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.

3.1.21
survey sampling
process of selecting a *sample* (3.1.8) of *sampling units* (3.1.5) from a target *population* (3.1.1) to obtain information regarding the characteristics of the items in the population

Note 1 to entry: Survey sampling in every day usage often implicitly suggests the methodology and can refer to the general field of expertise associated with investigating *finite populations* (3.1.2) via questionnaires, *opinion surveys* (3.1.23), political polls, and customer satisfaction surveys.

Note 2 to entry: This definition differs from the definition (sampling used in enumerative or analytic studies to estimate the values of one or more characteristics in a population or for estimating how those characteristics are distributed across the population) given in ISO 3534-2:2006, 1.2.18, which emphasizes the objective of survey sampling rather than the process itself as is given here.

3.1.22
pilot survey
preliminary small-scale *sample survey* (3.1.20)

Note 1 to entry: Prior to conducting the complete sample survey, a small-scale study is recommended to determine if there are difficulties in the instrument, questionnaire, or the *sampling design* (3.1.28). For example, in a questionnaire, it could be determined that a query is ambiguous so that the results obtained on that question will eventually be useless. There could also be key omissions or it could be determined that the instrument itself could be improved.

3.1.23
opinion survey
sample survey (3.1.20) with the objective of investigating public views

Note 1 to entry: Typically, an opinion survey is conducted by a written questionnaire, in-person interviews, telephone interviews, or by electronic media. Consequently, the results of such investigations can reflect subjective opinions or perspectives. Two common types of opinion surveys are political surveys (or polls) and marketing surveys.

3.1.24
sampling plan
sampling design (3.1.28) including implementation details

Note 1 to entry: Sampling plan and sampling design are distinguished here in the same sense as ISO 3534-3:2013 distinguished between experimental plan and experimental design. The sampling design is restricted to the selection of *sampling units* (3.1.5) according to a possibly complex probabilistic scheme, which then implies the manner of *estimation* (3.2.4) to be conducted. Sampling plan includes the sampling design and details on the complete implementation of the sampling process, including work assignments of personnel, preparation of questionnaires, and environmental conditions in the field.

3.1.25
sampling frame
complete collection of all *sampling units* (3.1.5) with their identification

Note 1 to entry: A sampling frame provides an explicit representation of the *population* (3.1.1) of interest for the study, generally consisting of a list. It is possible that the list may be incomplete or contain some errors. For example, an electoral register could contain individuals who have moved out of or into the voting district, contain individuals who are deceased since the construction of the sampling frame, could contain duplicate entries, or even contain convicted felons who are legally not entitled to vote and should be expunged from the records.

Note 2 to entry: This definition is different from ISO 3534-2:2006, 1.2.27 to emphasize the practical limitation in constructing the sampling frame which may be an incomplete representation of the target population.

EXAMPLE Depending on the background information on the population, the sampling frame can be simple but explicit or, in the extreme case, highly complex. In a famous survey, Reference [11] conducted a survey in eastern India in which the sampling frame comprised an enumeration of the fields, a list of villages, and a set of maps in different areas. From this heterogeneous material, a sampling frame was constructed and the study then proceeded.

3.1.26
dual frame
sampling frame (3.1.25) provided from two sources

EXAMPLE A sampling frame could include telephone numbers selected from both landlines and cell phones. It is not apparent in advance given a phone number as to which receiver it belongs, yet this information could prove valuable at the analysis stage. There is also a potential problem in selecting the same individual twice (once from each source).

3.1.27
area frame
sampling frame (3.1.25) consisting of non-overlapping geographic regions

Note 1 to entry: With the increasing availability of geographic information systems and portable global positioning systems, the capability to undertake *sampling* (3.1.16) projects with area frames has been improved. A practical motivation underlying area sampling is that, for many problems, there may be no current and accurate list of *population* (3.1.1) elements.

EXAMPLE If one wanted to measure candy sales in retail stores, one might choose a *sample* (3.1.8) of city blocks, and then audit sales of all retail outlets on those sample blocks.

3.1.28
sampling design
complete description of the structure of the *sampling* (3.1.16) and the subsequent analysis

Note 1 to entry: The description refers to the type of sampling to be undertaken [*simple random sampling* (3.1.12), *stratified sampling* (3.1.32), *cluster sampling* (3.1.38), and so forth] and the *sample size* (3.1.9) or sizes if various groupings of the *sampling units* (3.1.5) are considered. The description can also include the *selection probabilities* (3.1.15) for each sampling unit and possibly the probability of selection of each potential *sample* (3.1.8).

3.1.29
systematic sampling
sampling (3.1.16) according to a partially deterministic structural plan

Note 1 to entry: With systematic sampling, the randomization of sampling is restricted. An initial starting point in the *sampling frame* (3.1.25) could be randomly selected and then every k-th item thereafter to constitute the *sample* (3.1.8). In systematic sampling, there tends to be a fixed interval in time or space between items selected. If the sampling frame possesses a cyclic pattern of a comparable length to the sampling interval, then inadvertent biases or excessive variance estimates at the analysis phase could be introduced.

Note 2 to entry: This definition varies slightly from ISO 3534-2:2006, 1.3.12 by emphasizing that systematic sampling is not entirely random.

3.1.30
quasi-random sampling
systematic sampling (3.1.29) of every n-th entry from the *sampling frame* (3.1.25)

Note 1 to entry: Under certain conditions, largely governed by the method of compiling the sampling frame or list, a systematic sample of every n-th entry from a list will be equivalent for most practical purposes to a *random sample* (3.1.10). This method of *sampling* (3.1.16) is sometimes referred to as quasi-random sampling. It should be one type of systematic sampling, and it should be below the definition for “systematic sampling” in the concept diagram.

EXAMPLE Suppose a supermarket wants to study the buying habits of their customers. Using quasi-random sampling, the supermarket manager can choose every 15th customer entering the supermarket and conduct the study on this *sample* (3.1.8).

3.1.31
judgment sampling
purposive sampling
sampling (3.1.16) constituted at the discretion of the surveyer

Note 1 to entry: Judgment sampling is not *random sampling* (3.1.11) as some *sampling units* (3.1.5) have zero probability of selection while others are chosen whimsically. Even experts in the field are challenged to produce a viable judgment sample in comparison to those constituted through sound statistically based sampling methods.

Note 2 to entry: *Quota sampling* (3.1.45) is an example of judgment sampling in which, for example, interviewers choose their respondents to contain a specific number of men and a specific number of women in approximate age classes, the result of which provides the appropriate demographics at the expense of generalizing the results to any larger group.

Note 3 to entry: In exceptional circumstances, judgment sampling can be used effectively in conjunction with other sampling methods, if, for example, there is a specific subset of the *population* (3.1.1) (say senior management) for whom *sampling errors* (3.2.9) are not to be made from a public relations standpoint. Hence, all senior managers would be included in the judgment sample, while the remaining population is assessed with more appropriate sampling methods.

Note 4 to entry: In judgment sampling, sampling units are selected by considering the available auxiliary information (often subjectively) with a view to ensuring a *sample* (3.1.8) that is an adequate reflection of the population.

Note 5 to entry: Each of the terms *systematic sampling* (3.1.29), *quasi-random sampling* (3.1.30), and judgment sampling can be viewed as a form of controlled sampling, as described by Reference [5] and Reference [3] (Section 5A.6).

Note 6 to entry: In contrast to *probability sampling* (3.1.13), uncertainties cannot be evaluated for judgment sampling.

3.1.32
stratified sampling
sampling (3.1.16) such that the portions of the *sample* (3.1.8) are selected independently from the different *strata* (3.1.7) and that at least one *sampling unit* (3.1.5) is selected from each stratum

Note 1 to entry: In some cases, the portions are specified proportions determined in advance to increase the accuracy in estimating *population* (3.1.1) characteristics. In other cases, the strata cannot be established until after the sampling takes place owing to the absence of information on the proportions. This occurs when the frame does not contain information related to the basis of the stratification.

Note 2 to entry: Items from within each stratum are often selected by *random sampling* (3.1.11).

Note 3 to entry: If stratified sampling is used, then the corresponding *estimators* (3.2.3) of the *population parameters* (3.2.1) should take this into account.

Note 4 to entry: This definition differs slightly from ISO 3534-2:2006, 1.3.6 since it is necessary to select from each sample independently.

3.1.33
stratified simple random sampling
simple random sampling (3.1.12) from each *stratum* (3.1.7)

[SOURCE: ISO 3534-2:2006, 1.3.7]

Note 1 to entry: If the proportions of items selected from the differing strata are equal to the proportions of *population* (3.1.1) items in the strata, it is called proportional stratified simple random sampling.

3.1.34
proportional allocation
procedure in *stratified sampling* (3.1.32) to allocate the number of *sampling units* (3.1.5) to different *strata* (3.1.7) proportional to the number of sampling units in the strata

Note 1 to entry: An advantage of the use of proportional allocation is that the *estimator* (3.2.3) of the *population* (3.1.1) total does not require the use of strata weights. In other words, the contributions to the population total estimate, for example, is obtained by summing the observed values across all strata and then accounting for the sample size and population size. This is an example of a self-weighting estimate.

3.1.35
optimum allocation
procedure in *stratified sampling* (3.1.32) to allot the number of *sampling units* (3.1.5) to different *strata* (3.1.7) to optimize an objective function

Note 1 to entry: A variety of objective functions can be considered including those involving costs and others involving the precision in estimation. Optimize can mean minimize or maximize depending on the situation (e.g. minimize cost, maximize precision, maximize the number of samples for a fixed total cost, or minimize the variance of an estimator).

Note 2 to entry: Particularly in cases with repeated sampling (say on an annual basis), it could be prudent to compare the sample standard deviations within the strata to the assumed standard deviations used in obtaining the optimum allocation.

EXAMPLE Suppose the cost, C, of conducting a survey is

$$C = c_0 + \sum_h c_h n_h$$

where

c_0 is an overhead cost;

c_h is the unit cost for stratum h;

n_h is the sample size for stratum h (to be determined).

To meet a specified total cost C, then the optimum total sample size n summed across the strata is given by

$$n = \frac{(C - c_0)\sum N_h \sigma_h / \sqrt{c_h}}{\sum N_h \sigma_h \sqrt{c_h}}$$

The allocation in each stratum is given by

$$n_h = n \frac{N_h \sigma_h / \sqrt{c_h}}{\sum N_h \sigma_h / \sqrt{c_h}}$$

Reference [3] provides the optimal sample size if the goal is to meet a specified variance of the population mean.

3.1.36
Neyman allocation
optimum allocation (3.1.35) with the objective function equal to the variance of the *estimator* (3.2.3) of the population mean or total

Note 1 to entry: The purpose of the method is to maximize survey precision, given a fixed *sample size* (3.1.9). Neyman allocation, as the definition indicates, is actually a special case of optimal allocation (i.e. it minimizes the variance of the estimator of the population mean).

EXAMPLE With Neyman allocation, the optimum sample size for *stratum* (3.1.7)*h* would be:

$$n_h = n \frac{N_h \sigma_h}{(N_1 \sigma_1 + ... + N_k \sigma_k)}$$

where

k is the number of strata;

n_h is the sample size for stratum h;

n is the total sample size;

N_h is the *population* (3.1.1) size for stratum h;

σ_h is the standard deviation of stratum h.

More *sampling units* (3.1.5) are selected from strata exhibiting greater variability or which are relatively large.

3.1.37
poststratification
procedure in *stratified sampling* (3.1.32) to allocate the *sampling units* (3.1.5) to the different strata following the selection of a *random sample* (3.1.10)

Note 1 to entry: A possible application of poststratification occurs in conjunction with *simple random sampling* (3.1.12), in which the *stratum* (3.1.7) for each item in the *sample* (3.1.8) is identified after selection. If the strata sizes are known, then appropriate weights can be applied to the strata means. If the strata sizes are not known in advance, then the weights can be estimated according to the proportions observed from the simple random sample. Such a procedure, as suggested in this note, can be viewed as analogous to *proportional allocation* (3.1.34) using the observed proportions. For large samples and for cases where the errors in the estimated weights are minor, poststratification can be almost as precise as the results that would be obtained from proportional allocation.

3.1.38
cluster sampling
sampling (3.1.16) in which a *random sample* (3.1.10) of *clusters* (3.1.6) is selected and all the *sampling units* (3.1.5) which constitute the clusters are included in the *sample* (3.1.8)

[SOURCE: ISO 3534-2:2006, 1.3.9]

Note 1 to entry: The clusters are considered as the primary sampling units, as initially discussed in 3.1.5.

3.1.39
post cluster sampling
sampling (3.1.16) performed on *clusters* (3.1.6) from the initial *random sample* (3.1.10)

Note 1 to entry: The main difficulty faced in *cluster sampling* (3.1.38) is the lack of information relating to composition of clusters. In such situations, the clusters are built on the basis of initial random sample, then a final sampling is performed with these clusters as *sampling units* (3.1.5).

Note 2 to entry: Thus the name "post cluster" which essentially means that clusters are formed afterwards and not known beforehand. The term was coined by T. Dalenius, who first introduced the idea in his book *Sampling in Sweden*, pp. 156–158.

3.1.40
multi-stage sampling
sampling (3.1.16) in which the *sample* (3.1.8) is selected by stages, the *sampling units* (3.1.5) at each stage selected within the larger units from the preceding stage

[SOURCE: ISO 3534-2:2006, 1.3.16, modified — Used the word "selected" rather than "sampled" to avoid a circular definition.]

Note 1 to entry: Multi-stage sampling is different from multiple sampling. Multiple sampling is sampling by several criteria at the same time.

Note 2 to entry: The sampling method can be different for the various stages, such that the primary sample can be selected by *simple random sampling* (3.1.12) while the final sample is obtained through *systematic sampling* (3.1.29).

EXAMPLE In a multi-stage application, the first stage primary units could be provinces within a country, the second stage could be municipalities within provinces, and the third stage could be precincts within municipalities.

3.1.41
two-stage sampling
multi-stage sampling (3.1.40) with two stages

Note 1 to entry: After an initial sampling on *n sampling units* (3.1.5), another sampling is performed on the initial group of sampling units selected during the first sampling process.

EXAMPLE One could first generate a *sample* (3.1.8) of countries and then select a sample of provinces (alternatively, states, departments, or districts).

3.1.42
multi-stage cluster sampling
cluster sampling (3.1.38) with two or more stages, each *sampling* (3.1.16) being made on *clusters* (3.1.6), in which the clusters already obtained by the preceding *sample* (3.1.8) have been divided

[SOURCE: ISO 3534-2:2006, 1.3.11]

3.1.43
inverse sampling
sampling (3.1.16) that continues until pre-specified conditions have been met based upon the results so far obtained

Note 1 to entry: Inverse sampling is often used in surveys for rare items, e.g. rare diseases, in order to get a sufficient number of items in the *sample* (3.1.8) with the total number of items to be sampled being a random quantity.

EXAMPLE *Random sampling* (3.1.11) continues at a hospital until 10 instances of newborns having a rare congenital disorder occur. After the tenth such individual is identified, the incidence rate of this disorder can be estimated.

3.1.44
sampling proportional to size
sampling (3.1.16) such that the probability of selecting a *sampling unit* (3.1.5) is proportional to the size of an auxiliary variable

Note 1 to entry: Sampling proportional to size is most useful when the sampling units vary considerably in size because it ensures that the larger sampling units have a larger probability of getting into the *sample* (3.1.8). This method also facilitates planning for fieldwork because a predetermined number of respondents are interviewed in each unit selected, and staff can be allocated accordingly.

Note 2 to entry: If sampling proportional to size is used, then the corresponding *estimators* (3.2.3) of the *population parameters* (3.2.1) should take this into account. It can also be prudent to verify from the data that the auxiliary variable which is the basis for the selection probabilities is indeed correlated with the population characteristic.

Note 3 to entry: Sampling proportional to size is sometimes designated as “PPS sampling” (i.e. probability proportional to size). PPS sampling is particularly useful if the auxiliary variable is highly correlated with the population characteristic of interest.

3.1.45
quota sampling
stratified sampling (3.1.32) where the *sample* (3.1.8) is selected in a non-random manner that conforms to predefined structures

Note 1 to entry: Quota sampling is problematic since the selection of *sampling units* (3.1.5) being non-random precludes the use of standard inferential procedures. A particular concern with quota sampling is the evident *bias* (3.2.7) in cases, for example, where the interviewer selects sampling units with a particular conclusion in mind. In light of the non-randomness of the sampling, the usual methods of uncertainty estimation are not applicable.

Note 2 to entry: This definition is taken from ISO 3534-2:2006, 1.3.8 with the addition of the phrase “that conforms to predefined structures” to reflect the use of quota sampling in *survey sampling* (3.1.21).

3.2 Terms related to estimation

3.2.1
population parameter
unknown quantity corresponding to a characteristic of a *population* (3.1.1)

Note 1 to entry: The objectives of *sample surveys* (3.1.20) are typically to *estimate* (3.2.2) certain population parameters. Population parameters are usually symbolized by lower case Greek letters in italics.

Note 2 to entry: The definition given here differs from ISO 3534-2:2006, 1.2.2. The value of a population parameter could be known if a *census* (3.1.19) took place and each characteristic for each item in the population were then obtained without error. Prior to conducting the census, the population parameter would not be known.

EXAMPLE Examples of population parameters and the typical symbol used are population total (τ), population mean (μ), and population standard deviation (σ).

3.2.2
estimate
observed value of an *estimator* (3.2.3)

[SOURCE: ISO 3534-1:2006, 1.31]

Note 1 to entry: Estimate refers to a numerical value obtained from observed values. With respect to estimation of a *population parameter* (3.2.1), estimator refers to the statistic intended to estimate the population parameter and estimate refers to the result using observed values. Sometimes the adjective “point” is inserted before estimate to emphasize that a single value is being produced rather than an interval of values. Similarly, the adjective “interval” is inserted before estimate in cases where interval *estimation* (3.2.4) is taking place.

3.2.3
estimator
statistic used in estimation of the *population parameter* (3.2.1) based on a *sample* (3.1.8)

Note 1 to entry: An estimator takes into account the *sampling* (3.1.16) method and auxiliary information as available and pertinent. An estimator could be the sample average intended to *estimate* (3.2.2) the *population* (3.1.1) average which is appropriate for *simple random sampling* (3.1.12). Another estimator in the context of a *probability sample* (3.1.13) could also be intended to estimate the population average although the formula for this estimator would incorporate the probabilities of each item being selected in the sample.

Note 2 to entry: This definition differs slightly from ISO 3534-1:2006, 1.12, to place it in the context of sampling.

EXAMPLE The population total is commonly of interest. An estimator of the population total could be the average value in the sample multiplied by the total number of items in the population, which would be appropriate with a simple random sample.

3.2.4
estimation
process of obtaining an *estimate* (3.2.2) from an *estimator* (3.2.3)

Note 1 to entry: Estimation in ISO 3534-1:2006, 1.36 is understood in a rather broad context to include point estimation, interval estimation, or estimation of properties of *populations* (3.1.1). For the definition given here, the emphasis is on point estimation.

3.2.5
standard error
standard deviation of an *estimator* (3.2.3)

[SOURCE: ISO 3534-1:2006, 1.24]

Note 1 to entry: There is no (sensible) complementary term "non-standard" error. Standard error can be viewed as an abbreviation for the expression standard deviation of an estimator.

3.2.6
error of estimation
estimate (3.2.2) minus the true value of the *population parameter* (3.2.1) or *population* (3.1.1) property that it is intended to estimate

[SOURCE: ISO 3534-1:2006, 1.32 — modified as per Note 3 to entry below.]

Note 1 to entry: The population characteristic could be the population total, population average, or proportion of items having a specified property.

Note 2 to entry: Error of estimation could involve contributions due to *sampling* (3.1.16), measurement uncertainty, rounding, or other sources. In effect, error of estimation represents the bottom line performance of interest to practitioners. Determining the primary contributors to the error of estimation is a critical element in assessing the quality of a sampling activity.

Note 3 to entry: The definition given here differs slightly from that given in ISO 3534-1:2006, 1.32 in that it uses the terms "value of the population parameter" rather than merely "parameter", owing to the context of *survey sampling* (3.1.21).

3.2.7
bias
expectation of *error of estimation* (3.2.6)

[SOURCE: ISO 3534-1:2006, 1.33]

Note 1 to entry: The existence of bias can lead to unfortunate consequences in practice. For example, bias could lead to incorrect decisions from a political poll. On the other hand, in *survey sampling* (3.1.21), there are some *estimators* (3.2.3) (e.g. ratio, regression, and product estimators) that are slightly biased but may have small variability compared to the best unbiased estimator.

Note 2 to entry: An estimator of a *population parameter* (3.2.1) has zero bias if its average value over all possible samples equals this parameter. Some sampling textbooks refer to this property as design unbiased, although this usage is not universal.

Note 3 to entry: The term "expectation" is defined in ISO 3534-1:2006, 2.12.

3.2.8
unbiased estimator
estimator (3.2.3) having *bias* (3.2.7) equal to zero

[SOURCE: ISO 3534-1:2006, 1.34]

Note 1 to entry: Estimators that are unbiased are desirable in that on average, they give the correct value. Certainly, unbiased estimators provide a useful starting point in the search for "optimal" estimators of *population* (3.1.1) characteristics. The definition given here is of a statistical nature. In every day usage, practitioners try to avoid introducing bias into a study by ensuring, for example, that the *random sample* (3.1.10) is suitable for doing *estimation* (3.2.4) for the entire population. However, in many *survey sampling* (3.1.21) applications, unbiased estimators may not be readily available and in fact, certain estimators, although somewhat biased, may yield highly precise *estimates* (3.2.2) of *population parameters* (3.2.1).

3.2.9
sampling error
error of estimation (3.2.6) attributable to the process of *sampling* (3.1.16)

Note 1 to entry: Since not all items in the *population* (3.1.1) are considered, the summary value from the *sample* (3.1.8) will likely vary from sample to sample and may differ from the population value [if it were obtained from a *census* (3.1.19)]. Sampling error refers to that part of the difference between the *population parameter* (3.2.1) and its *estimate* (3.2.2) due to consideration of a sample only. It is recognized that most values of the characteristic in the sample will differ somewhat from the population parameter, which could only be determined through exhaustive consideration of all items in the population.

3.2.10
non-sampling error
error of estimation (3.2.6) attributable to all causes other than through the process of *sampling* (3.1.16)

Note 1 to entry: *Sample surveys* (3.1.20) can be fraught with difficulties in implementation. Basic mistakes can be made from inaccurate recording of subject responses to responses driven by a misunderstanding of a query leading to non-response. Various methods have been developed to overcome non-sampling errors. Sources of non-sampling error include deficiencies in the *samplingframe* (3.1.25), dishonesty in respondents, negligence on the part of the interviewers or survey takers, among many possibilities. Additionally, cultural biases, interview bias, and ambiguous questions pose further difficulties.

Note 2 to entry: Non-sampling errors could be human errors, possibly occurring during the statistical analysis. These errors can include, but are not limited to, data entry errors, biased questions in a questionnaire, biased processing/decision-making, inappropriate analysis, incorrect conclusions, and false information provided by respondents. The presence of these errors are likely to make the estimate less precise.

Note 3 to entry: Non-sampling errors are part of the total error that can arise from doing a statistical analysis. The remainder of the total error arises from *sampling error* (3.2.9). Unlike sampling error, increasing the *sample size* (3.1.9) will not have any effect on reducing non-sampling error. In fact, non-sampling errors can become more apparent with an increase in sample size (as the sampling error is reduced with increasing sample size). Unfortunately, it is virtually impossible to eliminate non-sampling errors entirely.

Note 4 to entry: Non-sampling errors occur from non-response, coding errors, computer processing errors, errors in the sampling frame, reporting errors, and other errors. Non-sampling errors are reduced through data editing, statistical adjustments for non-response, and close attention to detail. Non-sampling errors arise during the course of almost all survey activities [even a complete *census* (3.1.19)], such as respondents' different interpretation of questions, mistakes in processing results, or errors in the sampling frame. The non-response problem can result in biased *estimates* (3.2.2).

Note 5 to entry: Many techniques in *survey sampling* (3.1.21) (design or analysis stage) have been developed to remedy or to mitigate the non-sampling errors noted previously.

3.2.11
non-response
lack of an observed value for a *sampling unit* (3.1.5) included in the *sample* (3.1.8)

Note 1 to entry: Non-response represents missing values which can undermine the *sampling design* (3.1.28). Some missing values can be eventually replaced with an observation through persistence of the investigators. In spite of the best efforts of the investigators, non-response represents a potential source of *bias* (3.2.7) or *non-sampling error* (3.2.10). The observations could be missing due to difficulties in implementing the *sampling plan* (3.1.24) or due to refusals of some respondents (sampling units) to provide an observed value, possibly because they were not home at the time of the survey.

Note 2 to entry: The treatment of non-response is important to avoid reporting badly biased results. Methodology in *survey sampling* (3.1.21) has been developed to handle non-response. From a logistical standpoint, callbacks or repeated attempts to contact are used. For example, failure to return a questionnaire could lead to an in-person interview. Ultimately, having exhausted such methods at collecting the data, methods of imputation could be considered (zero imputation, mean value imputation, and regression imputation). The method of weighing adjustment can also be used to address the problem of non-response.

Note 3 to entry: Non-response can be particularly problematic in circumstances where specific groups of sampling units [possibly accounted for via *strata* (3.1.7)] yield very different response rates. For example, telephone surveys are likely to experience different response rates for residences depending on the employment situation.

EXAMPLE Situations that cause non-response include but are not limited to lack of coverage, not at home, and unable to answer. Lack of coverage occurs, for example, in agricultural settings where the production from homeowner gardens or farmer's market suppliers is not included.

3.2.12
question bias
bias (3.2.7) attributable to the nature of a query in the survey instrument

Note 1 to entry: The survey instrument could be a questionnaire or a script used by an interviewer.

EXAMPLE Question bias is introduced if the wording of a question is not appropriate and is misunderstood by the respondents or by certain subsets of the respondents. Question bias also occurs if the wording of a question is suggestive of an answer that the surveyer has a pre-disposition to obtain. Question bias is not the case in situations where the respondent is unwilling or reluctant to provide a truthful answer to an otherwise appropriate question (e.g. religious affiliation, sexual orientation, illicit drug use, and so forth). In such cases, randomized response procedures could be used.[16]

3.2.13
observational error
difference between a measured value of a quantity and its true value

Note 1 to entry: In statistics, an error is not a "mistake". Variability is an inherent part of things being measured and of the measurement process.

3.2.14
regression estimator
estimator (3.2.3) adjusted by a linear function of one or more auxiliary variables

Note 1 to entry: Regression estimators are designed to take advantage of auxiliary information available on the *sampling units* (3.1.5) and are strongly correlated with the population characteristic of interest. As the *auxiliary data* (3.2.15) are generally available for the entire *population* (3.1.1), *estimators* (3.2.3) can be improved by using this information.

EXAMPLE The study variable can be production in an agricultural field, while the auxiliary information in this case can be the number of plants in the agricultural field.

3.2.15
auxiliary data
additional information on *sampling units* (3.1.5) beyond the primary response variable

Note 1 to entry: The auxiliary data provides observed values for one or more auxiliary variables that could prove useful in the analysis stage in estimating population characteristics. In particular, *regression estimators* (3.2.14) and ratio *estimators* (3.2.16) rely on auxiliary data available on the entire *population* (3.1.1) coupled with the results from the *sample* (3.1.8).

Note 2 to entry: Auxiliary data can be used in the context of *sampling proportional to size* (3.1.44) where the size relates to an auxiliary variable.

Note 3 to entry: Yet another use of auxiliary data is to estimate the size of various *strata* (3.1.7) prior to conducting a *stratified sampling* (3.1.32) project.

EXAMPLE Other examples in this part of ISO 3534 that use auxiliary variables are included in (3.1.35), (3.1.36), (3.1.44), (3.2.14), and (3.2.16).

3.2.16
ratio estimator
estimator (3.2.3) that is one estimator divided by a second estimator, where the numerator is obtained from the *sample* (3.1.8) of observed values and the values in the denominator are available for the full *population* (3.1.1)

Note 1 to entry: Ratio estimators are commonly encountered in *finite population* (3.1.2)*sampling* (3.1.16). Ratio estimators also arise in *two-stage sampling* (3.1.41).

EXAMPLE 1 If the study variable is the number of bullocks on a farm and the auxiliary variable is the farm area, a ratio estimator can be used to estimate the mean number of bullocks per farm.

EXAMPLE 2 The study variable is the estimation of the total number of tractors in a district (net area sown is an auxiliary variable). Let there be N villages in the district out of which n villages have been randomly selected:

y_i : is the total number of tractors in the i-th randomly selected village.

x_i : is the net sown area in the i-th randomly selected village.

$R_i = y_i / x_i$: is the total number of tractors per unit net sown area in the i-th randomly selected village.

3.2.17
Horvitz-Thompson estimator
estimator (3.2.3) that weighs the observed values by the inverse of their inclusion probabilities

Note 1 to entry: The Horvitz-Thompson estimator is a general-purpose estimator which can be used for any probability *sampling plan* (3.1.24). The probability that a *sampling unit* (3.1.5) will be included in the *sample* (3.1.8) is denoted π_i. The Horvitz-Thompson estimator for the population total is given by

$$\sum_{i \in s} \frac{y_i}{\pi_i}$$

where

s is the set of indices for the sample;

y_i is the observed value for the item in the sample with index i;

π_i is the inclusion probability for this item.

This estimator is unbiased.

Note 2 to entry: In certain circumstances, the estimator of the variance of the Horvitz-Thompson estimator can be negative.[13]

Note 3 to entry: The Yates-Grundy estimator of the variance of the Horvitz-Thompson estimator under the Midzuno (1952) [12] scheme of sampling is always non-negative.

3.2.18
sampling fraction

proportion of *sampling units* (3.1.5) selected from a *population* (3.1.1), *sub-population* (3.1.3), *cluster* (3.1.6), or *stratum* (3.1.7) to the total sampling units in a population, sub-population, cluster, or stratum, respectively

Note 1 to entry: The sampling fraction in most situations will vary from one sub-population, cluster, or stratum to the next.

3.2.19
finite population correction

fpc

adjustment factor in *sampling without replacement* (3.1.18) from a *finite population* (3.1.2)

EXAMPLE In simple random sampling without replacement, the variance of the sample mean is (σ^2/n) $(1 - n/N)$. The second term is the finite population correction. As the *sampling fraction* (3.2.18) approaches 1, the finite population correction in turn approaches zero. In contrast, a correction is not warranted in *sampling with replacement* (3.1.17).

Annex A
(informative)

Methodology used to develop the vocabulary

A.1 General

The universality of application of this International Standard requires the employment of a coherent and harmonized vocabulary that is easily understandable by all potential users of applied statistics standards.

Concepts are not independent of one another, and an analysis of the relationships between concepts within the field of applied statistics and the arrangement of them into concept systems is a prerequisite of a coherent vocabulary. Such an analysis is used in the development of the vocabulary specified in this International Standard. Since the concept diagrams employed during the development process can be helpful in an informative sense, they are reproduced in Figures A.1 to A.3.

A.2 Content of a vocabulary entry and the substitution rule

The concept forms the unit of transfer between languages (including variants within one language, e.g. American English and British English). For each language, the most appropriate term for the universal transparency of the concept in that language, i.e. not a literal approach to translation, is chosen.

A definition is formed by describing only those characteristics that are essential to identify the concept. Information concerning the concept which is important but which is not essential to its description is put in one or more notes to the definition.

When a term is substituted by its definition, subject to minor syntax changes, there should be no change in the meaning of the text. Such a substitution provides a simple method for checking the accuracy of a definition. However, where the definition is complex in the sense that it contains a number of terms, substitution is best carried out taking one or, at most, two definitions at a time. Complete substitution of the totality of the terms will become difficult to achieve syntactically and unhelpful in conveying meaning.

A.3 Concept relationships and their graphical representation

A.3.1 General

In terminology work, the relationships between concepts are, as far as possible, based on the hierarchical formation of the characteristics of a species. This enables the most economical description of a concept by naming its species and describing the characteristics that distinguish it from its parent or sibling concepts. There are three primary forms of concept relationships indicated in this Annex: the hierarchical generic (A.3.2), the partitive (A.3.3), and the non-hierarchical associative (A.3.4).

A.3.2 Generic relation

Subordinate concepts within the hierarchy inherit all the characteristics of the superordinate concept and contain descriptions of these characteristics which distinguish them from the superordinate (parent) and coordinate (sibling) concepts, e.g. the relation of spring, summer, autumn, and winter to season. Generic relations are depicted by a fan or tree diagram without arrows (see Figure A.1).

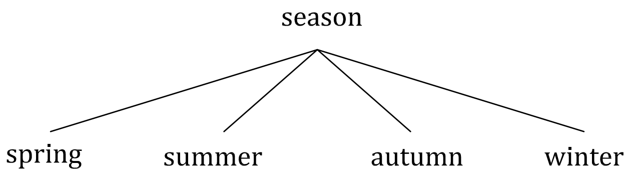

Figure A.1 — Graphical representation of a generic relation

A.3.3 Partitive relation

Subordinate concepts within the hierarchy form constituent parts of the superordinate concept, e.g. spring, summer, autumn, and winter can be defined as parts of the concept year. In comparison, it is inappropriate to define sunny weather (one possible characteristic of summer) as part of a year. Partitive relations are depicted by a rake, without arrows (see Figure A.2). Singular parts are depicted by one line, multiple parts by double lines.

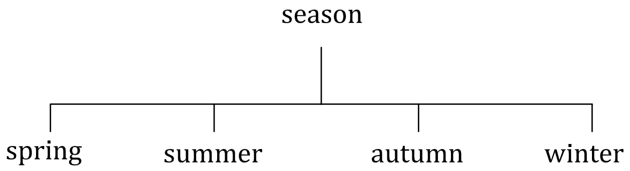

Figure A.2 — Graphical representation of a partitive relation

A.3.4 Associative relation

Associative relations cannot provide the economies in description that are present in generic and partitive relations but are helpful in identifying the nature of the relationship between one concept and another within a concept system, e.g. cause and effect, activity and location, activity and result, tool and function, material and product. An associative relation is depicted by a line with an arrowhead at each end (see Figure A.3). The exception is where sequential activities are involved. In this case, the single arrowhead is in the direction of flow.

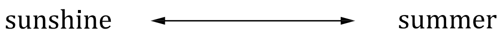

Figure A.3 — Graphical representation of an associative relation

Annex B
(informative)

Concept diagrams

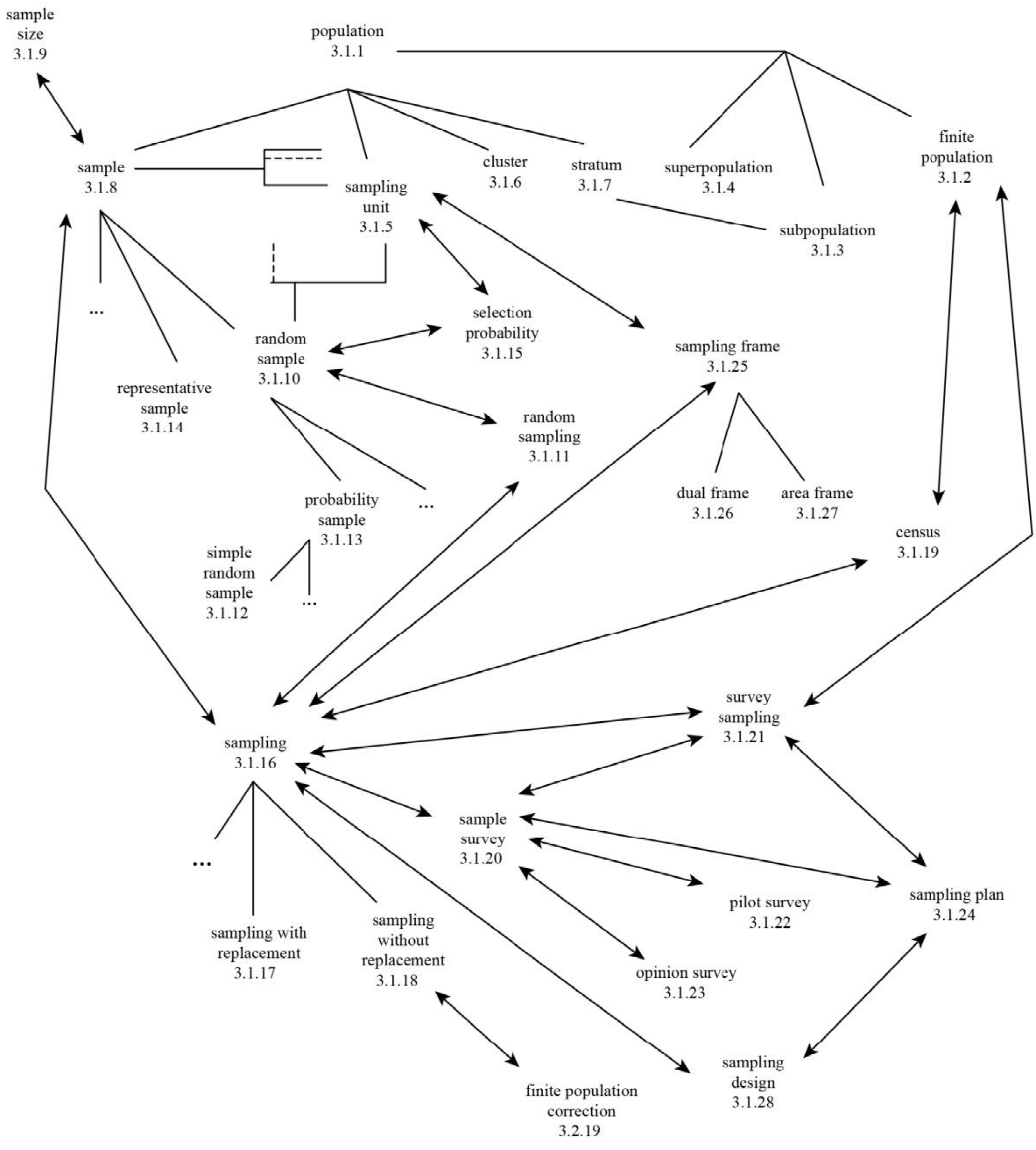

Figure B.1 — Basic sampling concepts

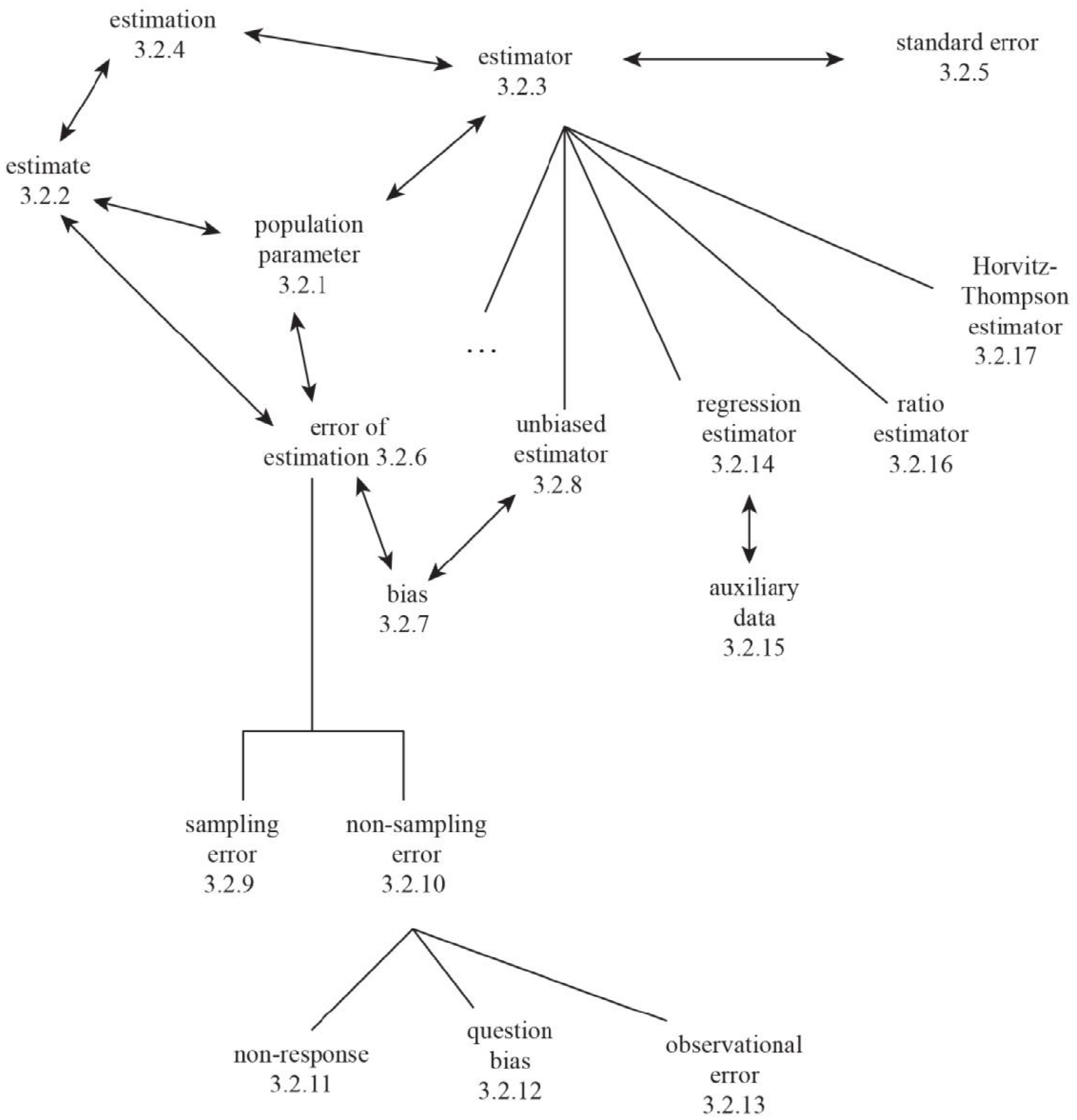

Figure B.2 — Estimation concepts

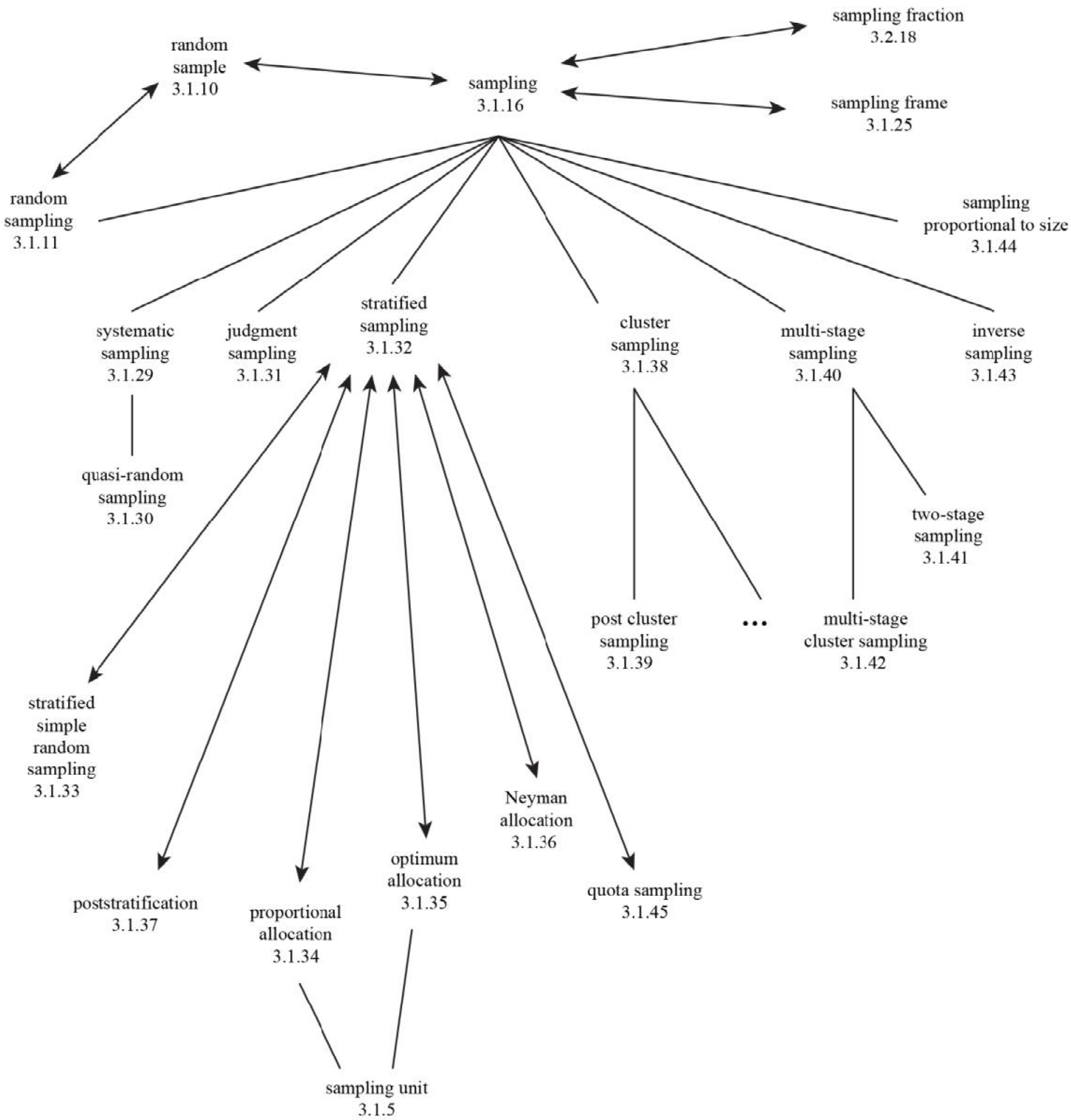

Figure B.3 — Sampling types

Annex C
(informative)

Index of sampling terms

3.1.28 sampling design

3.1.29 systematic sampling

3.1.30 quasi-random sampling

3.1.31 judgment sampling

3.1.32 stratified sampling

3.1.33 stratified simple random sampling

3.1.34 proportional allocation

3.1.35 optimum allocation

3.1.36 Neyman allocation

3.1.37 poststratification

3.1.38 cluster sampling

3.1.39 post cluster sampling

3.1.40 multi-stage sampling

3.1.41 two-stage sampling

3.1.42 multi-stage cluster sampling

3.1.43 inverse sampling

3.1.44 sampling proportional to size

3.1.45 quota sampling

Terms related to estimation

3.2.1 population parameter

3.2.2 estimate

3.2.3 estimator

3.2.4 estimation

3.2.5 standard error

3.2.6 error of estimation

3.2.7 bias

3.2.8 unbiased estimator

3.2.9 sampling error

3.2.10 non-sampling error

3.2.11 non-response

3.2.12 question bias

Annex D
(informative)

Alphabetical index of sampling terms

Bibliography

[1] Basu D. An Essay on the Logical Foundations of Survey Sampling, part one. In: *Foundations of Statistical Inference*, (Godambe V.P., & Sprott D.A. eds.). Holt, Rinehart, and Winston, Toronto, 1971

[2] Cassel C., Särndal C.-E., Wretman J.H. *Foundations of Inference in Survey Sampling*. Wiley, New York, 1977

[3] Cochran W.G. *Sampling Techniques*. Wiley, New York, Third Edition, 1977

[4] Dalenius T. *Elements of survey sampling*. Swedish Agency for Research Cooperation with Developing Countries, 1985

[5] Goodman R., & Kish L. Controlled Selection — A Technique in Probability Sampling. *J. Am. Stat. Assoc.* 1950, **45** pp. 439–448

[6] Kish L. *Questions/Answers (1978-1994) from the Survey Statistician*. International Association of Survey Statisticians, 1995

[7] Kruskal W., & Mosteller F. Representative Sampling I. Non-scientific Literature. *Int. Stat. Rev.* 1979, **47** pp. 13–24

[8] Kruskal W., & Mosteller F. Representative Sampling II. Scientific Literature, Excluding Statistics. *Int. Stat. Rev.* 1979, **47** pp. 111–122

[9] Kruskal W., & Mosteller F. Representative Sampling III. The Current Statistical Literature. *Int. Stat. Rev.* 1979, **47** pp. 245–265

[10] Kruskal W., & Mosteller F. Representative Sampling IV. The History of the Concept in Statistics. *Int. Stat. Rev.* 1979, **48** pp. 169–195

[11] Mahalanobis P.C. A Sample Survey of the Acreage Under Jute in Bengal. *Sankhya*. 1940, **4** pp. 511–530

[12] Midzuno H. On the Sampling System with Probability Proportionate to Sum of Size. *Ann. Inst. Stat. Math.* 1952, **3** pp. 99–107

[13] Raj D. *Sampling Theory*. McGraw Hill, New York, 1968

[14] Särndal C.-E., Swensson B., Wretman J. *Model Assisted Survey Sampling*. Springer, New York, 1992

[15] Sukhatme P.V., Sukhatme B.V., Sukhatme S., Asok C. *Sampling Theory of Surveys with Applications*. Iowa State University Press, Ames, 1984

[16] Warner S.L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Am. Stat. Assoc.* 1965, **60** pp. 63–69

[17] Yates F., & Grundy P.M. Selection Without Replacement from Within Strata with Probability Proportional to Size. *J. R. Stat. Soc., B*. 1953, **15** pp. 253–261

ICS 03.120.30;01.040.03

Price based on 30 pages