NORME INTERNATIONALE

ISO 16269-6

Deuxième édition
2014-01-15

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 16269-6:2014(F)

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

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 procédures utilisées pour élaborer le présent document et celles destinées à sa mise à jour sont décrites dans les Directives ISO/CEI, Partie 1. Il convient, en particulier de prendre note des différents critères d'approbation requis pour les différents types de documents ISO. Le présent document a été rédigé conformément aux règles de rédaction données dans les Directives ISO/CEI, Partie 2, www.iso.org/directives.

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. Les détails concernant les références aux droits de propriété intellectuelle ou autres droits analogues identifiés lors de l'élaboration du document sont indiqués dans l'Introduction et/ou sur la liste ISO des déclarations de brevets reçues, www.iso.org/patents.

Les éventuelles appellations commerciales utilisées dans le présent document sont données pour information à l'intention des utilisateurs et ne constituent pas une approbation ou une recommandation.

Le comité chargé de l'élaboration du présent document est l'ISO/TC 69, *Application des méthodes statistiques*.

Cette deuxième édition de l'ISO 16269 annule et remplace la première édition (ISO 16269-6:2005), 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 4: Détection et traitement des valeurs aberrantes*

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

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

— *Partie 8: Détermination des intervalles de prédiction*

Introduction

Un intervalle statistique de dispersion est un intervalle estimé, d'après un échantillon, pour lequel il est possible d'affirmer au niveau de confiance $1 - \alpha$, par exemple 0,95, qu'il contient au moins une proportion spécifiée 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 de la population. Inversement, la probabilité que cet intervalle contiendra moins que la proportion p 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.

Un intervalle statistique de dispersion dépend d'un niveau de confiance $1 - \alpha$ et d'une proportion donnée p de la population. Le niveau de confiance d'un intervalle statistique de dispersion est bien défini à partir d'un intervalle de confiance pour un paramètre. L'expression de la confiance d'un intervalle de confiance est la proportion des cas où l'intervalle de confiance contient la valeur vraie du paramètre, une proportion de $1 - \alpha$ des cas, dans une longue série d'échantillons aléatoires répétés dans des conditions identiques. De la même manière, l'expression de la confiance d'un intervalle statistique de dispersion indique qu'au moins une proportion p de la population est contenue dans l'intervalle dans une proportion de $1 - \alpha$ des cas d'une longue série d'échantillons aléatoires répétés dans des conditions identiques. Ainsi, si nous considérons la proportion établie p de la population comme un paramètre, l'idée sous-jacente aux intervalles statistiques de dispersion est similaire à l'idée sous-jacente aux intervalles de confiance.

Les intervalles statistiques 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.

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

La présente partie de l'ISO 16269 ne traite pas des méthodes paramétriques s'appliquant à des distributions autres qu'une loi normale. 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 dans le formulaire D de l'Annexe B.

Les limites statistiques de dispersion abordées dans la présente partie de l'ISO 16269 peuvent être utilisées pour comparer la capabilité naturelle d'un processus avec une ou deux limites de spécification données, soit une limite supérieure, U, soit une limite inférieure, L, ou encore les deux, dans la gestion statistique d'un processus.

Au-dessus de la limite de spécification supérieure, U, il y a la proportion de non-conformes supérieure, p_U (ISO 3534-2:2006, 2.5.4), et en dessous de la limite de spécification inférieure, L, il y a la proportion de non conformes inférieure, p_L (ISO 3534-2:2006, 2.5.5). La somme $p_U + p_L = p_t$ est appelée proportion de non conformes totale (ISO 3534-2:2006, 2.5.6). Entre les limites de spécification U et L, il y a la proportion de conformes $1 - p_t$.

Les idées sous-jacentes aux intervalles statistiques de dispersion sont plus répandues qu'on ne le pense généralement, par exemple dans l'échantillonnage pour acceptation par mesures et dans la gestion statistique d'un processus, comme on le verra dans les deux paragraphes qui suivent.

Dans l'échantillonnage pour acceptation par mesures, les limites U et/ou L sont connues, p_U, p_L ou p_t est spécifié en tant que limite de qualité acceptable (LQA), α est implicite et le lot est accepté s'il y a au moins un niveau de confiance implicite de $100(1 - \alpha)$ % que la LQA n'est pas dépassée.

Dans la gestion statistique d'un processus, les limites U et L sont fixées à l'avance et les proportions p_U, p_L et p_t sont soit calculées, lorsque la distribution est supposée connue, soit estimées. Il s'agit d'un exemple d'application de contrôle qualité, mais il existe de nombreuses autres applications d'intervalles statistiques de dispersion présentées dans de nombreux ouvrages tels que Hahn et Meeker.[13]

Par contre, pour les intervalles statistiques de dispersion dont il est question dans la présente partie de l'ISO 16269, le niveau de confiance pour l'estimateur de l'intervalle et la proportion de la distribution située dans les limites de l'intervalle (correspondant à la proportion de conformes mentionnée ci-dessus) sont fixés à l'avance, et les limites sont estimées. Ces limites peuvent être comparées à U et à L. Ainsi, l'adéquation des limites de spécification données U et L peut être comparée aux propriétés réelles du processus. Les intervalles statistiques de dispersion unilatéraux sont utilisés lorsque seule la limite de spécification supérieure, U, ou la limite de spécification inférieure, L, est appropriée, tandis que les intervalles bilatéraux sont utilisés lorsque les limites de spécification supérieure et inférieure sont prises en compte simultanément.

La terminologie relative à ces limites et intervalles différents est confuse car les «limites de spécification» étaient également autrefois appelées «limites de tolérance» (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 3534-2:2006, 3.1.3, seul le terme «limites de spécification» a été conservé pour ce concept.

La première édition de la présente norme contenait des tableaux détaillés du facteur k pour les intervalles de dispersion unilatéraux et bilatéraux lorsque la moyenne est inconnue, mais l'écart-type connu. Dans cette deuxième édition de la norme, ces tableaux sont supprimés. A leur place, les facteurs k exacts sont donnés à l'Annexe A lorsque l'un des paramètres de la loi normale est inconnu et l'autre paramètre est connu.

La première édition de la présente norme considérait les intervalles statistiques de dispersion en se fondant sur un seul échantillon de taille n. La présente édition considère les intervalles statistiques de dispersion pour m populations avec le même écart-type en se fondant sur des échantillons de chacune des m populations, chaque échantillon ayant la même taille n.

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 à un niveau de confiance spécifié. Les intervalles statistiques de dispersion unilatéraux et bilatéraux sont tous deux présentés l'intervalle statistique de dispersion unilatéral étant caractérisé par une limite supérieure ou par une limite inférieure, tandis que l'intervalle statistique de dispersion bilatéral possède à la fois une limite supérieure et une limite inférieure. Deux méthodes sont exposées: une méthode paramétrique, lorsque la caractéristique étudiée est distribuée selon une loi normale, et une méthode non paramétrique, lorsqu'on ne sait rien de la distribution si ce n'est qu'elle est continue. Il existe également une procédure permettant d'établir les intervalles statistiques de dispersion bilatéraux pour plus d'un échantillon normal avec une variance identique mais inconnue.

2 Références normatives

Les documents suivants, en totalité ou en partie, sont référencés de manière normative dans le présent document et sont indispensables pour son application. 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 3534-1:2006, *Statistique — Vocabulaire et symboles — Partie 1: Termes statistiques généraux et termes utilisés en calcul des probabilités*

ISO 3534-2:2006, *Statistique — Vocabulaire et symboles — Partie 2: Statistique appliquée*

3 Termes, définitions et symboles

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

3.1 Termes et définitions

3.1.1
intervalle statistique de dispersion
intervalle déterminé à partir d'un échantillon aléatoire de manière qu'un niveau de confiance spécifié puisse être défini et que l'intervalle couvre au moins une proportion spécifiée de la population échantillonnée

[SOURCE: ISO 3534-1:2006, 1.26]

Note 1 à l'article: Dans ce contexte, le niveau de confiance est la proportion à long terme des intervalles construits de cette manière qui incluront au moins la proportion spécifiée de la population échantillonnée.

3.1.2
limite statistique de dispersion
statistique représentant une borne d'un intervalle statistique de dispersion

[SOURCE: ISO 3534-1:2006, 1.27]

Note 1 à l'article: 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 une limite supérieure.

3.1.3
proportion de recouvrement
proportion des individus d'une population se trouvant à l'intérieur d'un intervalle statistique de dispersion

Note 1 à l'article: Ce concept ne doit pas être confondu 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 selon une loi normale

3.2 Symboles

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

$k_1(n; p; 1 - \alpha)$ — facteur utilisé pour déterminer les limites des intervalles unilatéraux, c'est-à-dire x_L ou x_U lorsque μ est connu et σ est inconnu

$k_2(n; p; 1 - \alpha)$ — facteur utilisé pour déterminer les limites des intervalles bilatéraux, c'est-à-dire x_L et x_U lorsque μ est connu et σ est inconnu

$k_3(n; p; 1 - \alpha)$ — facteur utilisé pour déterminer les limites des intervalles unilatéraux, c'est-à-dire x_L ou x_U lorsque μ est inconnu et σ est connu

$k_4(n; p; 1 - \alpha)$ — facteur utilisé pour déterminer les limites des intervalles bilatéraux, c'est-à-dire x_L et x_U lorsque μ est inconnu et σ est inconnu

$k_C(n; p; 1 - \alpha)$ — facteur utilisé pour déterminer x_L ou x_U lorsque les valeurs de μ et σ sont inconnues pour un intervalle statistique de dispersion unilatéral. Le suffixe C est choisi parce que ce facteur k est tabulé dans l'Annexe C.

$k_D(n; m; p; 1 - \alpha)$ — facteur utilisé pour déterminer x_{Li} et x_{Ui} ($i = 1,2,...,m$; $m \geq 2$) lorsque les valeurs des moyennes μ_i et la valeur de σ commun sont inconnues pour les m intervalles statistiques de dispersion bilatéraux. Le suffixe D est choisi parce que ce facteur k est tabulé dans l'Annexe D.

n — nombre d'observations dans l'échantillon (taille d'échantillon)

p — proportion minimale de la population affirmée comme se trouvant à l'intérieur de l'intervalle statistique de dispersion

u_p — fractile d'ordre p de la loi normale centrée réduite

x_j — $j^{\text{ème}}$ valeur observée ($j = 1,2,...,n$)

x_{ij} — $j^{\text{ème}}$ valeur observée ($j = 1,2,...,n$) of $i^{\text{ème}}$ valeur observée ($i = 1,2,...,m$)

$x_{\max}$	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, $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$
$\bar{x}_i$	moyenne du $i^{\text{ème}}$ échantillon $(i = 1,2,\ldots,m)$, $\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$
s	écart-type de l'échantillon, $s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2} = \sqrt{\frac{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2}{n(n-1)}}$
s_i	écart-type du $i^{\text{ème}}$ échantillon $(i = 1,2,\ldots,m)$, $s_i = \sqrt{\frac{1}{(n-1)}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2}$
s_P	écart-type de l'échantillon établi sur un ensemble de données; $s_P = \sqrt{\frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} s_i^2}$
$1 - \alpha$	niveau de confiance de l'affirmation selon laquelle la proportion de la population se trouvant à l'intérieur de l'intervalle de dispersion est supérieure ou égale au niveau spécifié p
μ	moyenne de la population
μ_i	moyenne de la $i^{\text{ème}}$ population ($i = 1,2,\ldots,m$)
σ	écart-type de la population

4 Méthodes

4.1 Population normale avec une moyenne et une variance connues

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

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

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

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

Dans les équations ci-dessus, μ_p est le fractile d'ordre p de la loi normale centrée réduite.

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

4.2 Population normale avec une moyenne inconnue et une variance connue

Lorsque l'un ou les deux paramètres de la loi normale sont inconnus, mais estimés à partir d'un échantillon aléatoire, des intervalles ayant des propriétés similaires à ceux du paragraphe 4.1 peuvent encore être construits. Supposons, par exemple, que la moyenne soit inconnue, mais que la variance soit connue. Il est alors possible de déterminer une constante k telle que l'intervalle, compris entre

$$x_{\mathrm{L}} = \overline{x} - k\sigma \text{ et } x_{\mathrm{U}} = \overline{x} + k\sigma$$

contienne *au moins* une proportion p de la population au niveau de confiance spécifié de $1 - \alpha$. Noter qu'il existe deux distinctions importantes par rapport à la situation de 4.1 où les paramètres étaient supposés connus. Premièrement, lorsqu'un ou plusieurs paramètres sont estimés, l'intervalle contient *au moins* une proportion p de la population, mais pas exactement une proportion p de la population. Deuxièmement, lorsque des paramètres sont estimés, la déclaration est vraie uniquement au niveau de confiance préalablement spécifié de $1 - \alpha$. Le facteur k dans l'expression des limites ci-dessus dépend des paramètres inconnus de la loi normale, de la proportion p, du niveau de confiance $1-\alpha$ et du nombre d'observations dans l'échantillon aléatoire lorsque l'un des paramètres de la loi normale est inconnu et que l'autre paramètre est connu, l'Annexe A donne les facteurs k exacts.

4.3 Population normale avec une moyenne inconnue et une variance inconnue

Les formulaires A et B, donnés dans l'Annexe B, sont applicables lorsque la moyenne et la variance de la population normale sont inconnues. Le formulaire A s'applique aux cas unilatéraux tandis que le formulaire B s'applique aux cas bilatéraux. Le formulaire A est utilisé avec les tableaux de facteurs k de l'Annexe C, ou en variante en utilisant la formule exacte relative au facteur k donnée en A.5. Le formulaire B est utilisé avec les facteurs k donnés dans la première colonne des tableaux de l'Annexe D. Les détails relatifs au calcul des facteurs k de l'Annexe D sont donnés à l'Annexe F.

4.4 Populations normales avec des moyennes inconnues et des variances identiques et inconnues

Le formulaire C, donné dans l'Annexe B, est applicable lorsque les moyennes et les variances des populations normales sont inconnues. En outre, les variances sont supposées être identiques pour toutes les populations étudiées, auquel cas on parle de variance commune.

4.5 Distribution continue quelconque de type inconnu

Si la caractéristique étudiée est une variable d'une population de forme inconnue, alors un intervalle statistique de dispersion peut être déterminé à partir des statistiques d'ordre $x_{(i)}$ d'un échantillon de n observations aléatoires indépendantes. La méthode indiquée dans le formulaire D, utilisée conjointement avec les Tableaux E.1 et E.2, donne les étapes permettant de déterminer la taille d'échantillon requis sur la base des statistiques d'ordre à utiliser, du niveau de confiance souhaité et du contenu souhaité.

NOTE 1 Les intervalles statistiques de dispersion dans lesquels le choix des bornes (sur la base des statistiques d'ordre) ne dépend pas de la population échantillonnée sont appelés intervalles statistiques de dispersion *non paramétriques*.

NOTE 2 La présente Norme internationale ne préconise pas de méthode pour les distributions d'un type connu autre qu'une loi normale. Toutefois, si la distribution est continue, la méthode non paramétrique peut être utilisée. Une sélection de références à la littérature scientifique pouvant aider à déterminer les intervalles de dispersion pour d'autres distributions est également fournie à la fin de ce document.

5 Exemples

5.1 Données pour les Exemples 1 et 2

Les formulaires A à B, donnés dans l'Annexe B, sont illustrés par les Exemples 1 et 2 à l'aide des valeurs numériques de l'ISO 2854:1976,[2] 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 plus faible que 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 et 2

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, issu 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 2854:1976, que les données ne contredisent pas l'hypothèse d'une distribution normale.

En utilisant le test graphique par diagrammes à surfaces pour les valeurs aberrantes décrit dans l'ISO 16269-4, il est également possible de conclure qu'aucune des valeurs de données ne peut être déclarée comme une valeur aberrante avec un niveau de signification α = 0,05.

Les données du Tableau 1 donnent les résultats suivants:

Taille d'échantillon: $n = 12$

Moyenne de l'échantillon: $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = 3\,024{,}1/12 = 252{,}01$

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

La présentation formelle des calculs sera donnée dans l'Exemple 1 en utilisant le formulaire A de l'Annexe B (intervalle unilatéral, variance et moyenne inconnues).

5.2 Exemple 1: Intervalle statistique de dispersion unilatéral avec variance et moyenne inconnues

Une limite x_L est requise de manière à ce qu'il soit possible d'affirmer au niveau de confiance $1 - \alpha$ = 0,95 (95 %) qu'au moins 0,95 (95 %) des charges de rupture des éléments du lot, mesurées dans les mêmes conditions, sont supérieures à x_L. Les résultats sont présentés de manière détaillée ci-dessous.

Détermination de l'intervalle statistique de dispersion de la proportion p:

a) intervalle unilatéral « à droite » (intervalle statistique de dispersion unilatéral inférieur)

Valeurs déterminées:

b) proportion de la population choisie pour l'intervalle statistique de dispersion: p = 0,95

c) niveau de confiance choisi: 1 − α = 0,95

d) Taille d'échantillon: n = 12

Valeur du facteur de dispersion, provenant du Tableau C.2: $k_C(n;p;1-\alpha)=2{,}736\,4$

Calculs:

$$\bar{x}=\frac{1}{n}\sum_{j=1}^{n}x_j=252{,}01$$

$$s=\sqrt{\frac{n\sum_{j=1}^{n}x_j^2-\left(\sum_{j=1}^{n}x_j\right)^2}{n(n-1)}}=35{,}545$$

$$k_C(n;p;1-\alpha)\times s=97{,}265\,3$$

Résultats: intervalle unilatéral «à droite» (intervalle statistique de dispersion unilatéral inférieur)

L'intervalle de dispersion qui contiendra au moins une proportion p de la population au niveau de confiance 1 − α a une limite inférieure:

$$x_L=\bar{x}-k_C(n;p;1-\alpha)\times s=154{,}7$$

5.3 Exemple 2: Intervalle statistique de dispersion bilatéral avec moyenne inconnue et variance inconnue

Supposons qu'il soit requis de calculer les limites x_L et x_U de manière qu'il soit possible d'affirmer au niveau de confiance 1 − α = 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.

Le Tableau D.4 donne pour un nombre d'échantillon, m = 1 ayant pour taille n = 12, le facteur de la limite statistique de dispersion bilatérale, k_D:

$$k_D(n;1;p;1-\alpha)=2{,}6703$$

d'où

$$x_L=\bar{x}-k_D(n;1;p;1-\alpha)\times s=252{,}01-2{,}6703\times 35{,}545=157{,}0$$

$$x_U=\bar{x}+k_D(n;1;p;1-\alpha)\times s=252{,}01+2{,}6703\times 35{,}545=347{,}0$$

5.4 Données pour les Exemples 3 et 4

Supposons que le pourcentage de solides dans chacun de quatre lots de levure de bière humide, provenant chacun d'un fournisseur différent, doive être déterminé. Les pourcentages des quatre lots sont distribués selon une loi normale avec des moyennes inconnues μ_i i = 1,2,3,4. Sur la base de l'expérience acquise avec ces fournisseurs, il peut être supposé que les variances sont identiques. Un essai pour les données suivantes ne donne aucune raison de supposer le contraire. Les données sont donc supposées avoir une variance commune σ^2. Le chercheur souhaite déterminer les intervalles statistiques de dispersion bilatéraux pour les pourcentages de solides dans chaque lot.

Les valeurs relatives à des échantillons aléatoires de taille n = 10 prélevés dans les quatre lots[14] sont indiquées dans le Tableau 2:

Tableau 2 — Données pour les Exemples 3 et 4

Valeurs en pourcentage

j/i	1	2	3	4	5	6	7	8	9	10
1	20	18	16	21	19	17	20	16	19	18
2	19	14	17	13	10	16	14	12	15	11
3	11	12	14	10	8	10	13	9	12	8
4	10	7	11	9	6	11	8	12	13	14

Noter que la $j^{\text{ème}}$ valeur du $i^{\text{ème}}$ échantillon est désignée par: x_{ij}.

Les résultats suivants sont obtenus:

Taille des échantillons: n = 10

Nombre d'échantillons: m = 4

Moyennes des échantillons de chacun des quatre lots:

$\bar{x}_1 = 184/10 = 18{,}4$; $\bar{x}_2 = 141/10 = 14{,}1$; $\bar{x}_3 = 107/10 = 10{,}7$; $\bar{x}_4 = 101/10 = 10{,}1$

Variances des échantillons de chacun des quatre lots:

$$s_1^2 = \frac{n\sum_{j=1}^{n} x_{1j}^2 - \left(\sum_{j=1}^{n} x_{1j}\right)^2}{n(n-1)} = \frac{264}{10\times 9} = 2{,}933\,3 \ ; \quad s_2^2 = \frac{n\sum_{j=1}^{n} x_{2j}^2 - \left(\sum_{j=1}^{n} x_{2j}\right)^2}{n(n-1)} = \frac{689}{10\times 9} = 7{,}655\,6$$

$$s_3^2 = \frac{n\sum_{j=1}^{n} x_{3j}^2 - \left(\sum_{j=1}^{n} x_{3j}\right)^2}{n(n-1)} = \frac{381}{10\times 9} = 4{,}233\,3 \ ; \quad s_4^2 = \frac{n\sum_{j=1}^{n} x_{4j}^2 - \left(\sum_{j=1}^{n} x_{4j}\right)^2}{n(n-1)} = \frac{609}{10\times 9} = 6{,}766\,7$$

Écart-type de l'échantillon établi sur l'ensemble des données:

$$s_{\mathrm{P}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} s_i^2} = \sqrt{\frac{1}{4}(2{,}933\,3 + 7{,}655\,6 + 4{,}233\,3 + 6{,}766\,7)} = 2{,}323\,2$$

Degrés de liberté de l'écart-type établi sur l'ensemble des données:

$f = m(n-1) = nm - m = 36$

5.5 Exemple 3: Intervalles statistiques de dispersion unilatéraux pour des populations séparées avec variance commune inconnue

Supposons que l'on souhaite calculer les intervalles statistiques de dispersion inférieurs pour les quatre fournisseurs, c'est-à-dire que l'on souhaite calculer les intervalles qui contiennent au moins une proportion p pour tous les fournisseurs. Le Tableau C ne peut pas fournir la réponse, mais les intervalles ont la même forme que celle indiquée dans l'Exemple 1, à savoir une constante multipliée par l'écart-type estimé soustrait de la moyenne estimée:

$$x_{\mathrm{L}i} = \bar{x}_i - k(n_i; f; p; 1-\alpha) \times s_{\mathrm{P}},$$

où la constante $k(n_i;f;p;1-\alpha)$ dépend de la taille du $i^{\text{ème}}$ échantillon et des degrés de liberté de l'écart-type établi sur l'ensemble des données. L'expression relative à la constante est issue de l'Article A.5, voir l'Équation (A.14):

$$k(n_i; f; p; 1-\alpha) = \frac{1}{\sqrt{n_i}} t_{1-\alpha}(\sqrt{n_i} u_p; f),$$

où $t_{1-\alpha}(\sqrt{n_i} u_p; f)$ désigne le quantile d'ordre $1 - \alpha$ de la loi de t non centrée avec le paramètre de décentrement $\sqrt{n_i} u_p$ et f degrés de liberté. La loi de t non centrée et en particulier ses quantiles sont disponibles dans des logiciels statistiques. Supposons qu'une proportion $p = 0{,}95$ et un niveau de confiance $1 - \alpha = 0{,}95$ soient souhaités. Dans ce cas, $n_{\mathrm{i}} = 10$ et $f = m(n - 1) = nm - m = 36$, et la constante est donc

$$k(10;36;0{,}95;0{,}95) = \frac{1}{\sqrt{10}} t_{0{,}95}(\sqrt{10} \times 1{,}6449;36) = 2{,}3471,$$

où le quantile d'ordre 0,95 de la loi normale centrée réduite $u_{0{,}95} = 1{,}6449$ est inséré.

Les valeurs données dans les tableaux de l'Annexe C sont les cas particuliers où les degrés de liberté sont égaux à la taille de l'échantillon moins 1, c'est-à-dire les degrés de liberté de l'écart-type sur la base d'un seul échantillon de taille n

$$k_{\mathrm{C}}(n;p;1-\alpha) = k(n;n-1;p;1-\alpha) = \frac{1}{\sqrt{n}} t_{1-\alpha}(\sqrt{n} u_p; n-1),$$

c'est-à-dire le cas particulier où le nombre de degrés de liberté de l'estimation de la variance est $n - 1$

Il s'ensuit que les limites statistiques de dispersion unilatérales calculées pour l'ensemble des quatre lots sont les suivantes:

Premier lot: $x_{\mathrm{L}1} = \bar{x}_1 - k\left(n_1;\nu;p;1-\alpha\right) \times s_{\mathrm{P}} = 18{,}40 - 2{,}3471 \times 2{,}3232 = 12{,}94$

Deuxième lot: $x_{\mathrm{L}2} = \bar{x}_2 - k\left(n_2;\nu;p;1-\alpha\right) \times s_{\mathrm{P}} = 14{,}10 - 2{,}3471 \times 2{,}3232 = 8{,}64$

Troisième lot: $x_{\mathrm{L}3} = \bar{x}_3 - k\left(n;\nu;p;1-\alpha\right) \times s_{\mathrm{P}} = 10{,}70 - 2{,}3471 \times 2{,}3232 = 4{,}66$

Quatrième lot: $x_{\mathrm{L}4} = \bar{x}_4 - k\left(n;\nu;p;1-\alpha\right) \times s_{\mathrm{P}} = 10{,}10 - 2{,}3471 \times 2{,}3232 = 4{,}06$

Si les limites statistiques de dispersion supérieures étaient requises, les mêmes grandeurs devraient être combinées, excepté que le produit de la constante par l'erreur-type devrait être ajouté à la moyenne estimée.

5.6 Exemple 4: Intervalles statistiques de dispersion bilatéraux pour des populations séparées ayant une variance commune inconnue

Cas 1 — Calcul pour tous les lots ($m = 4$)

