# INTERNATIONAL STANDARD

# ISO 24153

First edition
2009-12-01

## Random sampling and randomization procedures

*Modes opératoires d'échantillonnage et de répartition aléatoires*



Reference number
ISO 24153:2009(E)

# 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 24153 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 5, *Acceptance sampling*.



 

## Introduction

Random sampling and randomization procedures are the cornerstone to the validity of many statistical methods used in experimentation, whether for industrial quality control and improvement purposes or for designed experiments in the medical, biological, agricultural, or other scientific fields. Many statistical standards address the conduct of such experimentation. In particular, all of the following acceptance-sampling standards have been designed on the premise that random sampling is employed to select the required sampling units for lot disposition purposes:

ISO 2859 (all parts), *Sampling procedures for inspection by attributes*

ISO 3951 (all parts), *Sampling procedures for inspection by variables*

ISO 8422, *Sequential sampling plans for inspection by attributes*

ISO 8423, *Sequential sampling plans for inspection by variables for percent nonconforming (known standard deviation)*

ISO 13448 (all parts), *Acceptance sampling procedures based on the allocation of priorities principle (APP)*



ISO 14560, *Acceptance sampling procedures by attributes — Specified quality levels in nonconforming items per million*



ISO 18414, *Acceptance sampling procedures by attributes — Accept-zero sampling system based on credit principle for controlling outgoing quality*



ISO 21247, *Combined accept-zero sampling systems and process control procedures for product acceptance*

In addition, ISO 2859-3 and ISO 21247 include provisions for random sampling to be applied to determine whether a lot should be inspected or not under skip-lot sampling procedures, and to decide which units require inspection from a production process under continuous sampling plans, respectively. Consequently, it is of great importance to the valid operation of all of the above standards that sampling be effectively random in its application.

Although the principles of this International Standard are universally applicable where random sampling is required and the sampling units can be clearly defined, preferably on the basis of discrete items, there are many situations in which the material of interest does not lend itself to being quantified on a discrete-item basis, as in the case of a bulk material. In such situations, the user is advised to consult the following ISO International Standards for appropriate guidance:

ISO 11648 (all parts), *Statistical aspects of sampling from bulk materials*

# Random sampling and randomization procedures

## 1 Scope

This International Standard defines procedures for random sampling and randomization. Several methods are provided, including approaches based on mechanical devices, tables of random numbers, and portable computer algorithms.

This International Standard is applicable whenever a regulation, contract, or other standard requires random sampling or randomization to be used. The methods are applicable to such situations as

a) acceptance sampling of discrete units presented for inspection in lots,

b) sampling for survey purposes,

c) auditing of quality management system results, and

d) selecting experimental units, allocating treatments to them, and determining evaluation order in the conduct of designed experiments.



Information is also included to facilitate auditing or other external review of random sampling or randomization results where this is required by quality management personnel or regulatory bodies.



This International Standard does not provide guidance as to the appropriate random sampling or randomization procedures to be used for any particular experimental situation or give guidance with respect to possible sampling strategy selection or sample size determination. Other ISO standards (such as those listed in the Introduction) or authoritative references should be consulted for guidance in such areas.

## 2 Normative references

The following referenced documents are indispensable for the application of this document. 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, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

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

ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

ISO 80000-2, *Quantities and units — Part 2: Mathematical signs and symbols to be used in the natural sciences and technology*

## 3 Terms, definitions, and symbols

For the purposes of this document, the terms and definitions given in ISO 3534-1, ISO 3534-2, ISO 3534-3, and the following apply.

 

## 3.1 Terms and definitions

**3.1.1**
**cluster**
part of a **population** (3.1.6) divided into mutually exclusive groups of **sampling units** (3.1.13) related in a certain manner

[ISO 3534-2:2006, definition 1.2.28]

**3.1.2**
**cluster sampling**
**sampling** (3.1.12) in which a **random sample** (3.1.8) of **clusters** (3.1.1) is selected and all the **sampling units** (3.1.13) which constitute the clusters are included in the **sample** (3.1.11)

[ISO 3534-2:2006, definition 1.3.9]

**3.1.3**
**derangement**
**complete permutation**
permutation of elements where no element remains in its original position in the set (e.g. {3, 1, 2} is a derangement of {1, 2, 3})

**3.1.4**
**lot**
definite part of a **population** (3.1.6) constituted under essentially the same conditions as the population with respect to the **sampling** (3.1.12) purpose



NOTE The sampling purpose can, for example, be to determine lot acceptability, or to estimate the mean value of a particular characteristic.

[ISO 3534-2:2006, definition 1.2.4]



**3.1.5**
**multistage sampling**
**sampling** (3.1.12) in which the **sample** (3.1.11) is selected by stages, the **sampling units** (3.1.13) at each stage being sampled from the larger sampling units chosen at the previous stage

NOTE Multistage sampling is different from multiple sampling. Multiple sampling is sampling by several criteria at the same time.

[ISO 3534-2:2006, definition 1.3.10]

**3.1.6**
**population**
⟨reference⟩ totality of items under consideration

[ISO 3534-2:2006, definition 1.2.1]

**3.1.7**
**pseudo-independent random sampling**
**sampling** (3.1.12) where a **sample** (3.1.11) of $n$ **sampling units** (3.1.13) is taken from a **population** (3.1.6) in accordance with a table of random numbers or a computer algorithm designed such that each of the possible combinations of $n$ sampling units has a particular probability of being taken (see also 4.4)

**3.1.8**
**random sample**
**sample** (3.1.11) selected by **random sampling** (3.1.9)

[ISO 3534-2:2006, definition 1.2.25]

 

**3.1.9**
**random sampling**
**sampling** (3.1.12) where a **sample** (3.1.11) of $n$ **sampling units** (3.1.13) is taken from a **population** (3.1.6) in such a way that each of the possible combinations of $n$ sampling units has a particular probability of being taken

[ISO 3534-2:2006, definition 1.3.5]

**3.1.10**
**randomization**
process by which a set of items are set into a random order

NOTE If, from a **population** (3.1.6) consisting of the natural numbers 1 to $n$, numbers are drawn at random (i.e. in such a way that all numbers have the same chance of being drawn), one by one, successively, without replacement, until the population is exhausted, the numbers are said to be drawn "in random order".

If these $n$ numbers have been associated in advance with $n$ distinct units or $n$ distinct treatments that are then re-arranged in the order in which the numbers are drawn, the order of the units or treatments is said to be randomized.

**3.1.11**
**sample**
subset of a **population** (3.1.6) made up of one or more **sampling units** (3.1.13)

[ISO 3534-2:2006, definition 1.2.17]

**3.1.12**
**sampling**
act of drawing or constituting a **sample** (3.1.11)



[ISO 3534-2:2006, definition 1.3.1]



**3.1.13**
**sampling unit**
**unit**
one of the individual parts into which a **population** (3.1.6) is divided

NOTE 1 A sampling unit can contain one or more items, for example, a box of matches, but one test result will be obtained for it.

NOTE 2 A sampling unit can consist of discrete items or a defined amount of bulk material.

[ISO 3534-2:2006, definition 1.2.14]

**3.1.14**
**sampling with replacement**
**sampling** (3.1.12) in which each **sampling unit** (3.1.13) taken and observed is returned to the **population** (3.1.6) before the next sampling unit is taken

[ISO 3534-2:2006, definition 1.3.15]

**3.1.15**
**sampling without replacement**
**sampling** (3.1.12) in which each **sampling unit** (3.1.13) is taken from the **population** (3.1.6) once only without being returned to the population

[ISO 3534-2:2006, definition 1.3.16]

**3.1.16**
**seed**
numerical value or set of values used to initialize a **pseudo-independent random sampling** (3.1.7) algorithm or to establish a starting point in a table of random numbers

**3.1.17**
**simple random sample**
**sample** (3.1.11) selected by **simple random sampling** (3.1.18)

[ISO 3534-2:2006, definition 1.2.24]

**3.1.18**
**simple random sampling**
**sampling** (3.1.12) where a **sample** (3.1.11) of $n$ **sampling units** (3.1.13) is taken from a **population** (3.1.6) in such a way that all possible combinations of $n$ sampling units have the same probability of being taken

[ISO 3534-2:2006, definition 1.3.4]

**3.1.19**
**stratified sampling**
**sampling** (3.1.12) such that portions of the **sample** (3.1.11) are drawn from the different **strata** (3.1.21) and each stratum is sampled with at least one **sampling unit** (3.1.13)

[ISO 3534-2:2006, definition 1.3.6]

**3.1.20**
**stratified simple random sampling**
**simple random sampling** (3.1.18) from each **stratum** (3.1.21)



[ISO 3534-2:2006, definition 1.3.7]

**3.1.21**
**stratum**
mutually exclusive and exhaustive sub-population considered to be more homogeneous with respect to the characteristics investigated than the total **population** (3.1.6)

[ISO 3534-2:2006, definition 1.2.29]

## 3.2 Symbols

For the purposes of this document, the mathematical signs and symbols given in ISO 80000-2 and the following apply.

$d_i$ the $i$th (least significant) digit, or face value of a coin or die

$N$ lot size

$n$ sample size

$n_i$ the size of the $i$th sample

$U$ uniformly-distributed random real variable on the open range (0, 1)

$x_i$ the $i$th value of the variable $x$

$j!$ factorial $j$

$\lceil z \rceil$ ceiling function of $z$ (returns the smallest integer greater than or equal to real value $z$)

$\lfloor z \rfloor$ floor function of $z$ (returns the integer portion of real value $z$)

 

## 4 General

**4.1** Random sampling is a prerequisite to the correct application of most sampling plans in industrial use. Similarly, randomization, which uses the principles of random sampling, is indispensable in the conduct of designed experiments, as it increases the internal validity of an experiment, allowing statistical methods to be used in the interpretation of an experiment's results. The goal of random sampling is to provide a means of applying the results of probability theory to practical problems, while avoiding any form of bias. This goal is not attainable using certain other types of sampling. For example, sampling based on such concepts as personal intuition or judgment, haphazardness, or quota-achievement are inherently biased and consequently can lead to serious errors in the decision-making process, with no provision to assess risks. Equi-probable random sampling seeks to eliminate such bias by ensuring that each unit in a lot has the same probability of being selected (sampling with replacement) or, alternatively, that every possible sample of a given size from the lot has the same probability of being selected (sampling without replacement).

**4.2** Under equi-probable random sampling with replacement, the probability that a specific unit in a lot of $N$ units is selected at any given draw is always $1/N$. There are $N^n$ possible ordered random samples of $n$ units from $N$ units and, for completeness, there are $(N + n - 1)! / [n!\ (N - 1)!]$ possible different unordered random samples of $n$ units from $N$ units (see the note below).

Under simple random sampling without replacement, the probability that a unit in a lot is selected at a given draw is $1/N$ for the first draw, $1/(N - 1)$ for the second draw, $1/(N - 2)$ for the third draw, and so on. If $n$ units are randomly selected from a lot of $N$ units without replacement, then each combination of $n$ units has the same probability of selection as every other combination of $N$ units taken $n$ at a time. The number of possible different unordered random samples of $n$ units from a lot of $N$ units is $N! / [n!\ (N - n)!]$, which is the number of combinations of $N$ units taken $n$ at a time. It is equally noteworthy that the number of possible ordered random samples of $n$ units taken without replacement from a lot of $N$ units is $N! / (N - n)!$, which is equivalent to the number of possible permutations of $N$ units taken $n$ at a time. It should be noted that random sampling without replacement is the most common sampling strategy used in acceptance sampling applications.



NOTE Under sampling with replacement based on a sample of, say, 3 units from 5 units, the lists {1, 1, 2}, {1, 2, 1}, and {2, 1, 1} are different when order is considered (and technically referred to as multisets or bags), but the same when order is not considered.



**4.3** The goal of random sampling can only be achieved by adhering to rigorous procedures that have been carefully designed to achieve the intent of the definition. Several methods are presented in this International Standard to implement this goal. The mechanical device methods, in particular, assume that the coins and dice are unbiased, having been designed such that each side has the same probability of occurring during a toss or throw, and that the manner of tossing or throwing is being performed so as not to introduce bias. Furthermore, due to numerous deficiencies in the intrinsic implementations of random sampling methods in calculators and computer operating systems, programming languages, and software (see References [9], [10], [12], and [13] for further information), this International Standard has adopted a portable, proven method for generating random samples by computer. In addition, it should be noted that all of the methods below require that each distinct unit in a lot has been associated in advance with a distinct number from 1 to $N$, so that the sampling units identified as a result of the random sampling method can be unambiguously obtained from the lot.

**4.4** Finally, to reduce awkwardness in presentation, the adjective "pseudo-independent" will often be dropped when referring to such a random sampling procedure or method (see Reference [8]). Furthermore, the adjective “random” will be used frequently in the sense that the noun it modifies (often a number or permutation) is the output of a process that randomly generates such a number or permutation. In addition, when examples are provided, the sample sizes involved are artificially kept small with the goal of simply illustrating the concepts involved.

## 5 Random sampling — Mechanical device methods

### 5.1 Urn method

**5.1.1** Place $N$ distinctly-numbered but otherwise physically-identical objects (e.g. tickets, chips, or balls) into an urn to unambiguously represent each of the $N$ units in the lot and thoroughly mix the objects.

**5.1.2** For sampling without replacement, blindly select objects from the urn, one by one without returning them to the urn and optionally re-mixing the objects between successive draws, until the desired number $n$ of sampling units is obtained.

NOTE This method is commonly used by lottery agencies.

**5.1.3** For sampling with replacement, blindly select objects from the urn, one by one, returning each object to the urn after each draw and thoroughly re-mixing the objects between successive draws, until the desired number $n$ of sampling units is obtained. Using this method, the same unit may occur more than once in the sample.

### 5.2 Coin or die method

**5.2.1** Determine the number $m$ of coins or dice (or coin tosses or die throws) required, where $N$ is the lot size and $k$ is the number of sides of the device being used, according to the following equation:

$$m = \lceil \log_e N / \log_e k \rceil$$



**5.2.2** Where multiple coins or dice are used, clearly associate each coin or die with a specific position in the interpretation sequence of digits $d_i$. Where a single coin or die is used, assign the result of the first toss or throw to the most significant digit $d_m$, the second toss or throw to the next most significant digit $d_{m-1}$, and so on.



**5.2.3** Toss the coins or throw the dice and record the $m$ ordered results $d_i$. Translate the results to decimal integers through the following equation:

$$y = 1 + \sum_{i=1}^{m} (d_i - 1) k^{m-i}$$

**5.2.4** Repeat step 5.2.3, discarding all values that exceed $N$ and, in the case of sampling without replacement, all values that have already been selected, until the desired number $n$ of sampling units is obtained.

EXAMPLE 1 An inspector wishes to obtain a random sample of 4 units from a lot of 20 units and has a single coin available. From step 5.2.1, it is determined that $m = 5$ coin tosses are required to obtain each random number. It is decided in advance that a head will have a face value of 1 and a tail a face value of 2. The first sequence of tosses yields the multiset {1, 2, 1, 2, 2}, which through step 5.2.3 equates to $1 + (0)2^4 + (1)2^3 + (0)2^2 + (1)2^1 + (1)2^0 = 12$. The following three sequences of tosses yield the multisets {1, 2, 2, 2, 1}, {1, 1, 2, 2, 1}, and {2, 2, 1, 2, 2}, which equate to 15, 7, and 28, respectively. Since the value 28 exceeds the lot size, it needs to be discarded and additional sequences of tosses need to be performed until one more valid number is obtained to complete the random sample.

EXAMPLE 2 A random sample of 4 units from a lot of 50 units is required and the inspector has access to several six-sided dice of different colours. From step 5.2.1, it is determined that $m = 3$ dice are required to obtain each random number. The inspector chooses a blue, a green, and a red die and ranks them from most significant to least significant in that same order. However, it is evident upon examining the equation of 5.2.3 that numbers within the valid range from 1 to 50 will only result when the first die face is either 1 or 2. Consequently, some efficiency can be obtained by mapping the higher face values of the blue die to either 1 or 2 without distorting the outcome probabilities. The inspector decides in advance for odd face values on the blue die to be treated as 1 and even face values to be treated as 2. The first roll yields the multiset {3, 3, 4}, which through step 5.2.3 equates to $1 + (2)6^2 + (2)6^1 + (3)6^0 = 88$, which is too large but when transformed to {1, 3, 4} equates to 16. Three more rolls yield {6, 1, 3}, (which transforms to {2, 1, 3}), {5, 6, 6}, (which transforms to {1, 6, 6}), and {2, 5, 5}, which equate to 39, 36, and 65, respectively. Since the value 65 exceeds the lot size, it needs to be discarded and additional throws need to be made until one more valid number is obtained to complete the random sample.