NORME INTERNATIONALE

ISO 16269-6

Première édition
2005-04-01

Version corrigée
2006-01-01

Interprétation statistique des données —

Partie 6:
Détermination des intervalles statistiques de dispersion

Statistical interpretation of data —

Part 6: Determination of statistical tolerance intervals

Numéro de référence
ISO 16269-6:2005(F)

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.

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

Publié en Suisse

Sommaire

Page

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 16269-6 a été élaborée par le comité technique ISO/TC 69, *Application des méthodes statistiques*.

Cette première édition de l'ISO 16269-6 annule et remplace l'ISO 3207:1975, qui a fait l'objet d'une révision technique.

L'ISO 16269 comprend les parties suivantes, présentées sous le titre général *Interprétation statistique des données*:

— *Partie 6: Détermination des intervalles statistiques de dispersion*

— *Partie 7: Médiane — Estimation et intervalles de confiance*

— *Partie 8: Détermination des intervalles de prédiction*

Dans la présente version corrigée, le terme «tolérance» a été remplacé par «dispersion» et les termes «limite de spécification» ont été remplacés par «limite de tolérance» dans l'ensemble du document.

Introduction

Un intervalle statistique de dispersion est un intervalle estimé, d'après un échantillon, pour lequel il est possible d'affirmer avec un niveau de confiance $1 - \alpha$, par exemple 95 %, qu'il contient au moins une proportion donnée p d'individus de la population. Les limites d'un intervalle statistique de dispersion sont appelées limites statistiques de dispersion. Le niveau de confiance $1 - \alpha$ est la probabilité selon laquelle un intervalle statistique de dispersion construit de la manière spécifiée contiendra au moins une proportion p d'individus de la population. Inversement, la probabilité que cet intervalle contiendra moins que la proportion p d'individus de la population est α. La présente partie de l'ISO 16269 décrit les intervalles statistiques de dispersion unilatéraux et les intervalles statistiques de dispersion bilatéraux; un intervalle unilatéral est construit avec une limite inférieure ou une limite supérieure tandis qu'un intervalle bilatéral est construit avec une limite supérieure et une limite inférieure.

Les intervalles de dispersion sont fonction des observations de l'échantillon, c'est-à-dire des statistiques, et leurs valeurs seront généralement différentes pour des échantillons différents. Il est nécessaire que les observations soient indépendantes pour que les méthodes indiquées dans la présente partie de l'ISO 16269 soient valables.

La présente partie de l'ISO 16269 stipule deux types d'intervalles statistiques de dispersion: l'intervalle paramétrique et l'intervalle non paramétrique. L'approche paramétrique se fonde sur l'hypothèse selon laquelle la caractéristique étudiée dans la population a une distribution normale; ainsi, si l'hypothèse de normalité est avérée, le niveau de confiance avec lequel l'intervalle statistique de dispersion contient au moins une proportion p d'individus de la population ne peut être que de $1 - \alpha$. Pour les caractéristiques distribuées normalement, l'intervalle statistique de dispersion est déterminé à l'aide des formulaires A, B, C et D donnés dans l'Annexe A.

La présente partie de l'ISO 16269 ne traite pas des méthodes paramétriques s'appliquant à des distributions autres que les distributions normales. Si des écarts par rapport à la normalité sont suspectés dans la population, des intervalles statistiques de dispersion non paramétriques peuvent être construits. La procédure de détermination d'un intervalle statistique de dispersion pour une distribution continue quelconque est indiquée aux formulaires E et F de l'Annexe A.

Les limites de dispersion abordées dans la présente partie de l'ISO 16269 peuvent être utilisées pour comparer l'aptitude naturelle d'un processus avec une ou deux limites de tolérance données, soit une limite supérieure, U, soit une limite inférieure, L, ou encore les deux, dans la gestion d'un processus statistique. Cela est indiqué par le fait que ces limites de dispersion ont également été appelées limites naturelles du processus. Voir l'ISO 3534-2:1993, 3.2.4, ainsi que les remarques générales de l'ISO 3207, qui sera annulée et remplacée par la présente partie de l'ISO 16269.

Au-dessus de la limite de tolérance supérieure, U, il y a la fraction supérieure non conforme, p_U (ISO 3534-2:—, 3.2.5.5 et 3.3.1.4), et en dessous de la limite de tolérance inférieure, L, il y a la fraction inférieure non conforme, p_L (ISO 3534-2:—, 3.2.5.6 et 3.3.1.5). La somme $p_U + p_L = p_T$ est appelée fraction totale non conforme (ISO 3534-2:—, 3.2.5.7). Entre les limites de tolérance U et L, il y a la fraction conforme $1 - p_T$.

Dans la gestion du processus statistique, les limites U et L sont fixées à l'avance et les fractions p_U, p_L et p_T sont soit calculées, lorsque la distribution est supposée connue, soit estimées. Il existe beaucoup d'applications d'intervalles statistiques de dispersion, bien que l'exemple ci-dessus montre un exemple d'un problème de contrôle qualité. Des applications plus importantes et plus d'intervalles statistiques sont introduits dans de nombreux ouvrages tels que Hahn et Meeker [10].

Par contraste, pour les intervalles de dispersion dont il est question dans la présente partie de l'ISO 16269, le niveau de confiance pour l'estimateur d'intervalle et la proportion de distribution dans l'intervalle (correspondant à la fraction conforme mentionnée ci-dessus) sont fixés à l'avance, et les limites sont estimées.

Ces limites peuvent être comparées à U et à L. Ainsi la justesse des limites de tolérance données U et L peut être comparée aux propriétés réelles du processus. Les intervalles de dispersion unilatéraux sont utilisés uniquement lorsque la limite de tolérance supérieure, U, ou la limite de tolérance inférieure, L, est appropriée, tandis que les intervalles bilatéraux sont utilisés lorsque les limites supérieure et inférieure sont prises en compte simultanément.

La terminologie relative à ces limites et intervalles différents est confuse car les «limites de tolérance» étaient également autrefois appelées «limites de dispersion» (voir la Norme de terminologie ISO 3534-2:1993, 1.4.3, où ces deux termes, mais aussi le terme «valeurs limites», étaient utilisés comme synonymes pour désigner ce concept). Dans la dernière révision de l'ISO 3534-2:—, seul le terme «limites de tolérance» a été conservé pour désigner ce concept. En outre, le *Guide pour l'expression de l'incertitude de mesure (GUM)* [5] utilise le terme «facteur d'élargissement», défini comme un «facteur numérique utilisé comme multiplicateur de l'incertitude-type composée pour obtenir l'incertitude élargie». Cette utilisation du terme «élargissement» est différente de celle de la présente partie de l'ISO 16269.

Interprétation statistique des données —

Partie 6: Détermination des intervalles statistiques de dispersion

1 Domaine d'application

La présente partie de l'ISO 16269 décrit des méthodes permettant d'établir les intervalles statistiques de dispersion qui comprennent au moins une proportion spécifiée de la population avec un niveau de confiance spécifié. Des intervalles statistiques de dispersion unilatéraux et bilatéraux sont fournis, l'intervalle statistique de dispersion unilatéral étant caractérisé par une limite supérieure ou par une limite inférieure, tandis que l'intervalle statistique bilatéral possède à la fois une limite supérieure et une limite inférieure. Deux méthodes sont exposées: une méthode paramétrique, lorsque la caractéristique étudiée a une distribution normale, et une méthode non paramétrique, lorsque rien n'est connu de la distribution si ce n'est qu'elle est continue.

2 Références normatives

Les documents de référence suivants sont indispensables pour l'application du présent document. Pour les références datées, seule l'édition citée s'applique. Pour les références non datées, la dernière édition du document de référence s'applique (y compris les éventuels amendements).

ISO 3534-1, *Statistique — Vocabulaire et symboles — Partie 1: Probabilité et termes statistiques généraux*

ISO 3534-2:—[1)], *Statistique — Vocabulaire et symboles — Partie 2: Statistique appliquée*

3 Termes, définitions et symboles

3.1 Termes et définitions

Pour les besoins du présent document, les termes et définitions donnés dans l'ISO 3534-1 et l'ISO 3534-2 ainsi que les suivants s'appliquent.

3.1.1
intervalle statistique de dispersion
intervalle déterminé à partir d'un échantillon prélevé au hasard de manière qu'à un niveau de confiance spécifié, l'intervalle couvre au moins une proportion spécifiée de la population échantillonnée

NOTE Dans ce contexte, le niveau de confiance est la proportion à long terme d'intervalles construits de cette manière qui comprendra au moins la proportion également donnée de la population échantillonnée.

3.1.2
limite statistique de dispersion
statistique représentant un point limite d'un intervalle statistique de dispersion

NOTE Les intervalles statistiques de dispersion peuvent être soit unilatéraux, auquel cas ils ont une limite statistique de dispersion supérieure ou inférieure, soit bilatéraux, auquel cas ils possèdent les deux limites.

---

1) À publier. (Révision de l'ISO 3534-2:1993)

3.1.3
élargissement
proportion des individus d'une population se trouvant dans un intervalle statistique de dispersion

NOTE Ce concept est à ne pas confondre avec le concept de facteur d'élargissement utilisé dans le *Guide pour l'expression de l'incertitude de mesure (GUM)* [5].

3.1.4
population normale
population distribuée normalement

3.2 Symboles

Pour les besoins de la présente partie de l'ISO 16269, les symboles suivants s'appliquent.

i — suffixe d'une observation

$k_1(n; p; 1-\alpha)$ — facteur utilisé pour déterminer x_L ou x_U lorsque la valeur de σ est connue pour un intervalle de dispersion unilatéral

$k_2(n; p; 1-\alpha)$ — facteur utilisé pour déterminer x_L et x_U lorsque la valeur de σ est connue pour un intervalle de dispersion bilatéral

$k_3(n; p; 1-\alpha)$ — facteur utilisé pour déterminer x_L ou x_U lorsque la valeur de σ est inconnue pour un intervalle de dispersion unilatéral

$k_4(n; p; 1-\alpha)$ — facteur utilisé pour déterminer x_L et x_U lorsque la valeur de σ est inconnue pour un intervalle de dispersion bilatéral

n — nombre d'observations dans l'échantillon

p — proportion minimale de la population déclarée comme se trouvant dans l'intervalle statistique de dispersion

u_p — fractile d'ordre p de la distribution normale réduite

x_i — $i^{\text{ème}}$ valeur observée ($i = 1, 2, \ldots, n$)

$x_{\max}$ — valeur maximale des valeurs observées, $x_{\max} = \max\{x_1, x_2, \ldots, x_n\}$

$x_{\min}$ — valeur minimale des valeurs observées, $x_{\min} = \min\{x_1, x_2, \ldots, x_n\}$

x_L — limite inférieure de l'intervalle statistique de dispersion

x_U — limite supérieure de l'intervalle statistique de dispersion

$\bar{x}$ — moyenne de l'échantillon, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

s — écart-type de l'échantillon; $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n\left(n-1\right)}}$

$1 - \alpha$ niveau de confiance de la déclaration selon laquelle la proportion de la population se trouvant dans l'intervalle de dispersion est supérieure ou égale au niveau spécifié p

μ moyenne de la population

σ écart-type de la population

4 Méthodes

4.1 Population normale avec une variance et une moyenne connues

Lorsque les valeurs de la moyenne, μ, et de la variance, σ^2, d'une population normalement distribuée sont connues, la distribution de la caractéristique étudiée est complètement déterminée. Il y a exactement une proportion p de la population:

a) à la droite de $x_L = \mu - u_p \times \sigma$ (intervalle unilatéral);

b) à la gauche de $x_U = \mu + u_p \times \sigma$ (intervalle unilatéral);

c) entre $x_L = \mu - u_{(1+p)/2} \times \sigma$ et $x_U = \mu + u_{(1+p)/2} \times \sigma$ (intervalle bilatéral).

NOTE Dans la mesure où l'on sait que ces déclarations sont justes, elles sont faites avec un niveau de confiance de 100 %.

Dans les équations ci-dessus, u_p est le fractile d'ordre p de la distribution normale réduite. Les valeurs numériques de u_p sont indiquées aux dernières lignes des Tableaux B.1 à B.6 et C.1 à C.6.

4.2 Population normale avec une variance connue et une moyenne inconnue

Les formulaires A et B, donnés dans l'Annexe A, sont applicables lorsque la variance de la population normale est connue alors que la moyenne est inconnue. Le formulaire A s'applique aux cas unilatéraux tandis que le formulaire B s'applique aux cas bilatéraux.

4.3 Population normale avec une variance et une moyenne inconnues

Les formulaires C et D, donnés dans l'Annexe A, sont applicables lorsque la moyenne et la variance de la population normale sont inconnues. Le formulaire C s'applique aux cas unilatéraux tandis que le formulaire D s'applique aux cas bilatéraux.

4.4 Distribution continue quelconque de type inconnu

Si la caractéristique à l'étude est une variable continue provenant d'une population de forme inconnue et si un échantillon de n observations aléatoires et indépendantes de la caractéristique a été prélevé, alors un intervalle statistique de dispersion peut être déterminé à partir des observations ordonnées. La méthode indiquée aux formulaires E et F de l'Annexe A permet de déterminer l'élargissement ou l'effectif de l'échantillon nécessaires aux intervalles de dispersion déterminés à partir des valeurs extrêmes x_{min} et x_{max} de l'échantillon d'observations avec un niveau de confiance de $1 - \alpha$.

NOTE Les intervalles statistiques de dispersion qui ne sont pas fonction de la forme de la population échantillonnée sont appelés intervalles de dispersion *non paramétriques*.

La présente partie de l'ISO 16269 ne préconise pas de méthode pour les distributions d'un type connu autre que la distribution normale. Toutefois, si la distribution est continue, la méthode non paramétrique peut être utilisée. Une sélection de références à de la littérature scientifique pouvant aider à déterminer les intervalles de dispersion pour d'autres distributions est aussi fournie à la fin de ce document.

5 Exemples

5.1 Données

Les formulaires A à D, donnés dans l'Annexe A, sont illustrés par des exemples à l'aide des valeurs numériques de l'ISO 2854:1976, Article 2, paragraphe 1 des remarques introductives, Tableau X, fil 2: 12 mesures de la charge de rupture du fil en coton. Il convient de noter que le nombre d'observations, $n = 12$, indiqué pour ces exemples, est considérablement inférieur à celui recommandé dans l'ISO 2602 [1]. L'unité de mesure pour exprimer les données numériques et les calculs dans les différents exemples est le centinewton (voir Tableau 1).

Tableau 1 — Données pour les Exemples 1 à 4

Valeurs en centinewtons

x	228,6	232,7	238,8	317,2	315,8	275,1	222,2	236,7	224,7	251,2	210,4	270,7

Ces mesures proviennent d'un lot de 12 000 bobines, d'une même série de fabrication, emballées dans 120 boîtes contenant chacune 100 bobines. Douze boîtes de ce lot ont été prélevées au hasard et une bobine a été prise au hasard dans chacune de ces boîtes. Des éprouvettes de 50 cm de long ont été découpées dans le fil de ces bobines, à environ 5 m de l'extrémité libre. Les essais proprement dits ont été réalisés sur les parties centrales de ces éprouvettes. Des informations antérieures permettent de penser raisonnablement que les charges de rupture mesurées dans ces conditions ont une distribution pratiquement normale. Il est démontré, dans l'ISO 2954:1976, que les données ne contredisent pas l'hypothèse d'une distribution normale.

Les résultats suivants sont produits:

Effectif de l'échantillon: $n = 12$

Moyenne de l'échantillon: $\bar{x} = 3\,024{,}1/12 = 252{,}01$

Écart-type de l'échantillon: $$s = \sqrt{\frac{n\sum x^2 - \left(\sum x\right)^2}{n\,(n-1)}} = \sqrt{\frac{166\,772{,}27}{12 \times 11}} = \sqrt{1\,263{,}426\,3} = 35{,}545$$

La présentation formelle des calculs sera donnée uniquement pour le formulaire C de l'Annexe A (intervalle unilatéral, variance inconnue).

5.2 Exemple 1: intervalle statistique de dispersion unilatéral sous variance connue

Supposer que les mesures obtenues précédemment montrent que la dispersion est constante d'un lot à l'autre, provenant du même fournisseur, et qu'elle est représentée par un écart-type $\sigma = 33{,}150$, bien que la moyenne ne soit pas constante. Une limite x_L est requise de manière à ce qu'il soit possible d'affirmer avec un niveau de confiance $1 - \alpha = 0{,}95$ (95 %) qu'au moins 0,95 (95 %) des charges de rupture des individus du lot, mesurées dans les mêmes conditions, sont supérieures à x_L.

La Tableau B.4 donne

$$k_1\,(12;\, 0{,}95;\, 0{,}95) = 2{,}120$$

d'où

$$x_L = \bar{x} - k_1(n;\, p;\, 1-\alpha) \times \sigma = 252{,}01 - 2{,}120 \times 33{,}150 = 181{,}732$$

Une plus petite valeur de la limite inférieure, x_L, serait obtenue si une proportion plus importante de la population (par exemple, $p = 0{,}99$) et/ou un niveau de confiance supérieur (par exemple $1 - \alpha = 0{,}99$) étaient requis.

5.3 Exemple 2: intervalle statistique de dispersion bilatéral sous variance connue

Dans des conditions identiques à celles de l'Exemple 1, supposer que les limites x_L et x_U sont requises de manière qu'il soit possible d'affirmer avec un niveau de confiance $1 - \alpha = 0{,}95$ qu'au moins une proportion de $p = 0{,}90$ (90 %) des charges de rupture du lot se situe entre x_L et x_U.

Le Tableau C.4 donne

$$k_2\,(12;\,0{,}90;\,0{,}95) = 1{,}889$$

d'où

$$x_L = \bar{x} - k_2(n;\,p;\,1-\alpha) \times \sigma = 252{,}01 - 1{,}889 \times 33{,}150 = 189{,}390$$

$$x_U = \bar{x} + k_2(n;\,p;\,1-\alpha) \times \sigma = 252{,}01 + 1{,}889 \times 33{,}150 = 314{,}630$$

Il convient que la comparaison de cet exemple avec l'Exemple 1 fasse bien comprendre la différence entre le fait de garantir qu'au moins 90 % d'une population se situe entre les limites x_L et x_U et le fait de garantir qu'au plus 5 % se trouve au-delà de ces limites.

5.4 Exemple 3: intervalle statistique de dispersion unilatéral sous variance inconnue

Il est ici supposé que l'écart-type de la population est inconnu et doit être estimé à partir de l'échantillon. Les mêmes exigences seront supposées que pour le cas où l'écart-type est connu (Exemple 1); ainsi, $p = 0{,}95$ et $1 - \alpha = 0{,}95$. La présentation détaillée des résultats est donnée ci-après.

Détermination de l'intervalle statistique de dispersion de la proportion p a) Intervalle unilatéral «à droite» Valeurs déterminées: b) proportion de la population choisie pour l'intervalle de dispersion: $p = 0{,}95$ c) niveau de confiance choisi: $1 - \alpha = 0{,}95$ d) effectif de l'échantillon: $n = 12$ Valeur du facteur de dispersion, provenant du Tableau D.4: $k_3(n;\, p;\, 1 - \alpha) = 2{,}737$
Calculs: $\bar{x} = \sum x / n = 252{,}01$ $s = \sqrt{\dfrac{n\sum x^2 - (\sum x)^2}{n(n-1)}} = 35{,}545$ $k_3(n;\, p;\, 1 - \alpha) \times s = 97{,}286\,7$
Résultats: intervalle unilatéral «à droite» L'intervalle de dispersion qui contiendra au moins une proportion p de la population avec un niveau de confiance $1 - \alpha$ a une limite inférieure: $x_L = \bar{x} - k_3(n;\, p;\, 1-\alpha) \times s = 154{,}723$

5.5 Exemple 4: intervalle statistique de dispersion bilatéral sous variance inconnue

Dans les mêmes conditions que dans l'Exemple 2, supposer qu'il est requis de calculer les limites x_L et x_U de manière qu'il soit possible d'affirmer avec un niveau de confiance $1 - \alpha = 0{,}95$ que dans une proportion du lot au moins égale à $p = 0{,}90$ (90 %), la charge de rupture est comprise entre x_L et x_U.

La Tableau E.4 donne

$$k_4(n;\; p;\; 1-\alpha) = 2{,}671$$

d'où

$$x_L = \bar{x} - k_4(n;\, p;\, 1-\alpha) \times s = 252{,}01 - 2{,}671 \times 35{,}545 = 157{,}069$$

$$x_U = \bar{x} + k_4(n;\, p;\, 1-\alpha) \times s = 252{,}01 + 2{,}671 \times 35{,}545 = 346{,}951$$

Noter que la valeur de x_L est inférieure et que la valeur de x_U est supérieure à celles de l'Exemple 2 (variance connue) parce que l'utilisation de s à la place de σ nécessite une valeur de la constante de dispersion plus importante pour tenir compte de l'incertitude supplémentaire. Il est nécessaire d'assumer le fait de ne pas connaître l'écart-type de la population, σ, et l'extension de l'intervalle statistique de dispersion prend cela en

