

INTERNATIONAL
STANDARD

ISO
3534-1

NORME
INTERNATIONALE

Second edition
Deuxième édition
2006-10-15

Corrected version
Version corrigée
2007-09-15

Statistics — Vocabulary and symbols —

Part 1:

**General statistical terms and terms used
in probability**

Statistique — Vocabulaire et symboles —

Partie 1:

**Termes statistiques généraux et termes
utilisés en calcul des probabilités**



Reference number
Numéro de référence
ISO 3534-1:2006(E/F)

Documents
© ISO 2006
AVA Document Sharing Hub

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

PDF – Exonération de responsabilité

Le présent fichier PDF peut contenir des polices de caractères intégrées. Conformément aux conditions de licence d'Adobe, ce fichier peut être imprimé ou visualisé, mais ne doit pas être modifié à moins que l'ordinateur employé à cet effet ne bénéficie d'une licence autorisant l'utilisation de ces polices et que celles-ci y soient installées. Lors du téléchargement de ce fichier, les parties concernées acceptent de fait la responsabilité de ne pas enfreindre les conditions de licence d'Adobe. Le Secrétariat central de l'ISO décline toute responsabilité en la matière.

Adobe est une marque déposée d'Adobe Systems Incorporated.

Les détails relatifs aux produits logiciels utilisés pour la création du présent fichier PDF sont disponibles dans la rubrique General Info du fichier; les paramètres de création PDF ont été optimisés pour l'impression. Toutes les mesures ont été prises pour garantir l'exploitation de ce fichier par les comités membres de l'ISO. Dans le cas peu probable où surviendrait un problème d'utilisation, veuillez en informer le Secrétariat central à l'adresse donnée ci-dessous.



**COPYRIGHT PROTECTED DOCUMENT
DOCUMENT PROTÉGÉ PAR COPYRIGHT**

© ISO 2006

The reproduction of the terms and definitions contained in this International Standard is permitted in teaching manuals, instruction booklets, technical publications and journals for strictly educational or implementation purposes. The conditions for such reproduction are: that no modifications are made to the terms and definitions; that such reproduction is not permitted for dictionaries or similar publications offered for sale; and that this International Standard is referenced as the source document.

With the sole exceptions noted above, no other part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

La reproduction des termes et des définitions contenus dans la présente Norme internationale est autorisée dans les manuels d'enseignement, les modes d'emploi, les publications et revues techniques destinés exclusivement à l'enseignement ou à la mise en application. Les conditions d'une telle reproduction sont les suivantes: aucune modification n'est apportée aux termes et définitions; la reproduction n'est pas autorisée dans des dictionnaires ou publications similaires destinés à la vente; la présente Norme internationale est citée comme document source.

À la seule exception mentionnée ci-dessus, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'ISO à l'adresse ci-après ou du comité membre de l'ISO dans le pays du demandeur.

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/Publié en Suisse

Contents

Page

Foreword	v
Introduction	ix
Scope	1
1 General statistical terms	2
2 Terms used in probability	37
Annex A (informative) Symbols	81
Annex B (informative) Statistical concept diagrams	84
Annex C (informative) Probability concept diagrams	90
Annex D (informative) Methodology used in the development of the vocabulary	94
Bibliography	101
Alphabetical index	102
French alphabetical index (Index alphabétique)	104

Sommaire

Page

Avant-propos	vii
Introduction	x
Domaine d'application	1
1 Termes statistiques généraux	2
2 Termes utilisés en probabilité	37
Annexe A (informative) Symboles.....	81
Annexe B (informative) Diagrammes de concept.....	84
Annexe C (informative) Diagramme de concept de probabilité	90
Annexe D (informative) Méthodologie utilisée pour élaborer le vocabulaire	94
Bibliographie	101
Index alphabétique anglais (Alphabetical index)	102
Index alphabétique.....	104

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 3534-1 was prepared by Technical Committee ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 1, *Terminology and symbols*.

This second edition cancels and replaces the first edition (ISO 3534-1:1993), which has been technically revised.

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*

This corrected version of ISO 3534-1:2006 incorporates the following corrections.

- | | |
|------------------|--|
| 1.5 (Example 1) | “ Range (1.10)” was changed to “ sample range (1.10)”. |
| 1.18 (Note) | The last sentence was replaced. |
| 1.19 (Example) | The last number was changed to –0,64. |
| 1.21 (Note 3) | The last sentence was replaced. |
| 1.23 (Example 1) | 12,945 was changed to 12,948 and 257,118 was changed to 257,178. |
| 1.23 (Note 1) | The first sentence after the equation was replaced. |
| 1.23 (Note 2) | The last sentence was replaced. |
| 1.51 (Note) | The note was replaced. |
| 1.57 (Note) | A note was added. |
| 2.48 (Note 3) | The word “then” was removed. |
| 2.64 (Note) | “Multivariate distribution” was changed to “Multivariate normal distribution”. |
| 2.65 (Notes) | Note was changed to Note 1. Note 2 was added. |
| 2.68 (Example 2) | The third sentence was corrected. |

- Figure B.2 k was added after “sample moment of order”.
- Figure B.3 The spelling of “estimator” was corrected.
- Figure B.4 The order of the French and English terms for (1.48) was corrected.
- Figure B.5 Before (1.5), (1.53) and (1.54), “statistic” was replaced by “statistics”.
- Figure C.1 The “sample space” reference was changed to (2.1).
“probability (2.6)” was changed to “probability of an event A (2.5)”.
“conditional probability of A given B (2.6)” was changed to
“conditional probability $P(A|B)$ (2.6)”.
“independent event (2.4)” was changed to “independent events (2.4)”.
 $(-\infty, x)$ has been changed to $(-\infty, x]$.
- Figure C.2 “correlation (2.44)” was changed to “correlation coefficient (2.44)”.
- Figure C.4 The spelling of “exponential” was corrected.

Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (CEI) en ce qui concerne la normalisation électrotechnique.

Les Normes internationales sont rédigées conformément aux règles données dans les Directives ISO/CEI, Partie 2.

La tâche principale des comités techniques est d'élaborer les Normes internationales. Les projets de Normes internationales adoptés par les comités techniques sont soumis aux comités membres pour vote. Leur publication comme Normes internationales requiert l'approbation de 75 % au moins des comités membres votants.

L'attention est appelée sur le fait que certains des éléments du présent document peuvent faire l'objet de droits de propriété intellectuelle ou de droits analogues. L'ISO ne saurait être tenue pour responsable de ne pas avoir identifié de tels droits de propriété et averti de leur existence.

L'ISO 3534-1 a été élaborée par le comité technique ISO/TC 69, *Application des méthodes statistiques*, sous-comité SC 1, *Terminologie et symboles*.

Cette deuxième édition annule et remplace la première édition (ISO 3534-1:1993), qui a fait l'objet d'une révision technique.

L'ISO 3534 comprend les parties suivantes, présentées sous le titre général *Statistique — Vocabulaire et symboles*:

- *Partie 1: Termes statistiques généraux et termes utilisés en calcul des probabilités*
- *Partie 2: Statistique appliquée*
- *Partie 3: Plans d'expérience*

La présente version corrigée de l'ISO 3534-1:2006 comprend les corrections suivantes.

- | | |
|------------------|--|
| 1.5 (Exemple 1) | « l'étendue (1.10)» a été remplacé par « l'étendue d'échantillon (1.10)». |
| 1.18 (Note) | La dernière phrase a été modifiée et une phrase a été ajoutée. |
| 1.19 (Exemple) | Le dernier nombre a été modifié de sorte à lire –0,64. |
| 1.21 (Note 3) | «Pour les valeurs d'aplatissement» a été remplacé par «Pour des valeurs d'aplatissement» et la dernière phrase a été modifiée. |
| 1.23 (Exemple 1) | 12,945 a été changé en 12,948 et 257,118 a été changé en 257,178. |
| 1.23 (Note 1) | La première phrase après l'équation a été modifiée. |
| 1.23 (Note 2) | La dernière phrase a été modifiée. |
| 1.51 (Note) | La note a été modifiée. |
| 1.57 (Note) | Une note a été ajoutée. |
| 2.64 (Note) | «distributions à plusieurs variables» a été remplacé par «distributions normales à plusieurs variables». |

2.65 (Notes)	La note a été numérotée en Note 1. Une Note 2 a été ajoutée.
2.68 (Exemple 2)	La troisième phrase a été modifiée de sorte à lire: «Si les batteries sont opérationnelles, leurs temps de survie suivent ...».
Figure B.2	k a été ajouté après «moment d'échantillon d'ordre».
Figure B.4	L'ordre des termes français et anglais (1.48) a été corrigé.
Figure C.1	«étendu d'échantillon, Ω (2.68)» a été remplacé par «espace d'échantillon, Ω (2.1)». «probabilité (2.6)» a été remplacé par «probabilité d'un événement A (2.5)». «probabilité conditionnelle de A sachant B (2.6) a été remplacé par «probabilité conditionnelle $P(A B)$ (2.6)». $(-\infty, x)$ a été remplacé par $(-\infty, x]$.
Figure C.2	«corrélation (2.44)» a été remplacé par «coefficient de corrélation (2.44)».

Introduction

The current versions of ISO 3534-1 and ISO 3534-2 are intended to be compatible. They share the common goal of restricting their respective mathematical levels to the minimum levels necessary to attain coherent, correct and concise definitions. Part 1 on terms used in probability and statistics is fundamental, so by necessity is presented at a somewhat sophisticated mathematical level. Recognizing that users of ISO 3534-2 or other TC 69 standards on applied statistics may occasionally consult this part of ISO 3534 for the definition of certain terms, several of the terms are described in a less technical manner within the notes and are illustrated with examples. Although these informal descriptions are not a substitute for formal definitions, they may provide a working, layman definition of concepts, thus serving the needs of multiple users of these terminology standards. To accommodate further the applied user who normally would be involved with standards such as ISO 3534-2 or ISO 5725, for example, notes and examples are offered to make this part of ISO 3534 more accessible.

A well-defined and reasonably complete set of probability and statistical terms is essential to the development of and effective use of statistical standards. The definitions provided here must be sufficiently accurate and mathematically sophisticated to enable statistical standards developers to avoid ambiguities. Of course, more detailed explanations of concepts, their contexts and their realms of application can be found in introductory probability and statistics textbooks.

Concept diagrams are provided in an informative annex for each group of terms: 1) general statistical terms (in Annex B) and 2) terms used in probability (in Annex C). There are six concept diagrams for general statistical terms and four concept diagrams for terms related to probability. Some terms appear in multiple diagrams to provide a link from one set of concepts to another. Annex D provides a brief introduction to Concept Diagrams and their interpretation.

These diagrams were instrumental in constructing this revision as they assist in delineating the interrelationships of the various terms. These diagrams are also likely to be useful in translating the standard into other languages.

As a general comment with respect to much of the standard, unless otherwise indicated, the definitions relate to the one-dimensional (univariate) case. This provision is admitted here to eliminate the need to mention repetitively the one-dimensional scope for most of the definitions.

Introduction

Les versions actuelles de l'ISO 3534-1 et de l'ISO 3534-2 sont destinées à être compatibles. Elles partagent l'objectif commun de réduire leurs niveaux mathématiques au strict minimum pour obtenir des définitions cohérentes, correctes et concises. La présente partie de l'ISO 3534, relative aux termes utilisés dans le domaine des probabilités et des statistiques, est fondamentale et se présente de ce fait à un niveau mathématique relativement sophistiqué. Dans la mesure où les utilisateurs de l'ISO 3534-2 ou d'autres normes du TC 69 relatives à la statistique appliquée peuvent avoir à consulter la présente partie de l'ISO 3534, pour certains termes, les notes figurant à la suite des termes sont rédigées de manière moins technique et sont illustrées par des exemples. Bien que ces notes informelles ne remplacent nullement des définitions formelles, elles peuvent donner une définition de travail non technique des concepts et ainsi répondre aux besoins des nombreux utilisateurs de ces normes terminologiques. Des notes et des exemples sont ainsi proposés pour faciliter la compréhension de la présente partie de l'ISO 3534, et s'adapter davantage à l'utilisateur spécialiste qui, normalement, serait amené à consulter des normes telles que l'ISO 3534-2 ou l'ISO 5725.

Un ensemble correctement défini et suffisant de termes de probabilité et de termes statistiques est fondamental pour l'élaboration et l'utilisation efficace des normes statistiques. Les définitions du présent document doivent être suffisamment précises et d'un niveau mathématique assez sophistiqué pour permettre aux rédacteurs de normes statistiques d'éviter toute ambiguïté. De plus amples explications des concepts, de leurs contextes ainsi que de leurs domaines d'application peuvent naturellement être retrouvées en consultant les livres d'introduction à la probabilité et aux statistiques.

Des diagrammes de concepts sont fournis dans une annexe informative pour chaque groupe de termes: 1) termes statistiques généraux (voir Annexe B) et 2) termes utilisés en probabilité (voir Annexe C). Il existe six diagrammes de concepts pour les termes statistiques généraux et quatre diagrammes de concepts pour les termes relatifs à la probabilité. Certains des termes apparaissent dans plusieurs de ces diagrammes afin de fournir un lien permettant de passer d'un groupe de concepts à un autre groupe de concepts. L'Annexe D fournit une introduction succincte aux Diagrammes de concept et à leur interprétation.

Ces diagrammes ont été utilisés comme outils pour l'élaboration de cette révision dans la mesure où ils aident au cadrage des interrelations entre les différents termes. Ces diagrammes sont également susceptibles d'être utiles à la traduction de la norme dans d'autres langues.

D'une manière générale, eu égard au contenu global de la norme, et sauf indication contraire, les définitions font référence au cas à une dimension (à une variable). Cette disposition est admise dans le présent document afin d'éviter la répétition obligatoire de la portée unidimensionnelle de la plupart des définitions.

Statistics — Vocabulary and symbols —

Part 1: General statistical terms and terms used in probability

Scope

This part of ISO 3534 defines general statistical terms and terms used in probability which may be used in the drafting of other International Standards. In addition, it defines symbols for a limited number of these terms.

The terms are classified as:

- a) general statistical terms (Clause 1);
- b) terms used in probability (Clause 2).

Annex A gives a list of symbols and abbreviations recommended to be used for this part of ISO 3534.

The entries in this part of ISO 3534 are arranged in association with concept diagrams provided as Annexes B and C.

Statistique — Vocabulaire et symboles —

Partie 1: Termes statistiques généraux et termes utilisés en calcul des probabilités

Domaine d'application

La présente partie de l'ISO 3534 définit les termes statistiques généraux et les termes utilisés en calcul des probabilités susceptibles d'être utilisés dans la rédaction d'autres Normes internationales. En outre, elle définit un ensemble de symboles pour un nombre limité de ces termes.

Les termes sont classés sous les rubriques suivantes:

- a) termes statistiques généraux (Article 1);
- b) termes utilisés en calcul des probabilités (Article 2).

L'Annexe A donne une liste des symboles et abréviations utilisés dans la présente partie de l'ISO 3534.

Les entrées dans la présente partie de l'ISO 3534 sont classées en association avec les diagrammes de concepts fournis dans les Annexes B et C.

1 General statistical terms

1.1 population

totality of items under consideration

NOTE 1 A population may be real and finite, real and infinite or completely hypothetical. Sometimes the term "finite population" is used, especially in survey sampling. Likewise the term "infinite population" is used in the context of sampling from a continuum. In Clause 2, population will be viewed in a probabilistic context as the **sample space** (2.1).

NOTE 2 A hypothetical population allows one to imagine the nature of further data under various assumptions. Hence, hypothetical populations are useful at the design stage of statistical investigations, particularly for determining appropriate sample sizes. A hypothetical population could be finite or infinite in number. It is a particularly useful concept in inferential statistics to assist in evaluating the strength of evidence in a statistical investigation.

NOTE 3 The context of an investigation can dictate the nature of the population. For example, if three villages are selected for a demographic or health study, then the population consists of the residents of these particular villages. Alternatively, if the three villages were selected at random from among all of the villages in a specific region, then the population would consist of all residents of the region.

1.2 sampling unit

one of the individual parts into which a **population** (1.1) is divided

NOTE Depending on the circumstances the smallest part of interest may be an individual, a household, a school district, an administrative unit and so forth.

1.3 sample

subset of a **population** (1.1) made up of one or more **sampling units** (1.2)

NOTE 1 The sampling units could be items, numerical values or even abstract entities depending on the population of interest.

NOTE 2 The definition of sample in ISO 3534-2 includes an example of a sampling frame which is essential in drawing a random sample from a finite population.

1 Termes statistiques généraux

1.1 population

totalité des individus pris en considération

NOTE 1 Une population peut être réelle et finie, réelle et infinie ou totalement hypothétique. Le terme «population finie» est parfois utilisé, particulièrement dans l'échantillonnage d'enquête. De même le terme «population infinie» est utilisé dans le contexte de l'échantillonnage continu. Dans l'Article 2, la population est abordée dans le contexte de la probabilité comme l'**espace d'échantillon** (2.1).

NOTE 2 Une population hypothétique permet d'imaginer la nature des données futures sous différentes hypothèses. Les populations hypothétiques sont donc utiles pour l'étape de conception des enquêtes statistiques, particulièrement pour déterminer les tailles d'échantillon appropriées. Une population hypothétique peut être finie ou infinie en nombre. C'est un concept particulièrement utile dans le cas des statistiques déductives pour aider à évaluer la force de la preuve dans une enquête statistique.

NOTE 3 Le contexte d'une enquête peut dicter la nature de la population. Par exemple, si trois villages sont sélectionnés pour une enquête de démographie et de santé, la population est alors constituée des habitants de ces trois villages uniquement. Sinon, si ces trois villages sont sélectionnés aléatoirement parmi tous les villages d'une région donnée, la population est alors constituée de tous les habitants de la région.

1.2 unité d'échantillonnage

unité individuelle en laquelle une **population** (1.1) est divisée

NOTE Selon les circonstances, la plus petite partie considérée peut être un individu, un ménage, un district scolaire, une unité administrative, etc.

1.3 échantillon

sous-ensemble d'une **population** (1.1) constitué d'une ou de plusieurs unités **d'échantillonnage** (1.2)

NOTE 1 En fonction de la population considérée, les unités d'échantillonnage peuvent être des individus, des valeurs numériques ou encore des entités abstraites.

NOTE 2 La définition de ce terme donnée dans l'ISO 3534-2 comprend un exemple de base d'échantillonnage qui revêt un caractère essentiel pour prélever un échantillon au hasard d'une population finie.

1.4**observed value**

obtained value of a property associated with one member of a **sample** (1.3)

NOTE 1 Common synonyms are "realization" and "datum". The plural of datum is data.

NOTE 2 The definition does not specify the genesis or how this value has been obtained. The value may represent one realization of a **random variable** (2.10), but not exclusively so. It may be one of several such values that will be subsequently subjected to statistical analysis. Although proper inferences require some statistical underpinnings, there is nothing to preclude computing summaries or graphical depictions of observed values. Only when attendant issues such as determining the probability of observing a specific set of realizations does the statistical machinery become both relevant and essential. The preliminary stage of an analysis of observed values is commonly referred to as data analysis.

1.5**descriptive statistics**

graphical, numerical or other summary depiction of **observed values** (1.4)

EXAMPLE 1 Numerical summaries include **average** (1.15), **sample range** (1.10), **sample standard deviation** (1.17), and so forth.

EXAMPLE 2 Examples of graphical summaries include boxplots, diagrams, Q-Q plots, normal quantile plots, scatterplots, multiple scatterplots and histograms.

1.4**valeur observée**

valeur obtenue d'une propriété associée à un élément d'un **échantillon** (1.3)

NOTE 1 Des synonymes courants sont les termes «réalisation» et «donnée».

NOTE 2 La définition ne précise pas l'origine ni la manière dont cette valeur a été obtenue. La valeur peut représenter une réalisation d'une **variable aléatoire** (2.10) mais pas de manière exclusive. Il peut s'agir d'une des différentes valeurs qui seront soumises ultérieurement à une analyse statistique. Bien que des inférences appropriées nécessitent des validations statistiques, rien n'empêche de procéder à des calculs de synthèses ou à des représentations graphiques des valeurs observées. Ce n'est que dans le cas de problèmes concomitants tels que la détermination de la probabilité d'observer un ensemble spécifique de réalisations que les mécanismes statistiques deviennent à la fois pertinents et fondamentaux. L'étape préliminaire d'une analyse de valeurs observées fait souvent référence à une analyse de données.

1.5**statistique descriptive**

description graphique, numérique ou autre analyse de synthèse des **valeurs observées** (1.4)

EXAMPLE 1 Les synthèses numériques comprennent la **moyenne** (1.15), **l'étendue d'échantillon** (1.10), **l'écart-type d'échantillon** (1.17), etc.

EXAMPLE 2 Des exemples de représentations graphiques synthétiques sont les diagrammes à surfaces, les diagrammes, les graphiques Q-Q, les diagrammes de quantile normal, les nuages de points, les nuages de points multidimensionnels et les histogrammes.

1.6

random sample

sample (1.3) which has been selected by a method of random selection

NOTE 1 This definition is less restrictive than that given in ISO 3534-2 to allow for infinite populations.

NOTE 2 When the sample of n sampling units is selected from a finite **sample space** (2.1), each of the possible combinations of n sampling units will have a particular **probability** (2.5) of being taken. For survey sampling plans, the particular probability for each possible combination may be calculated in advance.

NOTE 3 For survey sampling from a finite sample space, a random sample can be selected by different sampling plans such as stratified random sampling, systematic random sampling, cluster sampling, sampling with probability of sampling proportional to the size of an auxiliary variable and many other possibilities.

NOTE 4 The definition generally refers to actual **observed values** (1.4). These observed values are considered as realizations of **random variables** (2.10), where each observed value corresponds to one random variable. When **estimators** (1.12), test statistics for **statistical tests** (1.48) or **confidence intervals** (1.28) are derived from a random sample, the definition accommodates reference to the random variables arising from abstract entities in the sample rather than the actual observed values of these random variables.

NOTE 5 Random samples from infinite populations are often generated by repeated draws from the sample space, leading to a sample consisting of independent, identically distributed random variables using the interpretation of this definition mentioned in Note 4.

1.7

simple random sample

(finite population) **random sample** (1.6) such that each subset of a given size has the same probability of selection

NOTE This definition is in harmony with the definition given in ISO 3534-2, although the wording here is slightly different.

1.6

échantillon aléatoire

échantillon (1.3) prélevé selon une méthode de sélection aléatoire

NOTE 1 Cette définition est moins limitative que celle donnée dans l'ISO 3534-2 et permet d'admettre des populations infinies.

NOTE 2 Lorsque l'échantillon de n unités d'échantillonnage est prélevé d'un **espace d'échantillon** (2.1) fini, chacune des combinaisons possibles de n unités d'échantillonnage aura une **probabilité** (2.5) particulière d'être prélevée. Pour les plans d'échantillonnage d'enquête, la probabilité particulière pour chaque combinaison possible peut être calculée à l'avance.

NOTE 3 Pour l'échantillonnage d'enquête réalisé à partir d'un espace d'échantillon fini, il est possible de prélever un échantillon aléatoire par différents plans d'échantillonnage tels que l'échantillonnage stratifié aléatoire, l'échantillonnage systématique aléatoire, l'échantillonnage en grappe, l'échantillonnage avec probabilité d'échantillonnage proportionnel à l'effectif d'une variable auxiliaire et de nombreuses autres possibilités.

NOTE 4 La définition fait généralement référence aux **valeurs observées** (1.4) réellement. Ces valeurs observées sont considérées comme des réalisations de **variables aléatoires** (2.10), où chaque valeur observée correspond à une variable aléatoire. Lorsque des **estimateurs** (1.12), des statistiques de test pour **tests statistiques** (1.48) ou des **intervalles de confiance** (1.28) sont dérivés d'un échantillon aléatoire, la définition fait référence aux variables aléatoires issues d'entités abstraites dans l'échantillon plutôt qu'aux valeurs observées réellement de ces variables aléatoires.

NOTE 5 Des échantillons aléatoires prélevés de populations infinies sont souvent générés par des prélèvements répétés de l'espace d'échantillon et donnent lieu à un échantillon composé de variables aléatoires indépendantes distribuées de manière identique en utilisant l'interprétation de cette définition indiquée dans la Note 4.

1.7

échantillon simple aléatoire

(population finie) **échantillon aléatoire** (1.6) tel que chacun des sous-ensembles d'une taille donnée ait la même probabilité de sélection

NOTE Cette définition concorde avec celle donnée dans l'ISO 3534-2, même si le libellé est légèrement différent.

1.8 statistic

completely specified function of **random variables** (2.10)

NOTE 1 A statistic is a function of random variables in a **random sample** (1.6) in the sense given in Note 4 of 1.6.

NOTE 2 Referring to Note 1, if $\{X_1, X_2, \dots, X_n\}$ is a random sample from a **normal distribution** (2.50) with unknown **mean** (2.35) μ and unknown **standard deviation** (2.37) σ , then the expression $(X_1 + X_2 + \dots + X_n)/n$ is a statistic, the **sample mean** (1.15), whereas $[(X_1 + X_2 + \dots + X_n)/n] - \mu$ is not a statistic as it involves the unknown value of the **parameter** (2.9) μ .

NOTE 3 The definition given here is a technical one, corresponding to the treatment found in mathematical statistics. In application settings, the plural of statistic, namely statistics, can refer to the technical discipline involving the analysis activities described in ISO/TC 69 International Standards.

1.8 statistique

fonction totalement spécifiée de **variables aléatoires** (2.10)

NOTE 1 Une statistique est une fonction des variables aléatoires d'un **échantillon aléatoire** (1.6) au sens donné dans la Note 4 en 1.6.

NOTE 2 En référence à la Note 1, si $(X_1, X_2, \dots, X_n)$ est un échantillon aléatoire issu d'une **loi normale** (2.50) à **moyenne** (2.35) μ inconnue et **écart-type** (2.37) σ inconnu, l'expression $(X_1 + X_2 + \dots + X_n)/n$ est une statistique, la **moyenne d'échantillon** (1.15), alors que $[(X_1 + X_2 + \dots + X_n)/n] - \mu$ n'est pas une statistique dans la mesure où elle implique la valeur inconnue du **paramètre** (2.9) μ .

NOTE 3 La présente définition relève du domaine technique de la statistique mathématique. Dans les applications, le pluriel de «statistique», nommé «statistiques», peut faire référence à la discipline technique comprenant les activités d'analyse de données décrites dans les Normes internationales de l'ISO/TC 69.

1.9 order statistic

statistic (1.8) determined by its ranking in a non-decreasing arrangement of **random variables** (2.10)

EXAMPLE Let the observed values of a sample be 9, 13, 7, 6, 13, 7, 19, 6, 10, and 7. The observed values of the order statistics are 6, 6, 7, 7, 7, 9, 10, 13, 13, 19. These values constitute realizations of $X_{(1)}$ through $X_{(10)}$.

NOTE 1 Let the **observed values** (1.4) of a **random sample** (1.6) be $\{x_1, x_2, \dots, x_n\}$ and once sorted in non-decreasing order designated as $x_{(1)} \leq \dots \leq x_{(k)} \leq \dots \leq x_{(n)}$. Then $(x_{(1)}, \dots, x_{(k)}, \dots, x_{(n)})$ is the observed value of the order statistic $(X_{(1)}, \dots, X_{(k)}, \dots, X_{(n)})$ and $x_{(k)}$ is the observed value of the k^{th} order statistic.

NOTE 2 In practical terms, obtaining the order statistics for a data set amounts to sorting the data as formally described in Note 1. The sorted form of the data set then lends itself to obtaining useful summary statistics as given in the next few definitions.

NOTE 3 Order statistics involve sample values identified by their position after ranking in non-decreasing order. As in the example, it is easier to understand the sorting of sample values (realizations of random variables) rather than the sorting of unobserved random variables. Nevertheless, one can conceive of random variables from a **random sample** (1.6) being arranged in a non-decreasing order. For example, the maximum of n random variables can be studied in advance of its realized value.

NOTE 4 An individual order statistic is a statistic which is a completely specified function of a random variable. This function is simply the identity function with the further identification of position or rank in the sorted set of random variables.

NOTE 5 Tied values pose a potential problem especially for discrete random variables and for realizations that are reported to low resolution. The word "non-decreasing" is used rather than "ascending" as a subtle approach to the problem. It should be emphasized that tied values are retained and not collapsed into the single tied value. In the example above, the two realizations of 6 and 6 are tied values.

NOTE 6 Ordering takes place with reference to the real line and not to the absolute values of the random variables.

NOTE 7 The complete set of order statistics consist of an n dimensional random variable, where n is the number of observations in the sample.

NOTE 8 The components of the order statistic are also referred to as order statistics but with a qualifier that gives the number in the sequence of ordered values of the sample.

NOTE 9 The minimum, the maximum, and for odd-numbered sample sizes, the **sample median** (1.13), are special cases of order statistics. For example, for sample size 11, $X_{(1)}$ is the minimum, $X_{(11)}$ is the maximum and $X_{(6)}$ is the sample median.

1.9 statistique d'ordre

statistique (1.8) déterminée par le rang qu'elle occupe dans un ordre non décroissant de **variables aléatoires** (2.10)

EXEMPLE Soit les valeurs observées d'un échantillon 9, 13, 7, 6, 13, 7, 19, 6, 10, et 7. Les valeurs observées de la statistique d'ordre sont 6, 6, 7, 7, 7, 9, 10, 13, 13, 19. Ces valeurs constituent les réalisations de $X_{(1)}$ à $X_{(10)}$.

NOTE 1 Soit les **valeurs observées** (1.4) d'un **échantillon aléatoire** (1.6) $\{x_1, x_2, \dots, x_n\}$ rangées dans un ordre non décroissant $x_{(1)} \leq \dots \leq x_{(k)} \leq \dots \leq x_{(n)}$. Alors $(x_{(1)}, \dots, x_{(k)}, \dots, x_{(n)})$ est la valeur observée de la statistique d'ordre $(X_{(1)}, \dots, X_{(k)}, \dots, X_{(n)})$ et $x_{(k)}$ est la valeur observée de la $k^{\text{ème}}$ statistique d'ordre.

NOTE 2 Dans la pratique, la statistique d'ordre pour un ensemble de données est obtenue en ordonnant les données tel que décrit dans la Note 1. La forme ordonnée de l'ensemble de données permet ensuite d'obtenir des statistiques synthétiques sommaires telles que décrites dans les quelques définitions suivantes.

NOTE 3 Les statistiques d'ordre impliquent des valeurs d'échantillon identifiées par leur position lorsqu'elles sont rangées par ordre non décroissant. Comme dans l'exemple, il est facile de comprendre le classement des valeurs d'échantillon (réalisations de variables aléatoires) plutôt que le classement des variables aléatoires non observées. Cependant, il est possible d'envisager de pouvoir ordonner dans un ordre non décroissant des variables aléatoires issues d'un **échantillon aléatoire** (1.6). Par exemple, le maximum de n variables aléatoires peut être étudié avant que sa valeur soit réalisée.

NOTE 4 Une statistique d'ordre individuelle est une statistique qui est une fonction totalement spécifiée d'une variable aléatoire. Cette fonction est tout simplement la fonction d'identité composée avec l'identification de position ou de rang dans l'ensemble ordonné des variables aléatoires.

NOTE 5 Les valeurs ex-æquo soulèvent un problème potentiel, notamment pour des variables aléatoires discrètes et des réalisations qui sont exprimées avec une faible résolution. L'utilisation de l'expression «non décroissant» au lieu de «ascendant» constitue une approche subtile du problème. Il convient de noter que les valeurs ex-æquo sont conservées et non groupées en une valeur unique. Dans l'exemple ci-dessus, les deux réalisations de 6 et 6 sont des valeurs ex-æquo.

NOTE 6 Le classement est réalisé algébriquement et non avec les valeurs absolues des variables aléatoires.

NOTE 7 L'ensemble complet de statistiques d'ordre est constitué d'une variable aléatoire de n dimensions, où n est le nombre d'observations dans l'échantillon.

NOTE 8 Les composantes de la statistique d'ordre sont également désignées comme des statistiques d'ordre mais avec un qualificatif qui donne le rang dans la séquence des valeurs ordonnées de l'échantillon.

NOTE 9 La taille d'échantillon minimale, la taille d'échantillon maximale et, pour les tailles d'échantillon impaires, la **taille d'échantillon médiane** (1.13) sont des cas particuliers de statistiques d'ordre. Par exemple, pour la taille d'échantillon 11, $X_{(1)}$ est la taille d'échantillon minimale, $X_{(11)}$ est la taille d'échantillon maximale et $X_{(6)}$ est la taille d'échantillon médiane.

1.10 sample range

largest **order statistic** (1.9) minus the smallest order statistic

EXAMPLE Continuing with the example from 1.9, the observed sample range is $19 - 6 = 13$.

NOTE In statistical process control, the sample range is often used to monitor the dispersion over time of a process, particularly when the sample sizes are relatively small.

1.11 mid-range average (1.15) of smallest and largest order statistics (1.9)

EXAMPLE The observed mid-range for the example values in 1.9 is $(6+19)/2 = 12,5$.

NOTE The mid-range provides a quick and simple assessment of the middle of small data sets.

1.12 estimator

$\hat{\theta}$

statistic (1.8) used in **estimation** (1.36) of the parameter θ

NOTE 1 An estimator could be the **sample mean** (1.15) intended to estimate the population **mean** (2.35), which could be denoted by μ . For a **distribution** (2.11) such as the **normal distribution** (2.50), the “natural” estimator of the population mean μ is the sample mean.

NOTE 2 For estimating a population property [e.g. the **mode** (2.27) for a **univariate distribution** (2.16)], an appropriate estimator could be a function of the estimator(s) of the parameter(s) of a distribution or could be a complex function of a **random sample** (1.6).

NOTE 3 The term “estimator” is used here in a broad sense. It includes the point estimator for a parameter, as well as the interval estimator which is possibly used for prediction (sometimes referred to as a predictor). Estimator also can include functions such as kernel estimators and other special purpose statistics. Additional discussion is provided in the notes to 1.36.

1.10 étendue d'échantillon

plus grande **statistique d'ordre** (1.9) moins la plus petite statistique d'ordre

EXEMPLE En reprenant l'exemple donné en 1.9, l'étendue d'échantillon observée est $19 - 6 = 13$.

NOTE En maîtrise statistique des processus, l'étendue d'échantillon est souvent utilisée pour surveiller la dispersion dans le temps d'un processus, notamment lorsque les effectifs d'échantillon sont relativement petits.

1.11 milieu de l'étendue moyenne (1.15) de la plus petite et de la plus grande des statistiques d'ordre (1.9)

EXEMPLE Le milieu de l'étendue observée pour les valeurs de l'exemple donné en 1.9 est $(6 + 19)/2 = 12,5$.

NOTE Le milieu de l'étendue fournit une évaluation rapide et simple du milieu d'un petit jeu de données.

1.12 estimateur

$\hat{\theta}$

statistique (1.8) utilisée pour une **estimation** (1.36) du paramètre θ

NOTE 1 Un estimateur peut être la **moyenne d'échantillon** (1.15) destinée à estimer la **moyenne** (2.35) de population qui peut être notée μ . Pour une **distribution** (2.11) telle que la **loi normale** (2.50), l'estimateur «naturel» de la moyenne de la population μ est la moyenne d'échantillon.

NOTE 2 Pour estimer une propriété de population [par exemple le **mode** (2.27) d'une **distribution à une variable** (2.16)], un estimateur approprié peut être une fonction du ou des estimateurs du ou des paramètres d'une distribution ou une fonction complexe d'un **échantillon aléatoire** (1.6).

NOTE 3 Le terme «estimateur» est utilisé ici dans un sens large. Il comprend l'estimateur ponctuel d'un paramètre, ainsi que l'estimateur par intervalle éventuellement utilisé pour la prédiction (parfois appelé prédicteur). Il peut également comprendre des fonctions telles que l'estimateur noyau et d'autres statistiques à usage spécial. Des commentaires supplémentaires sont fournis dans les notes en 1.36.

1.13

sample median

$[(n+1)/2]$ th order statistic (1.9), if the sample size (see ISO 3534-2:2006, 1.2.26) n is odd; sum of the $(n/2)$ th and $[(n/2) + 1]$ th order statistics divided by 2, if the sample size n is even

EXAMPLE Continuing with the example of 1.9, the value of 8 is a realization of the sample median. In this case (even sample size of 10), the 5th and 6th values were 7 and 9, whose average equals 8. In practice, this would be reported as "the sample median is 8", although strictly speaking, the sample median is defined as a random variable.

NOTE 1 For a random sample (1.6) of sample size n whose random variables (2.10) are arranged in non-decreasing order from 1 to n , the sample median is the $(n+1)/2$ th random variable if the sample size is odd. If the sample size n is even, then the sample median is the average of the $(n/2)$ th and $(n+1)/2$ th random variables.

NOTE 2 Conceptually, it may seem impossible to conduct an ordering of random variables which have not yet been observed. Nevertheless, the structure for understanding order statistics can be established so that upon observation, the analysis may proceed. In practice, one obtains observed values and through sorting the values, one obtains realizations of the order statistics. These realizations can then be interpreted from the structure of order statistics from a random sample.

NOTE 3 The sample median provides an estimator of the middle of a distribution, with half of the sample to each side of it.

NOTE 4 In practice, the sample median is useful in providing an estimator that is insensitive to very extreme values in a data set. For example, median incomes and median housing prices are frequently reported as summary values.

1.13

médiane d'échantillon

$[(n+1)/2]$ ème statistique d'ordre (1.9), si l'effectif d'échantillon (ISO 3534-2:2006, 1.2.26) n est impair; somme de la $(n/2)$ ème et de la $[(n/2) + 1]$ ème statistique d'ordre divisée par 2, si l'effectif d'échantillon n est pair

EXEMPLE En reprenant l'exemple donné en 1.9, la valeur de 8 est une réalisation de la médiane d'échantillon. Dans ce cas (effectif d'échantillon pair de 10), lorsque les cinquième et sixième valeurs sont 7 et 9, leur moyenne est égale à 8. Dans la pratique, ce serait reporté comme «la médiane d'échantillon est 8», bien que strictement parlant, la médiane d'échantillon est définie par une variable aléatoire.

NOTE 1 Pour un échantillon aléatoire (1.6) d'effectif n dont les variables aléatoires (2.10) sont rangées par ordre non décroissant de 1 à n , la médiane d'échantillon est la $(n+1)/2$ ème variable aléatoire si l'effectif d'échantillon est impair. Si l'effectif d'échantillon n est pair, la médiane d'échantillon est alors la moyenne des $(n/2)$ ème et $(n+1)/2$ ème variables aléatoires.

NOTE 2 En théorie, il peut paraître impossible de ranger des variables aléatoires qui n'ont pas encore été observées. Cependant, la structure permettant de comprendre la statistique d'ordre peut être établie de manière à pouvoir réaliser l'analyse dès l'observation. Dans la pratique, on obtient des valeurs observées et en rangeant les valeurs, il est possible de parvenir à des réalisations de la statistique d'ordre. Ces réalisations peuvent ensuite être interprétées à partir de la structure de la statistique d'ordre d'un échantillon aléatoire.

NOTE 3 La médiane d'échantillon fournit un estimateur du centre d'une distribution avec la moitié de l'échantillon de chaque côté.

NOTE 4 Dans la pratique, l'utilité de la médiane d'échantillon est de fournir un estimateur insensible aux valeurs extrêmes d'un ensemble de données. Par exemple, les revenus médians et les prix médians des logements sont fréquemment utilisés comme des valeurs synthétiques.

1.14**sample moment of order k**

$$E(X^k)$$

sum of k^{th} power of **random variables** (2.10) in a **random sample** (1.6) divided by the number of observations in the **sample** (1.3)

NOTE 1 For a random sample of sample size n , i.e. $\{X_1, X_2, \dots, X_n\}$, the sample moment of order k , $E(X^k)$, is

$$\frac{1}{n} \sum_{i=1}^n X_i^k$$

NOTE 2 Furthermore, this concept can be described as the sample moment of order k about zero.

NOTE 3 The sample moment of order 1 will be seen in the next definition to be the **sample mean** (1.15).

NOTE 4 Although the definition is given for arbitrary k , commonly used instances in practice involve $k = 1$ [**sample mean** (1.15)], $k = 2$ [associated with the **sample variance** (1.16) and **sample standard deviation** (1.17)], $k = 3$ [related to **sample coefficient of skewness** (1.20)] and $k = 4$ [related to **sample coefficient of kurtosis** (1.21)].

NOTE 5 The “ E ” in $E(X^k)$ comes from the “expected value” or “expectation” of the random variable X .

1.14**moment d'échantillon d'ordre k**

$$E(X^k)$$

somme de la $k^{\text{ème}}$ puissance des **variables aléatoires** (2.10) d'un **échantillon aléatoire** (1.6), divisée par le nombre d'observations dans l'**échantillon** (1.3)

NOTE 1 Pour un échantillon aléatoire d'effectif n , c'est-à-dire $\{X_1, X_2, \dots, X_n\}$, le moment d'échantillon d'ordre k , $E(X^k)$, est:

$$\frac{1}{n} \sum_{i=1}^n X_i^k$$

NOTE 2 En outre, ce concept peut être décrit comme le moment d'échantillon d'ordre k par rapport à zéro.

NOTE 3 Le moment d'échantillon d'ordre 1 sera considéré dans la définition suivante comme la **moyenne d'échantillon** (1.15).

NOTE 4 Bien que la définition considère une valeur quelconque de k les exemples les plus couramment utilisés dans la pratique impliquent des valeurs de $k = 1$ [**moyenne d'échantillon** (1.15)], $k = 2$ [associée à la **variance d'échantillon** (1.16) et à l'**écart-type d'échantillon** (1.17)], $k = 3$ [relative au **coefficient d'asymétrie d'échantillon** (1.20)] et $k = 4$ [relative au **coefficient d'aplatissement d'échantillon** (1.21)].

NOTE 5 Le « E » dans $E(X^k)$ provient de la «valeur espérée» ou de «l'espérance mathématique» de la variable aléatoire X .

1.15

sample mean average

arithmetic mean

sum of **random variables** (2.10) in a **random sample** (1.6) divided by the number of terms in the sum

EXEMPLE Continuing with the example from 1.9, the realization of the sample mean is 9,7 as the sum of the observed values is 97 and the sample size is 10.

NOTE 1 Considered as a statistic, the sample mean is a function of random variables from a random sample in the sense given in Note 3 of 1.8. One must distinguish this estimator from the numerical value of the sample mean calculated from the **observed values** (1.4) in the random sample.

NOTE 2 The sample mean considered as a statistic is often used as an estimator for the population **mean** (2.35). A common synonym is arithmetic mean.

NOTE 3 For a random sample of sample size n , i.e. $\{X_1, X_2, \dots, X_n\}$, the sample mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

NOTE 4 The sample mean can be recognized as the sample moment of order 1.

NOTE 5 For sample size 2, the sample mean, the **sample median** (1.13) and **mid-range** (1.11) are the same.

1.15

moyenne d'échantillon moyenne

moyenne arithmétique

somme des **variables aléatoires** (2.10) dans un **échantillon aléatoire** (1.6) divisée par le nombre de termes de la somme

EXEMPLE En reprenant l'exemple donné en 1.9, la réalisation de la moyenne d'échantillon est 9,7 lorsque la somme des valeurs observées est égale à 97 et l'effectif d'échantillon est de 10.

NOTE 1 Considérée comme une statistique, la moyenne d'échantillon est une fonction des variables aléatoires d'un échantillon aléatoire au sens donné dans la Note 3 en 1.8. Il faut différencier cet estimateur de la valeur numérique de la moyenne d'échantillon calculée à partir des **valeurs observées** (1.4) dans l'échantillon aléatoire.

NOTE 2 La moyenne d'échantillon considérée comme une statistique est souvent utilisée comme un estimateur de la **moyenne** (2.35) d'une population. Un synonyme courant est la moyenne arithmétique.

NOTE 3 Pour un échantillon aléatoire d'effectif n , c'est-à-dire, $\{X_1, X_2, \dots, X_n\}$, la moyenne d'échantillon est:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

NOTE 4 La moyenne d'échantillon peut être considérée comme le moment d'échantillon d'ordre 1.

NOTE 5 Pour une taille d'échantillon de 2, la moyenne d'échantillon, la **médiane d'échantillon** (1.13) et le **milieu de l'étendue** (1.11) sont les mêmes.

1.16 sample variance

s^2

sum of squared deviations of **random variables** (2.10) in a **random sample** (1.6) from their **sample mean** (1.15) divided by the number of terms in the sum minus one

EXAMPLE Continuing with the numerical example of 1.9, the sample variance can be computed to be 17,57. The sum of squares about the observed sample mean is 158,10 and the sample size 10 minus 1 is 9, giving the appropriate denominator.

NOTE 1 Considered as a **statistic** (1.8), the sample variance s^2 is a function of random variables from a random sample. One has to distinguish this **estimator** (1.12) from the numerical value of the sample variance calculated from the **observed values** (1.4) in the random sample. This numerical value is called the empirical sample variance or the observed sample variance and is usually denoted by s^2 .

NOTE 2 For a random sample of sample size n , i.e. $\{X_1, X_2, \dots, X_n\}$ with sample mean $\bar{X}$ the sample variance is:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

NOTE 3 The sample variance is a statistic that is "almost" the average of the squared deviations of the **random variables** (2.10) from their sample mean (only "almost" since $n - 1$ is used rather than n in the denominator). Using $n - 1$ provides an **unbiased estimator** (1.34) of the population **variance** (2.36).

NOTE 4 The quantity $n - 1$ is known as the **degrees of freedom** (2.54).

NOTE 5 The sample variance can be recognized to be the 2nd sample moment of the **standardized sample random variables** (1.19).

1.16 variance d'échantillon

s^2

somme des carrés des écarts des **variables aléatoires** (2.10) d'un **échantillon aléatoire** (1.6) par rapport à leur **moyenne d'échantillon** (1.15), divisée par le nombre de termes de la somme moins un

EXEMPLE En reprenant l'exemple numérique donné en 1.9, la variance d'échantillon peut être calculée comme égale à 17,57. La somme des carrés par rapport à la moyenne d'échantillon observée est égale à 158,10 et l'effectif d'échantillon 10 moins 1 est de 9, donnant le dénominateur approprié.

NOTE 1 Considérée comme une **statistique** (1.8), la variance d'échantillon s^2 est une fonction des variables aléatoires d'un échantillon aléatoire. Il faut différencier cet **estimateur** (1.12) de la valeur numérique de la variance d'échantillon calculée à partir des **valeurs observées** (1.4) dans l'échantillon aléatoire. Cette valeur numérique est appelée la variance d'échantillon empirique ou la variance d'échantillon observée et est généralement notée s^2 .

NOTE 2 Pour un échantillon aléatoire d'effectif n , c'est-à-dire $(X_1, X_2, \dots, X_n)$ avec pour moyenne d'échantillon $\bar{X}$, la variance d'échantillon est:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

NOTE 3 La variance d'échantillon est une statistique qui est «presque» la moyenne des carrés des écarts des **variables aléatoires** (2.10) par rapport à leur moyenne d'échantillon (étant donné que $n - 1$ est utilisé plutôt que n dans le dénominateur). L'utilisation de $n - 1$ fournit un **estimateur sans biais** (1.34) de la **variance** (2.36) de la population.

NOTE 4 La grandeur $n - 1$ est appelée **degrés de liberté** (2.54).

NOTE 5 La variance d'échantillon peut être reconnue comme étant le deuxième moment de distribution d'échantillonnage des **variables aléatoires centrées réduites d'échantillon** (1.19).

1.17 sample standard deviation

S
non-negative square root of the **sample variance** (1.16)

EXAMPLE Continuing with the numerical example of 1.9, the observed sample standard deviation is 4,192 since the observed sample variance is 17,57.

NOTE 1 In practice, the **sample standard deviation** is used to estimate the **standard deviation** (2.37). Here again, it should be emphasized that S is also a **random variable** (2.10) and not a realization from a **random sample** (1.6).

NOTE 2 The sample standard deviation is a measure of the dispersion of a **distribution** (2.11).

1.18 sample coefficient of variation sample standard deviation (1.17) divided by the sample mean (1.15)

NOTE As with the **coefficient of variation** (2.38), the utility of this statistic is limited to populations that are positive valued. The sample coefficient of variation is commonly reported as a percentage. It is particularly applicable where variation increases in proportion to the mean.

1.19 standardized sample random variable random variable (2.10) minus its sample mean (1.15) divided by the sample standard deviation (1.17)

EXAMPLE For the example of 1.9, the observed sample mean is 9,7 and the observed sample standard deviation is 4,192. Hence, the observed standardized random variables (to two decimal places) are:

−0,17; 0,79; −0,64; −0,88; 0,79; −0,64; 2,22; −0,88;
0,07; −0,64.

NOTE 1 The standardized sample random variable is distinguished from its theoretical counterpart **standardized random variable** (2.33). The intent of standardizing is to transform random variables to have zero means and unit standard deviations, for ease in interpretation and comparison.

NOTE 2 Standardized observed values have an observed mean of zero and an observed standard deviation of 1.

1.17 écart-type d'échantillon

S
racine carrée de la **variance d'échantillon** (1.16)

EXEMPLE En reprenant l'exemple numérique donné en 1.9, l'écart-type d'échantillon observé est égal à 4,192 étant donné que la variance d'échantillon observée est égale à 17,57.

NOTE 1 Dans la pratique, l'écart-type d'échantillon est utilisé pour estimer l'écart-type (2.37). De nouveau, il convient de noter que S est une **variable aléatoire** (2.10) et non une réalisation d'un **échantillon aléatoire** (1.6).

NOTE 2 L'écart-type d'échantillon est une mesure de la dispersion d'une **distribution** (2.11).

1.18 coefficient de variation d'échantillon écart-type d'échantillon (1.17) divisé par la moyenne d'échantillon (1.15)

NOTE À l'instar du **coefficient de variation** (2.38), l'utilité de cette statistique est limitée aux populations à valeurs positives. Le coefficient de variation d'échantillon est couramment exprimé en pourcentage. Il s'applique en particulier lorsque la variation croît proportionnellement à la moyenne.

1.19 variable aléatoire centrée réduite d'échantillon variable aléatoire (2.10) moins sa moyenne d'échantillon (1.15) divisée par l'écart-type d'échantillon (1.17)

EXEMPLE Pour l'exemple donné en 1.9, la moyenne d'échantillon observée est égale à 9,7 et l'écart-type d'échantillon observé est de 4,192. Ainsi, les variables aléatoires centrées réduites observées (à deux décimales) sont:

−0,17; 0,79; −0,64; −0,88; 0,79; −0,64; 2,22; −0,88;
0,07; −0,64.

NOTE 1 La variable aléatoire centrée réduite d'échantillon est différenciée de sa contrepartie théorique, la **variable aléatoire centrée réduite** (2.33). Le centrage-réduction a pour objet de transformer des variables aléatoires de manière à obtenir des moyennes égales à zéro et des écarts-types unitaires et faciliter l'interprétation et la comparaison.

NOTE 2 Les valeurs observées centrées réduites ont une moyenne observée de zéro et un écart-type observé de 1.

1.20

sample coefficient of skewness

arithmetic mean of the third power of the **standardized sample random variables** (1.19) from a **random sample** (1.6)

EXAMPLE Continuing with the example from 1.9, the observed sample coefficient of skewness can be computed to be 0,971 88. For a sample size such as 10 in this example, the sample coefficient of skewness is highly variable, so it must be used with caution. Using the alternative formula in Note 1, the computed value is 1,349 83.

NOTE 1 The formula corresponding to the definition is

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^3$$

Some statistical packages use the following formula for the sample coefficient of skewness to correct for **bias** (1.33):

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n Z_i^3$$

where

$$Z_i = \frac{X_i - \bar{X}}{S}$$

For a large sample size, the distinction between the two estimates is negligible. The ratio of the unbiased to the biased estimate is 1,389 for $n = 10$, 1,031 for $n = 100$ and 1,003 for $n = 1\,000$.

NOTE 2 Skewness refers to lack of symmetry. Values of this statistic close to zero suggest that the underlying distribution is approximately symmetric, whereas non-zero values would likely correspond to a distribution having occasional extreme values on one side of the centre of the distribution. Skewed data would also be reflected in values of the **sample mean** (1.15) and **sample median** (1.13) that are dissimilar. Positively skewed (right-skewed) data indicate the possible presence of a few extreme, large observations. Similarly, negatively skewed (left-skewed) data indicate the possible presence of a few extreme, small observations.

NOTE 3 The sample coefficient of skewness can be recognized to be the 3rd sample moment of the **standardized sample random variables** (1.19).

1.20

coefficient d'asymétrie d'échantillon

moyenne arithmétique de la troisième puissance des **variables aléatoires centrées réduites d'échantillon** (1.19) pour un **échantillon aléatoire** (1.6)

EXEMPLE En reprenant l'exemple donné en 1.9, le coefficient d'asymétrie d'échantillon observé peut être calculé comme égal à 0,971 88. Pour l'effectif d'échantillon de 10 le coefficient d'asymétrie d'échantillon est fortement variable et est donc à utiliser avec prudence. En utilisant la formule alternative de la Note 1, la valeur calculée est 1,349 83.

NOTE 1 La formule correspondant à la définition est la suivante:

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^3$$

Certains progiciels statistiques utilisent la formule suivante pour corriger le **biais** (1.33) dans le coefficient d'asymétrie d'échantillon:

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n Z_i^3$$

où

$$Z_i = \frac{X_i - \bar{X}}{S}$$

Pour une grande taille d'échantillon, la différence entre les deux estimations est négligeable. Le rapport entre l'estimation avec biais et celle sans biais est 1,389 pour $n = 10$, 1,031 pour $n = 100$ et 1,003 pour $n = 1\,000$.

NOTE 2 L'asymétrie désigne le manque de symétrie. Des valeurs de cette statistique proches de zéro suggèrent que la distribution sous-jacente est approximativement symétrique, alors que des valeurs non nulles pourraient correspondre à une distribution présentant des valeurs extrêmes particulières situées d'un côté du centre de la distribution. Les données asymétriques pourraient également correspondre à des valeurs de la **moyenne d'échantillon** (1.15) et de la **médiane d'échantillon** (1.13) qui sont dissemblables. Les données positivement asymétriques (asymétriques vers la droite) indiquent l'éventuelle présence de quelques fortes observations extrêmes. Dans le même ordre d'idée, des données négativement asymétriques (asymétriques vers la gauche) indiquent l'éventuelle présence de quelques faibles observations extrêmes.

NOTE 3 Le coefficient d'asymétrie d'échantillon peut être considéré comme le troisième moment d'échantillon des **variables aléatoires centrées réduites d'échantillon** (1.19).

1.21

sample coefficient of kurtosis

arithmetic mean of the fourth power of the **standardized sample random variables** (1.19) from a **random sample** (1.6)

EXEMPLE Continuing with the example from 1.9, the observed sample coefficient of kurtosis can be computed to be 2,674 19. For such a sample size as 10 in this example, the sample coefficient of kurtosis is highly variable, so it must be used with caution. Statistical packages use various adjustments in computing the sample coefficient of kurtosis (see Note 3 of 2.40). Using the alternate formula given in Note 1, the computed value is 0,436 05. The two values 2,674 19 and 0,436 05 are not comparable directly. To do so, take $2,674\ 19 - 3$ (to relate to the kurtosis of the normal distribution which is 3) which equals $-0,325\ 81$ which now can be appropriately compared to 0,436 05.

NOTE 1 The formula corresponding to the definition is:

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^4$$

Some statistical packages use the following formula for the sample coefficient of kurtosis to correct for **bias** (1.33) and to indicate the deviation from the kurtosis of the normal distribution (which equals 3):

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n Z_i^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where

$$Z_i = \frac{X_i - \bar{X}}{S}$$

The second term in the expression is approximately 3 for large n . Sometimes the kurtosis is reported as a value as defined in 2.40 minus 3 to emphasize comparisons to the normal distribution. Obviously, a practitioner needs to be aware of the adjustments, if any, in statistical package computations.

NOTE 2 Kurtosis refers to the heaviness of the tails of a (unimodal) distribution. For the **normal distribution** (2.50), the sample coefficient of kurtosis is approximately 3, subject to sampling variability. In practice, the kurtosis of the normal distribution provides a benchmark or baseline value. **Distributions** (2.11) with values smaller than 3 have lighter tails than the normal distribution; distributions with values larger than 3 have heavier tails than the normal distribution.

1.21

coefficient d'aplatissement d'échantillon

moyenne arithmétique de la quatrième puissance des **variables aléatoires centrées réduites d'échantillon** (1.19) pour un **échantillon aléatoire** (1.6)

EXEMPLE En reprenant l'exemple donné en 1.9, le coefficient d'aplatissement d'échantillon observé peut être calculé comme égal à 2,674 19. Pour l'effectif d'échantillon de 10 le coefficient d'aplatissement d'échantillon est fortement variable et est donc à utiliser avec prudence. Les progiciels statistiques utilisent différents ajustements pour calculer le coefficient d'aplatissement (voir Note 3 en 2.40). En utilisant la formule alternative donnée dans la Note 1, la valeur calculée est 0,436 05. Les deux valeurs 2,674 19 et 0,436 05 ne sont pas directement comparables. Pour ce faire, il convient de prendre $2,674\ 19 - 3$ (pour se reporter à l'aplatissement de la loi normale, qui est de 3), qui est égal à $-0,325\ 81$ qui peut maintenant être comparé correctement à 0,436 05.

NOTE 1 La formule correspondant à la définition est la suivante:

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^4$$

Certains progiciels statistiques utilisent la formule suivante dans le calcul du coefficient d'asymétrie d'échantillon afin de corriger le **biais** (1.33) et indiquer l'écart de la loi normale par rapport à l'aplatissement (qui est égal à 3):

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n Z_i^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

où

$$Z_i = \frac{X_i - \bar{X}}{S}$$

Le second terme de l'expression est approximativement 3 pour un grand n . L'aplatissement est parfois reporté comme une valeur telle que définie en 2.40 moins 3 pour améliorer la comparaison avec la loi normale. Évidemment, un praticien a besoin d'être informé des ajustements, le cas échéant, dans les calculs informatiques de statistique.

NOTE 2 L'aplatissement fait référence à la lourdeur des queues d'une distribution unimodale. Pour la **loi normale** (2.50), le coefficient d'aplatissement d'échantillon est d'environ 3 sous réserve de la variabilité d'échantillonnage. Dans la pratique, l'aplatissement de la loi normale fournit une valeur repère ou de référence. Les **distributions** (2.11) à valeurs inférieures à 3 ont des queues plus légères que la loi normale; les distributions à valeurs supérieures à 3 ont des queues plus lourdes que la loi normale.

NOTE 3 For observed values of kurtosis much larger than 3, the possibility exists that the underlying distribution has genuinely heavier tails than the normal distribution. A sample could be contaminated by observations from another source or from a coding error.

NOTE 4 The sample coefficient of kurtosis can be recognized to be the 4th sample moment of the standardized sample random variables.

1.22

sample covariance

S_{XY}

sum of products of deviations of pairs of **random variables** (2.10) in a **random sample** (1.6) from their **sample means** (1.15) divided by the number of terms in the sum minus one

EXAMPLE 1 Consider the following numerical illustration using 10 observed 3-tuples (triplets) of values. For this example, consider only x and y .

Table 1 — Results for Example 1

i	1	2	3	4	5	6	7	8	9	10
x	38	41	24	60	41	51	58	50	65	33
y	73	74	43	107	65	73	99	72	100	48
z	34	31	40	28	35	28	32	27	27	31

The observed sample mean for X is 46,1 and for Y is 75,4. The sample covariance is equal to

$$[(38 - 46,1) \times (73 - 75,4) + (41 - 46,1) \times (74 - 75,4) + \dots + (33 - 46,1) \times (48 - 75,4)]/9 = 257,178$$

EXAMPLE 2 In the table of the previous example, consider only y and z . The observed sample mean for Z is 31,3. The sample covariance is equal to

$$[(73 - 75,4) \times (34 - 31,3) + (74 - 75,4) \times (31 - 31,3) + \dots + (48 - 75,4) \times (31 - 31,3)]/9 = -54,356$$

NOTE 3 Pour des valeurs d'aplatissement observées beaucoup plus élevées que 3, il est possible que la distribution sous-jacente présente des queues largement plus lourdes que la loi normale. Un échantillon peut être contaminé par des observations provenant d'une autre source ou d'une erreur de codage.

NOTE 4 Le coefficient d'aplatissement d'échantillon peut être considéré comme le quatrième moment d'échantillon des variables aléatoires centrées réduites d'échantillon.

1.22

covariance d'échantillon

S_{XY}

somme des produits des écarts des couples de **variables aléatoires** (2.10) d'un **échantillon aléatoire** (1.6) par rapport à leur **moyenne d'échantillon** (1.15), divisée par le nombre de termes de la somme moins un

EXEMPLE 1 Soit la représentation numérique suivante utilisant 10 valeurs observées à 3-tuples (triplets). Dans cet exemple ne sont considérés que x et y .

Tableau 1 — Résultats pour l'Exemple 1

i	1	2	3	4	5	6	7	8	9	10
x	38	41	24	60	41	51	58	50	65	33
y	73	74	43	107	65	73	99	72	100	48
z	34	31	40	28	35	28	32	27	27	31

La moyenne d'échantillon observée est de 46,1 pour X et de 75,4 pour Y . La covariance d'échantillon est égale à

$$[(38 - 46,1) \times (73 - 75,4) + (41 - 46,1) \times (74 - 75,4) + \dots + (33 - 46,1) \times (48 - 75,4)]/9 = 257,178$$

EXEMPLE 2 Dans le tableau de l'exemple précédent, ne considérer que y et z . La moyenne d'échantillon observée pour Z est de 31,3. La covariance d'échantillon est égale à

$$[(73 - 75,4) \times (34 - 31,3) + (74 - 75,4) \times (31 - 31,3) + \dots + (48 - 75,4) \times (31 - 31,3)]/9 = -54,356$$

NOTE 1 Considered as a **statistic** (1.8) the sample covariance is a function of pairs of **random variables** $[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ from a random sample of size n in the sense given in Note 3 of 1.6. This **estimator** (1.12) needs to be distinguished from the numerical value of the sample covariance calculated from the observed pairs of values of the **sampling units** (1.2) $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ in the random sample. This numerical value is called the empirical sample covariance or the observed sample covariance.

NOTE 2 The sample covariance S_{XY} is given as:

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

NOTE 3 Using $n-1$ provides an **unbiased estimator** (1.34) of the population **covariance** (2.43).

NOTE 4 The example in Table 1 consists of three variables whereas the definition refers to a pair of variables. In practice, it is common to encounter situations with multiple variables.

NOTE 1 Considérée comme une **statistique** (1.8), la covariance d'échantillon est une fonction des couples des **variables aléatoires** $[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ d'un échantillon aléatoire d'effectif n au sens donné dans la Note 3 en 1.6. Cet **estimateur** (1.12) doit être différencié de la valeur numérique de la covariance d'échantillon calculée à partir des couples observés des valeurs des **unités d'échantillonnage** (1.2) $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ dans l'échantillon aléatoire. Cette valeur numérique est appelée covariance d'échantillon empirique ou covariance d'échantillon observée.

NOTE 2 La covariance d'échantillon S_{XY} est donnée par:

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

NOTE 3 L'utilisation de $n-1$ fournit un **estimateur sans biais** (1.34) de la **covariance** (2.43) de la population.

NOTE 4 L'exemple dans le Tableau 1 est constitué de trois variables alors que la définition fait référence à un couple de variables. En pratique, il est fréquent de rencontrer des situations avec des variables multiples.

1.23 sample correlation coefficient

r_{xy}
sample covariance (1.22) divided by the product of the corresponding **sample standard deviations** (1.17)

EXAMPLE 1 Continuing with Example 1 of 1.22, the observed standard deviation is 12,948 for X and 21,329 for Y . Hence, the observed sample correlation coefficient (for X and Y) is given by:

$$257,178/(12,948 \times 21,329) = 0,931\ 2$$

EXAMPLE 2 Continuing with Example 2 of 1.22, the observed standard deviation is 21,329 for Y and 4,165 for Z . Hence, the observed sample correlation coefficient (for Y and Z) is given by:

$$-54,356/(21,329 \times 4,165) = -0,612$$

NOTE 1 Notationally, the sample correlation coefficient is computed as:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

This expression is equivalent to the ratio of the sample covariance to the square root of the product of the sample variances. Sometimes the symbol r_{xy} is used to denote the sample correlation coefficient. The observed sample correlation coefficient is based on realizations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

NOTE 2 The observed sample correlation coefficient can take on values in $[-1, 1]$, with values near 1 indicating strong positive correlation and values near -1 indicating strong negative correlation. The sample correlation coefficient indicates the degree of linear relationship between two variables, with values near -1 or 1 indicating a strong linear relationship while values near 0 indicate a weak linear relationship.

1.23 coefficient de corrélation d'échantillon

r_{xy}
covariance d'échantillon (1.22) divisée par le produit des **écarts-types d'échantillon** (1.17) correspondants

EXEMPLE 1 En reprenant l'Exemple 1 donné en 1.22, l'écart-type observé est de 12,948 pour X et de 21,329 pour Y . Ainsi, le coefficient de corrélation d'échantillon observé (pour X et Y) est donné par:

$$257,178/(12,948 \times 21,329) = 0,931\ 2$$

EXEMPLE 2 En reprenant l'Exemple 2 donné en 1.22, l'écart-type observé est de 21,329 pour Y et de 4,165 pour Z . Ainsi, le coefficient de corrélation d'échantillon observé (pour Y et Z) est donné par:

$$-54,356/(21,329 \times 4,165) = -0,612$$

NOTE 1 En termes de notation, le coefficient de corrélation d'échantillon est obtenu de la manière suivante:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

L'expression est équivalente au rapport de la covariance d'échantillon à la racine carrée du produit des variances d'échantillon. Parfois le symbole r_{xy} est utilisé pour désigner le coefficient de corrélation d'échantillon. Le coefficient de corrélation d'échantillon observé est fondé sur les réalisations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

NOTE 2 Le coefficient de corrélation d'échantillon observé peut prendre des valeurs dans $[-1, 1]$, les valeurs proches de 1 indiquant une forte corrélation positive et les valeurs proches de -1 indiquant une forte corrélation négative. Le coefficient de corrélation d'échantillon renseigne sur le degré de dépendance linéaire entre deux variables, les valeurs proches de -1 ou de 1 indiquant une forte dépendance linéaire alors que les valeurs proches de 0 indiquent une faible dépendance linéaire.

1.24 standard error

 $\sigma_{\hat{\theta}}$

standard deviation (2.37) of an **estimator** (1.12) $\hat{\theta}$

EXAMPLE If the **sample mean** (1.15) is the estimator of the population **mean** (2.35) and the standard deviation of a single **random variable** (2.10) is σ , then the standard error of the sample mean is $\sigma/\sqrt{n}$ where n is the number of observations in the sample. An estimator of the standard error is $S/\sqrt{n}$ where S is the **sample standard deviation** (1.17).

NOTE 1 In practice, the standard error provides a natural estimate of the standard deviation of an estimator.

NOTE 2 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”. Commonly, in practice, standard error is implicitly referring to the standard deviation of the sample mean. The notation for the standard error of the sample mean is $\sigma_{\bar{X}}$.

1.25 interval estimator

interval, bounded by an upper limit **statistic** (1.8) and a lower limit statistic

NOTE 1 One of the end points could be $+\infty$, $-\infty$ or a natural limit of the value of a **parameter**. For example, 0 is a natural lower limit for an **interval estimator** of the population **variance** (2.36). In such cases, the intervals are commonly referred to as one-sided intervals.

NOTE 2 An interval estimator can be given in conjunction with **parameter** (2.9) **estimation** (1.36). The interval estimator is presumed to contain a parameter on a stated proportion of occasions, under conditions of repeated sampling, or in some other probabilistic sense.

NOTE 3 Three common types of interval estimators include **confidence intervals** (1.28) for parameter(s), **prediction intervals** (1.30) for future observations, and **statistical tolerance intervals** (1.26) on the proportion of a **distribution** (2.11) contained.

1.24 erreur type

 $\sigma_{\hat{\theta}}$

écart-type (2.37) d'un **estimateur** (1.12) $\hat{\theta}$

EXAMPLE Si la **moyenne d'échantillon** (1.15) est l'estimateur de la population **moyenne** (2.35) et l'écart-type d'une **variable aléatoire** unique (2.10) est σ , alors l'erreur type de la moyenne d'échantillon est $\sigma/\sqrt{n}$ où n est le nombre d'observations dans l'échantillon. Un estimateur de l'erreur type est $S/\sqrt{n}$ où S est l'**écart-type de l'échantillon** (1.17).

NOTE 1 En pratique, l'erreur type fournit une estimation naturelle de l'écart-type d'un estimateur.

NOTE 2 Il n'y a pas de terme (approprié) complémentaire pour erreur «non type». L'erreur type peut être vue comme une abréviation de l'expression écart-type d'un estimateur. Couramment, en pratique, l'erreur type fait simplement référence à l'écart-type de la moyenne d'échantillon. La notation pour l'erreur type de la moyenne d'échantillon est $\sigma_{\bar{X}}$.

1.25 estimateur par intervalle

intervalle, borné par une **statistique** (1.8) de borne supérieure et une statistique de borne inférieure

NOTE 1 L'une des bornes peut être $+\infty$, $-\infty$ ou une borne naturelle de la valeur d'un **paramètre**. Par exemple, 0 est une borne inférieure naturelle d'un **estimateur par intervalle** de la **variance** (2.36) de la population. Dans ce cas, les intervalles sont couramment désignés comme intervalles unilatéraux.

NOTE 2 Un estimateur par intervalle peut être donné dans le cadre d'une **estimation** (1.36) de **paramètre** (2.9). Un estimateur par intervalle est supposé inclure ce paramètre dans une proportion déclarée d'épreuves, dans des conditions d'échantillonnage répété ou dans un autre sens probabiliste.

NOTE 3 Les trois types courants d'estimateurs par intervalle sont les **intervalles de confiance** (1.28) pour un ou des paramètres, les **intervalles de prédiction** (1.30) pour des observations futures et les **intervalles statistiques de dispersion** (1.26) pour la proportion contenue d'une **distribution** (2.11) contenue.

1.26 statistical tolerance interval

interval determined from a **random sample** (1.6) in such a way that one may have a specified level of confidence that the interval covers at least a specified proportion of the sampled **population** (1.1)

NOTE The confidence in this context is the long-run proportion of intervals constructed in this manner that will include at least the specified proportion of the sampled population.

1.27 statistical tolerance limit statistic (1.8) representing an end point of a statistical tolerance interval (1.26)

NOTE Statistical tolerance intervals may be either

- one-sided (with one of its limits fixed at the natural boundary of the random variable), in which case they have either an upper or a lower statistical tolerance limit, or
- two-sided, in which case they have both.

A natural boundary of the random variable may provide a limit for a one-sided limit.

1.28 confidence interval interval estimator (1.25) (T_0, T_1) for the parameter (2.9) θ with the statistics (1.8) T_0 and T_1 as interval limits and for which it holds that $P[T_0 < \theta < T_1] \geq 1 - \alpha$

NOTE 1 The confidence reflects the proportion of cases that the confidence interval would contain the true parameter value in a long series of repeated **random samples** (1.6) under identical conditions. A confidence interval does not reflect the **probability** (2.5) that the observed interval contains the true value of the parameter (it either does or does not contain it).

NOTE 2 Associated with this confidence interval is the attendant performance characteristic $100(1 - \alpha) \%$, where α is generally a small number. The performance characteristic, which is called the confidence coefficient or confidence level, is often 95 % or 99 %. The inequality $P[T_0 < \theta < T_1] \geq 1 - \alpha$ holds for any specific but unknown population value of θ .

1.26 intervalle statistique de dispersion intervalle déterminé à partir d'un échantillon aléatoire (1.6) de sorte qu'il existe un niveau spécifié de confiance que l'intervalle couvre au moins une proportion donnée de la population (1.1) échantillonnée

NOTE La confiance dans ce contexte est la proportion à long terme des intervalles construits de cette manière qui comprendront au moins la proportion donnée de la population échantillonnée.

1.27 limite statistique de dispersion statistique (1.8) représentant une borne d'un intervalle statistique de dispersion (1.26)

NOTE Les intervalles statistiques de dispersion peuvent être soit

- unilatéraux (avec l'une des limites fixée à la limite naturelle de la variable aléatoire), auquel cas ils ont une limite statistique de dispersion inférieure ou supérieure, soit
- bilatéraux, auquel cas ils ont une limite inférieure et supérieure.

Une limite naturelle de la variable aléatoire peut fournir une limite dans le cas d'une limite unilatérale.

1.28 intervalle de confiance estimateur par intervalle (1.25) (T_0, T_1) pour le paramètre (2.9) θ avec les statistiques (1.8) T_0 et T_1 comme limites d'intervalle et pour lequel il est stipulé que $P[T_0 < \theta < T_1] \geq 1 - \alpha$

NOTE 1 La confiance reflète la proportion de cas où l'intervalle de confiance contiendrait la valeur vraie du paramètre dans une longue série d'échantillons aléatoires (1.6) répétés dans des conditions identiques. Un intervalle de confiance ne reflète pas la **probabilité** (2.5) que l'intervalle observé contienne la valeur vraie du paramètre (qu'il la contienne ou non).

NOTE 2 Cet intervalle de confiance est associé à la caractéristique de performance concomitante $100(1 - \alpha) \%$, où α est généralement un petit nombre. La caractéristique de performance, appelée niveau de confiance, est souvent de 95 % à 99 %. L'inégalité $P[T_0 < \theta < T_1] \geq 1 - \alpha$ s'applique à toute valeur de population spécifique mais inconnue de θ .

1.29

one-sided confidence interval

confidence interval (1.28) with one of its end points fixed at $+\infty$, $-\infty$, or a natural fixed boundary

NOTE 1 Definition 1.28 applies with either T_0 set at $-\infty$ or T_1 set at $+\infty$. One-sided confidence intervals arise in situations where interest focuses strictly on one direction. For example, in audio volume testing for safety concerns in cellular telephones, an upper confidence limit would be of interest indicating an upper bound for the volume produced under presumed safe conditions. For structural mechanical testing, a lower confidence limit on the force at which a device fails would be of interest.

NOTE 2 Another instance of one-sided confidence intervals occurs in situations where a parameter has a natural boundary such as zero. For a **Poisson distribution** (2.47) involved in modelling customer complaints, zero is a lower bound. As another example, a confidence interval for the reliability of an electronic component could be (0,98, 1), where 1 is the natural upper boundary limit.

1.30

prediction interval

range of values of a variable, derived from a **random sample** (1.6) of values from a continuous population, within which it can be asserted with a given confidence that no fewer than a given number of values in a further random sample from the same **population** (1.1) will fall

NOTE Commonly, interest focuses on a single further observation arising from the same situation as the observations which are the basis of the prediction interval. Another practical context is regression analysis in which a prediction interval is constructed for a spectrum of independent values.

1.31

estimate

observed value (1.4) of an **estimator** (1.12)

NOTE Estimate refers to a numerical value obtained from observed values. With respect to **estimation** (1.36) of a **parameter** (2.9) from an hypothesized **probability distribution** (2.11), estimator refers to the **statistic** (1.8) intended to estimate the 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 is taking place.

1.29

intervalle de confiance unilatéral

intervalle de confiance (1.28) dont l'une des bornes est fixée à $+\infty$, $-\infty$ ou une limite fixée naturellement

NOTE 1 La définition 1.28 s'applique tant à T_0 fixé à $-\infty$ qu'à T_1 fixé à $+\infty$. Les intervalles de confiance unilatéraux s'appliquent lorsque l'intérêt ne porte que sur une seule direction. Par exemple, pour des essais de volume sonore de téléphones cellulaires réalisés pour des raisons de sécurité, l'intérêt d'une borne de confiance supérieure serait d'indiquer une borne supérieure pour le volume généré dans des conditions de sécurité présumées. Pour des essais mécaniques de structure, l'intérêt d'une borne de confiance inférieure serait de déterminer la force à laquelle un dispositif présente une défaillance.

NOTE 2 Un autre exemple d'intervalles de confiance unilatéraux est celui appliqué lorsqu'un paramètre a une borne naturelle égale à zéro. Pour une **loi de Poisson** (2.47) utilisée pour modéliser des plaintes d'abonnés, zéro est une borne inférieure. Pour autre exemple, un intervalle de confiance pour la fiabilité d'un composant électronique pourrait être (0,98, 1), où 1 est la limite supérieure naturelle.

1.30

intervalle de prédiction

étendue de valeurs d'une variable, dérivée d'un **échantillon aléatoire** (1.6) de valeurs d'une population continue, à l'intérieur de laquelle il peut être assuré avec une confiance donnée que pas moins d'un nombre donné de valeurs d'un échantillon aléatoire ultérieur de la même **population** (1.1) ne tombera

NOTE Généralement, l'intérêt se porte sur une seule observation ultérieure provenant de la même situation que les observations qui sont la base de l'intervalle de prédiction. Un autre contexte pratique est l'analyse de régression dans laquelle un intervalle de prédiction est construit à partir d'un spectre de valeurs indépendantes.

1.31

estimation (résultat)

valeur observée (1.4) d'un **estimateur** (1.12)

NOTE L'estimation fait référence à une valeur numérique obtenue à partir de valeurs observées. S'agissant de l'**estimation** (1.36) d'un **paramètre** (2.9) sur la base d'une **loi de probabilité** (2.11) supposée, l'estimateur fait référence à la **statistique** (1.8) destinée à estimer le paramètre et l'estimation fait référence au résultat obtenu avec des valeurs observées. L'estimation est parfois associée à l'adjectif «ponctuelle» pour souligner qu'une seule valeur est produite plutôt qu'un intervalle de valeurs. De même, l'estimation est associée à l'expression «par intervalle» lorsqu'il s'agit de réaliser une estimation par intervalle.

1.32**error of estimation**

estimate (1.31) minus the **parameter** (2.9) or population property that it is intended to estimate

NOTE 1 Population property may be a function of the parameter or parameters or another quantity related to the **probability distribution** (2.11).

NOTE 2 Estimator error could involve contributions due to sampling, measurement uncertainty, rounding, or other sources. In effect, estimator error represents the bottom line performance of interest to practitioners. Determining the primary contributors to estimator error is a critical element in quality improvement efforts.

1.32**erreur d'estimation**

estimation (1.31) moins le **paramètre** (2.9) ou la propriété de la population à estimer

NOTE 1 La propriété de la population peut être une fonction du paramètre ou des paramètres ou une autre quantité relative à la **loi de probabilité** (2.11).

NOTE 2 L'erreur d'estimation peut être due à l'une des causes suivantes: échantillonnage, incertitude de mesure, arrondi ou autres sources d'erreur. En pratique, l'erreur d'estimation représente pour les praticiens la performance de base à considérer. La détermination des contributions principales à l'erreur d'estimation constitue un élément critique dans les efforts d'amélioration de la qualité.

1.33**bias**

expectation (2.12) of **error of estimation** (1.32)

NOTE 1 This definition differs from ISO 3534-2:2006 (3.3.2) and VIM:1993 (5.25 and 5.28). Here bias is used in a generic sense as indicated in Note 1 in 1.34.

NOTE 2 The existence of bias can lead to unfortunate consequences in practice. For example, underestimation of the strength of materials due to bias could lead to unexpected failures of a device. In survey sampling, bias could lead to incorrect decisions from a political poll.

1.33**biais**

espérance mathématique (2.12) d'une **erreur d'estimation** (1.32)

NOTE 1 La présente définition diffère de celles données dans l'ISO 3534-2:2006 (3.3.2) et le VIM:1993 (5.25 et 5.28). Le biais est utilisé ici dans un sens générique tel qu'indiqué dans la Note 1 en 1.34.

NOTE 2 L'existence d'un biais peut avoir des conséquences fâcheuses dans la pratique. Par exemple, le fait de sous-estimer la résistance de matériaux à cause d'un biais peut avoir comme conséquence des défaillances inattendues au niveau d'un dispositif. Dans l'échantillonnage d'enquête, le biais peut conduire à des décisions incorrectes prises sur la base d'un sondage politique.

1.34**unbiased estimator**

estimator (1.12) having **bias** (1.33) equal to zero

EXAMPLE 1 For a **random sample** (1.6) of n independent **random variables** (2.10), each with the same **normal distribution** (2.50) with **mean** (2.35) μ and **standard deviation** (2.37) σ , the **sample mean** $\bar{X}$ (1.15) and the **sample variance** (1.16) S^2 are unbiased estimators for the mean μ and the **variance** (2.36) σ^2 , respectively.

EXAMPLE 2 As is mentioned in Note 1 to 1.37 the **maximum likelihood estimator** (1.35) of the variance σ^2 uses the denominator n instead of $n - 1$ and thus is a biased estimator. In applications, the **sample standard deviation** (1.17) receives considerable use but it is important to note that the square root of the sample variance using $n - 1$ is a biased estimator of the population **standard deviation** (2.37).

EXAMPLE 3 For a random sample of n independent pairs of random variables, each pair with the same **bivariate normal distribution** (2.65) with **covariance** (2.43) equal to $\rho\sigma_{XY}$, the **sample covariance** (1.22) is an unbiased estimator for population covariance. The maximum likelihood estimator uses n instead of $n - 1$ in the denominator and thus is biased.

NOTE 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 parameters. The definition given here is of a statistical nature.

In everyday usage, practitioners try to avoid introducing bias into a study by ensuring, for example, that the random sample is representative of the population of interest.

1.34**estimateur sans biais**

estimateur (1.12) dont le **biais** (1.33) est égal à zéro

EXAMPLE 1 Pour un **échantillon aléatoire** (1.6) de n **variables aléatoires** (2.10) indépendantes ayant chacune la même **loi normale** (2.50) avec **moyenne** (2.35) μ et **écart-type** (2.37) σ , la **moyenne d'échantillon** (1.15) $\bar{X}$ et la **variance d'échantillon** (1.16) S^2 sont des estimateurs sans biais respectivement pour la moyenne μ et la **variance** (2.36) σ^2 .

EXAMPLE 2 Comme indiqué dans la Note 1 en 1.37, l'**estimateur du maximum de vraisemblance** (1.35) de la variance σ^2 utilise le dénominateur n au lieu de $n - 1$ et de ce fait est un estimateur biaisé. Dans la pratique, l'**écart-type d'échantillon** (1.17) est largement utilisé mais il est important de noter que la racine carrée de la variance d'échantillon utilisant $n - 1$ est un estimateur biaisé de l'**écart-type** (2.37) de la population.

EXAMPLE 3 Pour un échantillon aléatoire de n couples indépendants de variables aléatoires ayant chacun la même **loi normale à deux variables** (2.65), avec une **covariance** (2.43) égale à $\rho\sigma_{XY}$, la **covariance d'échantillon** (1.22) est un estimateur sans biais pour la covariance de la population. L'estimateur du maximum de vraisemblance utilise le dénominateur n au lieu de $n - 1$ et, de ce fait, est biaisé.

NOTE Les estimateurs **sans biais** sont utiles car, en moyenne, ils donnent la valeur correcte. Les estimateurs sans biais fournissent sans conteste un point de départ utile pour la recherche d'estimateurs «optimaux» des paramètres d'une population. La présente définition est de nature statistique.

Dans la pratique quotidienne, les praticiens tentent d'éviter d'introduire un biais dans une étude en s'assurant, par exemple, que l'échantillon aléatoire est représentatif de la population considérée.

1.35**maximum likelihood estimator**

estimator (1.12) assigning the value of the **parameter** (2.9) where the **likelihood function** (1.38) attains or approaches its highest value

NOTE 1 Maximum likelihood estimation is a well-established approach for obtaining parameter estimates where a **distribution** (2.11) has been specified [for example, **normal** (2.50), **gamma** (2.56), **Weibull** (2.63), and so forth]. These estimators have desirable statistical properties (for example, invariance under monotone transformation) and in many situations provide the estimation method of choice. In cases in which the maximum likelihood estimator is biased, a simple **bias** (1.33) correction sometimes takes place. As mentioned in Example 2 of 1.34 the maximum likelihood estimator for the **variance** (2.36) of the normal distribution is biased but it can be corrected by using $n - 1$ rather than n . The extent of the bias in such cases decreases with increasing sample size.

NOTE 2 The abbreviation MLE is commonly used both for maximum likelihood estimator and maximum likelihood estimation with the context indicating the appropriate choice.

1.36**estimation**

procedure that obtains a statistical representation of a **population** (1.1) from a **random sample** (1.6) drawn from this population

NOTE 1 In particular, the procedure involved in progressing from an **estimator** (1.12) to a specific **estimate** (1.31) constitutes estimation.

NOTE 2 Estimation is understood in a rather broad context to include point estimation, interval estimation or estimation of properties of populations.

NOTE 3 Frequently, a statistical representation refers to the estimation of a **parameter** (2.9) or parameters or a function of parameters from an assumed model. More generally, the representation of the population could be less specific, such as statistics related to impacts from natural disasters (casualties, injuries, property losses and agricultural losses — all of which an emergency manager might wish to estimate).

NOTE 4 Consideration of **descriptive statistics** (1.5) could suggest that an assumed model provides an inadequate representation of the data, such as indicated by a measure of the goodness of fit of the model to the data. In such cases, other models could be considered and the estimation process continued.

1.35**estimateur du maximum de vraisemblance**

estimateur (1.12) qui attribue la valeur du **paramètre** (2.9) lorsque la **fonction de vraisemblance** (1.38) atteint ou approche sa valeur la plus élevée

NOTE 1 L'estimation du maximum de vraisemblance est une méthode bien établie pour obtenir des estimations de paramètre lorsqu'une **distribution** (2.11) a été spécifiée [par exemple, **loi normale** (2.50), **loi gamma** (2.56), **loi de Weibull** (2.63), etc.]. Ces estimateurs ont des propriétés statistiques utiles (par exemple, invariance en transformée monotone) et fournissent dans de nombreux cas la méthode d'estimation de choix. Lorsque l'estimateur du maximum de vraisemblance est biaisé, une correction simple de **biais** (1.33) peut avoir lieu. Comme indiqué dans l'Exemple 2 en 1.34, l'estimateur du maximum de vraisemblance pour la variance (2.36) de la loi normale est biaisé mais il peut être corrigé en utilisant $n - 1$ au lieu de n . L'étendue du biais dans de tels cas décroît avec l'augmentation de la taille de l'échantillon.

NOTE 2 L'abréviation EMV est couramment utilisée pour l'estimateur du maximum de vraisemblance et l'estimation du maximum de vraisemblance, le contexte indiquant l'usage.

1.36**estimation (opération)**

opération ayant pour but, à partir d'un **échantillon aléatoire** (1.6) prélevé d'une **population** (1.1), d'obtenir une représentation statistique de cette population

NOTE 1 En particulier, l'opération de passage d'un **estimateur** (1.12) à une **estimation** (1.31) spécifique constitue une estimation.

NOTE 2 Le terme «estimation» est considéré dans un large contexte pour comprendre l'estimation ponctuelle, l'estimation par intervalle ou l'estimation de propriétés de populations.

NOTE 3 Une représentation statistique fait souvent référence à l'estimation d'un **paramètre** (2.9) ou de paramètres ou d'une fonction de paramètres issus d'un modèle supposé. D'une manière plus générale, la représentation de la population peut être moins précise, comme les statistiques relatives aux impacts des catastrophes naturelles (victimes, dommages corporels, pertes de biens, ainsi que les pertes agricoles — tout ce qu'un responsable de l'urgence peut être en droit de vouloir estimer).

NOTE 4 La mise en œuvre d'une **statistique descriptive** (1.5) peut suggérer qu'un modèle supposé fournisse une représentation inappropriée des données, telle qu'indiquée par une mesure de l'adéquation du modèle aux données. Dans ce cas, d'autres modèles peuvent être considérés et le processus d'estimation poursuivi.

1.37

maximum likelihood estimation

estimation (1.36) based upon the **maximum likelihood estimator** (1.35)

NOTE 1 For the **normal distribution** (2.50), the **sample mean** (1.15) is the **maximum likelihood estimator** (1.35) of the **parameter** (2.9) μ while the **sample variance** (1.16), using the denominator n rather than $n - 1$, provides the maximum likelihood estimator of σ^2 . The denominator $n - 1$ is typically used since this value provides an **unbiased estimator** (1.34).

NOTE 2 Maximum likelihood estimation is sometimes used to describe the derivation of an **estimator** (1.12) from the likelihood function.

NOTE 3 Although in some cases, a closed-form expression emerges using maximum likelihood estimation, there are other situations in which the maximum likelihood estimator requires an iterative solution to a set of equations.

NOTE 4 The abbreviation MLE is commonly used both for maximum likelihood estimator and maximum likelihood estimation with the context indicating the appropriate choice.

1.38

likelihood function

probability density function (2.26) evaluated at the **observed values** (1.4) and considered as a function of the **parameters** (2.9) of the **family of distributions** (2.8)

EXAMPLE 1 Consider a situation in which ten items are selected at random from a very large **population** (1.1) and 3 of the items are found to have a specific characteristic. From this sample, an intuitive **estimate** (1.31) of the population proportion having the characteristic is 0,3 (3 out of 10). Under a **binomial distribution** (2.46) model, the likelihood function (probability mass function as a function of p with n fixed at 10 and x at 3) achieves its maximum at $p = 0,3$, thus agreeing with intuition.

[This can be further verified by plotting the probability mass function of the **binomial distribution** (2.46) $120 p^3 (1 - p)^7$ versus p .]

EXAMPLE 2 For the **normal distribution** (2.50) with known **standard deviation** (2.37), it can be shown in general that the likelihood function takes its maximum at μ equal to the sample mean.

1.37

estimation du maximum de vraisemblance

estimation (1.36) fondée sur l'estimateur du maximum de vraisemblance (1.35)

NOTE 1 Pour la **loi normale** (2.50), la **moyenne d'échantillon** (1.15) est l'estimateur du maximum de vraisemblance (1.35) du **paramètre** (2.9) μ tandis que la **variance d'échantillon** (1.16), utilisant le dénominateur n plutôt que $n - 1$ fournit l'estimateur du maximum de vraisemblance de σ^2 . Le dénominateur $n - 1$ est généralement utilisé car cette valeur fournit un **estimateur sans biais** (1.34).

NOTE 2 L'expression «estimation du maximum de vraisemblance» est parfois utilisée pour décrire l'obtention d'estimateurs (1.12) à l'aide de la fonction de vraisemblance.

NOTE 3 Bien que dans certains cas l'utilisation de l'estimation du maximum de vraisemblance donne lieu à une expression analytique, il existe d'autres situations où l'estimateur du maximum de vraisemblance nécessite d'appliquer une solution itérative à un ensemble d'équations.

NOTE 4 L'abréviation EMV est couramment utilisée pour l'estimateur du maximum de vraisemblance et l'estimation du maximum de vraisemblance, le contexte indiquant l'usage.

1.38

fonction de vraisemblance

fonction de densité de probabilité (2.26) évaluée au niveau des **valeurs observées** (1.4) et considérée en fonction des **paramètres** (2.9) de la **famille de distributions** (2.8)

EXAMPLE 1 Soit une situation où dix individus sont sélectionnés au hasard à partir d'une **population** (1.1) très large et 3 des individus se révèlent avoir une caractéristique spécifique. Sur la base de cet échantillon, une **estimation** (1.31) intuitive de la proportion de population ayant la caractéristique est de 0,3 (3 sur 10). Dans le cadre d'un modèle de **loi binomiale** (2.46), la fonction de vraisemblance (fonction de masse de probabilité en fonction de p avec n fixé à 10 et x à 3) atteint son maximum à $p = 0,3$ ce qui concorde avec l'intuition. [Cela peut ensuite être vérifié en traçant la fonction de masse de probabilité de la **loi binomiale** (2.46) $120 p^3 (1 - p)^7$ par rapport à p .]

EXAMPLE 2 Pour la **loi normale** (2.50) à **écart-type** (2.37) connu, il est démontré que la fonction de vraisemblance atteint son maximum lorsque μ est égal à la moyenne d'échantillon.

1.39

profile likelihood function

likelihood function (1.38) as a function of a single **parameter** (2.9) with all other parameters set to maximize it

1.39

fonction de vraisemblance partielle

fonction de vraisemblance (1.38) considérée en fonction d'un **paramètre** (2.9) simple avec tous les autres paramètres fixés pour lui permettre d'atteindre le maximum

1.40

hypothesis

H

statement about a **population** (1.1)

NOTE Commonly the statement about the population concerns one or more **parameters** (2.9) in a **family of distributions** (2.8) or about the family of distributions.

1.40

hypothèse

H

formulation sur une **population** (1.1)

NOTE La formulation sur la population concerne généralement un ou plusieurs **paramètres** (2.9) d'une **famille de distributions** (2.8) ou de la famille de distributions.