristic. Results obtained in the random sample provide the basis for understanding the *population* (3.1.1) as a whole. In particular, a random sample is required for the use of inferential statistical methods in the context of survey sampling.

Note 5 to entry: The definition given in this entry is, as noted, the same as that given in ISO 3534-1:2006, 1.6. This definition presumes that the concept of random selection is understood from the context of probability theory. Less formally, randomness in survey sampling involves a chance mechanism in the choice of sampling units placed into the sample in contrast to a systematic or deterministic manner.

3.1.11
random sampling
act of forming a *random sample* (3.1.10)

Note 1 to entry: The *sampling* (3.1.16) of *n sampling units* (3.1.5) is taken from a *population* (3.1.1) in such a way that each of the possible combinations of *n* sampling units has a particular probability of being taken which can be difficult or impossible to determine. This definition differs from that given in ISO 3534-2:2006, 1.3.5.

EXAMPLE A computer program that employs a random number generator could be used to obtain a random sample from a registry of individuals in a country or province.

3.1.12
simple random sample
probability sample (3.1.13) with each subset of a given size having the same probability of selection

Note 1 to entry: This definition is in harmony with the definition given in ISO 3534-2:2006, 1.2.24, although the wording here is slightly different.

3.1.13
probability sample
random sample (3.1.10) in which the probability of selection of every possible *sample* (3.1.8) can be determined

Note 1 to entry: The *selection probability* (3.1.15) of each *sampling unit* (3.1.5) can be determined.

EXAMPLE In *sampling proportional to size* (3.1.44), the selection probability is related to a specific *auxiliary variable* (3.2.15).

3.1.14
representative sample
sample (3.1.8) for which the observed values have the same distribution as that in the *population* (3.1.1)

Note 1 to entry: The notion of representative sample is fraught with controversy, with some survey practitioners rejecting the term altogether. Reference [6] noted the following six categories of meanings of representative sampling in the non-statistical literature and attributed these to References [7], [8], [9], and [10] from their series of articles in the International Statistical Review:

1) general, unjustified acclaim, approbation for the data;

2) absence of selective factors;

3) mirror or miniature of the population. The sample has the same distributions as the population;

4) typical or ideal case;

5) coverage of the population. Samples designed to reflect variation, especially among strata;

6) probability sampling. A formal sampling scheme to give every element a known, positive probability of selection.

Note 2 to entry: This definition extends the definition given in ISO 3534-2:2006, 1.2.35 to a wider class of sampling than random sampling, to include for example *judgment sampling* (3.1.31). The starting point in ISO 3534-2:2006 with representative sampling is a *random sample* (3.1.10), whereas in this definition, the starting point is sample.

3.1.15
selection probability
number expressing the chance that a specific *sampling unit* (3.1.5) will be chosen

3.1.16
sampling
act of forming a *sample* (3.1.8)

Note 1 to entry: The general term "forming" is used since samples could arise from a random generation process, a physical process, or a scheme that has little or no stochastic basis. Subsequent statistical inference necessitates a random method for generating the sample, but sampling itself includes historically some methods that have practical deficiencies.

Note 2 to entry: Sampling method is occasionally used as a synonym for sampling, although this practice is not universal. Sampling method is also linked to *sampling plan* (3.1.24).

Note 3 to entry: This definition does not use "drawing" (in contrast to ISO 3534-2:2006, 1.3.1) as this suggests a physical process of forming the sample that is not a necessary requirement.

3.1.17
sampling with replacement
sampling (3.1.16) in which each *sampling unit* (3.1.5) is taken and observed, returned to the *population* (3.1.1) before the next sampling unit is sampled

[SOURCE: ISO 3534-2:2006, 1.3.15]

Note 1 to entry: In this case, the same sampling unit may appear more than once in the *sample* (3.1.8). It is possible that groups of units are sampled and then returned to the population before additional sampling occurs. Of course, the actual procedure used has a bearing on the associated probabilities of selection.

Note 2 to entry: In sampling with replacement, the *selection probabilities* (3.1.15) for the sampling units remains unchanged for each selection and the selections are independent.

3.1.18
sampling without replacement
sampling (3.1.16) in which each *sampling unit* (3.1.5) is included in the *sample* (3.1.8) from the *population* (3.1.1) once only

Note 1 to entry: In *finite population* (3.1.2) sampling (especially for small population sizes), the generally preferred procedure is to use sampling without replacement since more precise estimators can be obtained. However, these gains are at the expense of possibly complicated formulae for estimated variances.

Note 2 to entry: In sampling without replacement, the *selection probabilities* (3.1.15) for the sampling units change from one selection to the next (depending on the outcomes) and, consequently, the selections are not independent.

Note 3 to entry: This definition is similar but not identical to the definition given in ISO 3534-2:2006, 1.3.16.

3.1.19
census
examination of each *sampling unit* (3.1.5) in a specified *finite population* (3.1.2)

Note 1 to entry: Many countries conduct a full headcount on a regular basis (for example, every 10 years). These efforts are challenging since, for example, some parts of the population can be illusive (e.g. illegal immigrants) and are difficult to include in the census. A census can also be conducted exhaustively for finite populations to eliminate *sampling error* (3.2.9) at the cost of additional effort over that incurred in *sampling* (3.1.16). A census may not be appropriate for countries that maintain a complete registry of its population (e.g. Denmark).

Note 2 to entry: Generally, a census is conducted with respect to a fixed timepoint and with respect to specified characteristics. For example, the decennial census in the United States is conducted to characterize the population as of 1 April of the year of the census.

Note 3 to entry: The intent of a census is to examine every sampling unit in the finite population, but it is recognized that practical difficulties may occur in reaching each sampling unit and further in obtaining complete information on the individual units.

3.1.20
sample survey
examination or analytic study of a *finite population* (3.1.2) using *survey sampling* (3.1.21)

Note 1 to entry: Sample survey comprises a vast array of tools and techniques for investigating the properties and nature of a *population* (3.1.1). Questionnaires, interviews, mail surveys, and so forth come to mind as instruments for collecting information on a population. The methodology of sample survey involves the careful selection of samples in order to maximize the amount of information gleaned on the population relative to the effort expended.

Note 2 to entry: The objective of a sample survey is to gain extensive information and knowledge about the population. The results of a sample survey can be of much higher quality than that which could be obtained from a *census* (3.1.19) of the population, owing to the focusing of efforts and expertise on a subset of the population. In the case of a census, the possibly overwhelming nature of exhaustively examining every item in the population could lead to many *non-sampling errors* (3.2.10) that undermine the entire effort.