NORME INTERNATIONALE

ISO 16269-8

Première édition
2004-09-15

Interprétation statistique des données —

Partie 8: **Détermination des intervalles de prédiction**

Statistical interpretation of data —
Part 8: Determination of prediction intervals

Numéro de référence
ISO 16269-8:2004(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 16269-8 a été élaborée par le comité technique ISO/TC 69, *Application des méthodes statistiques*.

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

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

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

— *Partie 8: Détermination des intervalles de prédiction*

Introduction

Les intervalles de prédiction sont précieux lorsqu'il est souhaité ou exigé de prédire les résultats sur un échantillon futur d'un nombre donné d'éléments discrets, à partir des résultats obtenus sur un échantillon antérieur d'éléments produits dans des conditions identiques. Ils sont particulièrement utiles pour les ingénieurs qui doivent pouvoir établir des limites sur la performance d'un nombre relativement petit d'éléments manufacturés. Cet aspect revêt une importance croissante compte tenu du passage récent, dans certaines industries, à la production à petite échelle.

Bien que le premier compte rendu sur les intervalles de prédiction et leurs applications remonte à 1973, un manque de conscience quant à leur valeur subsiste, ce qui est peut être dû en partie au fait que les travaux de recherche sont difficiles d'accès pour l'utilisateur potentiel et en partie à une confusion avec les intervalles de confiance et les intervalles statistiques de tolérance. L'objectif de la présente partie de l'ISO 16269 est donc double:

— préciser les différences entre intervalles de prédiction, intervalles de confiance et intervalles statistiques de tolérance;

— définir des procédures pour certain des types les plus utiles d'intervalles de prédiction, avec l'appui de tables détaillées qui ont été récemment calculées.

Pour des informations sur les intervalles de prédiction n'entrant pas dans le domaine d'application de la présente partie de l'ISO 16269, le lecteur peut se référer à la Bibliographie.

Interprétation statistique des données —

Partie 8: Détermination des intervalles de prédiction

1 Domaine d'application

La présente partie de l'ISO 16269 spécifie des méthodes pour déterminer des intervalles de prédiction pour une variable unique dont la loi est continue. Ces intervalles sont des étendues de valeurs de la variable, calculées à partir d'un échantillon aléatoire d'effectif n, pour lesquelles une prédiction se rapportant à un nouvel échantillon aléatoire d'effectif m de la même population peut être faite avec une confiance spécifiée.

Trois différents types de population sont considérés, à savoir:

a) à loi normale avec écart-type inconnu;

b) à loi normale avec écart-type connu;

c) à loi continue mais de forme inconnue.

Pour chacun de ces trois types de population, deux méthodes sont présentées, l'une pour les intervalles de prédiction unilatéraux, l'autre pour les intervalles de prédiction bilatéraux symétriques. Tous les cas présentent un choix entre six niveaux de confiance.

Les méthodes présentées pour les cas a) et b) peuvent aussi être utilisées pour des populations distribuées selon des lois non normales qu'il est possible de transformer à la normalité.

Pour les cas a) et b), les tableaux présentés dans la présente partie de l'ISO 16269 sont limités aux intervalles de prédiction contenant *toutes* les nouvelles valeurs échantillonnées m de la variable. Pour le cas c), les tableaux se rapportent à des intervalles de prédiction qui contiennent au moins $m - r$ valeurs sur les m valeurs suivantes, où r prend les valeurs de 0 à 10 ou de 0 à $m - 1$, la plus petite étendue étant retenue.

Pour les populations à loi normale, une procédure est également donnée pour calculer les intervalles de prédiction relatifs à la moyenne de m nouvelles observations.

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, *Statistique — Vocabulaire et symboles — Partie 2: Maîtrise statistique de la qualité*

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 de prédiction
étendue de valeurs d'une variable, déterminée à partir d'un échantillon aléatoire d'une population, de façon à pouvoir avoir un niveau de confiance spécifié tel que pas moins d'un nombre donné de valeurs d'un nouvel échantillon aléatoire de taille donnée de la même population y appartiendra

NOTE Dans ce contexte, le niveau de confiance est la proportion à long terme des intervalles construits de manière à avoir cette propriété.

3.1.2
statistiques d'ordre
valeurs d'un échantillon identifiées par leur position après avoir été rangées par ordre de grandeur non décroissant

NOTE Dans la présente partie de l'ISO 16269, les valeurs de l'échantillon données dans leur ordre de sélection sont notées $x_1, x_2, \ldots, x_n$. Après avoir été rangées par ordre non décroissant, elles sont notées $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$, où $x_{[1]} \leqslant x_{[2]} \leqslant \ldots \leqslant x_{[n]}$. Le terme «non décroissant» est utilisé de préférence à «croissant» afin d'inclure le cas où deux valeurs ou plus sont égales, au moins dans l'intervalle de l'erreur de mesure. Des entiers distincts et contigus sont assignés en indice aux valeurs de l'échantillon qui sont égales les unes aux autres lorsqu'elles sont représentées comme des statistiques d'ordre.

3.2 Symboles

a limite inférieure des valeurs de la variable dans la population

α probabilité maximale nominale que plus de r observations du nouvel échantillon aléatoire d'effectif m seront hors de l'intervalle de prédiction

b limite supérieure des valeurs de la variable dans la population

C niveau de confiance, exprimé en pourcentage: $C = 100\,(1 - \alpha)$

k coefficient d'intervalle de prédiction

m effectif du nouvel échantillon aléatoire auquel la prédiction s'applique

n effectif de l'échantillon aléatoire à partir duquel l'intervalle de prédiction est calculé

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

r nombre maximal spécifié d'observations du nouvel échantillon aléatoire d'effectif m qui seront hors de l'intervalle de prédiction

T_1 limite inférieure de prédiction

T_2 limite supérieure de prédiction

x_i $i^{\text{ème}}$ observation dans un échantillon aléatoire

$x_{[i]}$ statistique de i ème ordre

$\bar{x}$ moyenne de l'échantillon: $\bar{x} = \sum_{i=1}^{n} x_i / n$

4 Intervalles de prédiction

4.1 Généralités

Un *intervalle de prédiction bilatéral* est un intervalle de la forme (T_1, T_2), où $T_1 < T_2$; T_1 et T_2 sont calculées à partir d'un échantillon aléatoire d'effectif n et sont appelées respectivement *limite inférieure de prédiction* et *limite supérieure de prédiction*.

Si a et b sont respectivement les limites inférieure et supérieure de la variable dans la population, un *intervalle de prédiction unilatéral* sera de la forme (T_1, b) ou (a, T_2).

NOTE 1 Pour des raisons pratiques, a est souvent pris égal à zéro pour les variables qui ne peuvent être négatives et b est souvent pris égal à l'infini pour les variables n'ayant pas de limite supérieure naturelle.

NOTE 2 Parfois, une population est traitée comme normale dans le but de déterminer un intervalle de prédiction, même si elle a une limite finie. Cela peut paraître incohérent, car la loi normale s'étend de l'infini négatif à l'infini positif. Toutefois, dans la pratique, beaucoup de populations ayant une limite finie sont approximées par une distribution normale.

La signification pratique d'un intervalle de prédiction se rapportant à des valeurs d'un échantillon individuel est la suivante: l'expérimentateur déclare qu'un nouvel échantillon de m valeurs prélevé au hasard dans la même population aura au plus r valeurs hors de l'intervalle, tout en admettant une petite probabilité nominale que cette assertion puisse être erronée. La probabilité nominale qu'un intervalle construit de la sorte satisfasse à l'assertion est appelée niveau de confiance.

La signification pratique d'un intervalle de prédiction se rapportant à la moyenne d'un échantillon est la suivante: l'expérimentateur déclare que la *moyenne* d'un nouvel échantillon de m valeurs prélevé au hasard dans la même population sera comprise dans l'intervalle, tout en admettant une petite probabilité nominale que cette assertion puisse être erronée. De nouveau, la probabilité nominale qu'un intervalle construit de la sorte satisfasse à l'assertion est appelée niveau de confiance.

La présente partie de l'ISO 16269 présente des procédures applicables à une population de distribution normale pour $r = 0$ et des procédures applicables à la moyenne d'un nouvel échantillon d'une population de distribution normale. Elle fournit aussi des procédures applicables à des populations dont la loi est de forme inconnue pour $r = 0, 1, \ldots, 10$ ou $r = 0$ à $m - 1$, la plus petite étendue étant retenue. Dans tous les cas, les tableaux présentent des coefficients d'intervalle de prédiction ou des effectifs d'échantillon qui donnent *au moins* le niveau déclaré de confiance. En général, le niveau de confiance réel est légèrement supérieur au niveau déclaré.

Les limites des intervalles de prédiction pour des populations de distribution normale sont situées à k fois l'écart-type de l'échantillon (ou l'écart-type de la population lorsque ce dernier est connu) à partir de la moyenne de l'échantillon, k étant le coefficient d'intervalle de prédiction. Dans le cas où l'écart-type de la population n'est pas connu, la valeur de k devient très importante pour de petites valeurs de n en combinaison avec de grandes valeurs de m et des niveaux de confiance élevés. Il convient d'éviter chaque fois que possible d'utiliser de grandes valeurs de k, par exemple supérieures à 10 ou à 15, car les intervalles de prédiction résultants seront probablement trop étendus pour être d'une quelconque utilité, autre que d'indiquer que l'échantillon initial était trop petit pour donner une information utile sur des valeurs futures. De plus, pour de grandes valeurs de k, il se peut que l'intégrité des intervalles de prédiction résultants soit fortement compromise par des écarts, même minimes, par rapport à la normalité. Les valeurs de k jusqu'à 250 sont incluses dans les tableaux principalement pour montrer avec quelle rapidité k décroît à mesure que l'effectif n de l'échantillon initial croît.

Pour les intervalles de prédiction se rapportant aux valeurs individuelles dans un nouvel échantillon, il est possible d'utiliser le Formulaire A pour organiser les calculs relatifs à une population de distribution normale, et le Formulaire C lorsque la population a une loi de forme inconnue. Le Formulaire B est donné pour aider à calculer un intervalle de prédiction relatif à la moyenne d'un nouvel échantillon d'une population de distribution normale.

Les Annexes A à D fournissent des tableaux de coefficients d'intervalle de prédiction. Les Annexes E et F fournissent des tableaux d'effectifs d'échantillon quand la population est d'une forme de distribution non connue. L'Annexe G donne la procédure d'interpolation dans les tableaux, lorsque la combinaison requise de n, de m et du niveau de confiance n'est pas représentée dans les tableaux. L'Annexe H présente la théorie statistique sous-jacente aux tableaux.

4.2 Comparaison avec d'autres types d'intervalles statistiques

4.2.1 Choix du type d'intervalle

Dans la pratique, des prédictions sont souvent exigées pour un nombre *fini* d'observations sur la base des résultats obtenus sur un échantillon aléatoire initial. Telles sont les circonstances dans lesquelles la présente partie de l'ISO 16269 est appropriée. Il existe parfois une confusion avec d'autres types d'intervalles statistiques. Les paragraphes 4.2.2 et 4.2.3 sont présentés pour préciser les distinctions.

4.2.2 Comparaison avec un intervalle statistique de tolérance

Un intervalle de prédiction pour des valeurs d'un échantillon individuel est un intervalle, calculé à partir d'un échantillon aléatoire d'une population, pour lequel un énoncé de confiance est possible concernant le *nombre* maximal de valeurs dans un nouvel *échantillon* aléatoire de la population qui seront hors de l'intervalle. Un intervalle statistique de tolérance (tel que défini dans l'ISO 16269-6) est aussi un intervalle calculé à partir d'un échantillon aléatoire d'une population pour lequel un énoncé de confiance est possible; toutefois, l'énoncé se rapporte dans ce cas à la *proportion* maximale de valeurs de la *population* situées hors de l'intervalle (ou, de façon équivalente, à la proportion *minimale* de valeurs de la population situées *dans* l'intervalle).

NOTE 1 Une constante d'intervalle statistique de tolérance est la limite d'une constante d'intervalle de prédiction, puisque l'effectif m de l'échantillon futur tend vers l'infini, tandis que le nombre, r, d'individus dans l'échantillon futur en dehors de l'intervalle reste une proportion constante de m, fourni par $r > 0$. Cela est illustré dans le Tableau 1 pour un niveau de confiance de 95 % pour des intervalles unilatéraux et bilatéraux quand r/m = 0,1.

Cependant, il n'y a pas une telle analogie entre les constantes d'intervalle statistique de tolérance et les constantes d'intervalle de prédiction pour r = 0, cas sur lequel se focalise principalement la présente partie de l'ISO 16269.

Tableau 1 — Exemples de constantes d'intervalles de prédiction

r	1	2	5	10	20	50	100	1 000	Constantes d'intervalle statistique de tolérance pour une proportion minimale de 0,9 pour la population couverte
m	10	20	50	100	200	500	1 000	10 000	
	Constantes d'intervalle de prédiction								
Intervalles unilatéraux	1,887	1,846	1,767	1,718	1,686	1,663	1,655	1,647	1,646
Intervalles bilatéraux	2,208	2,172	2,103	2,061	2,034	2,014	2,007	2,000	2,000

NOTE 2 Le cas r = 0 est particulièrement important dans les applications relatives à la sécurité.

4.2.3 Comparaison avec un intervalle de confiance relatif à la moyenne

L'intervalle de prédiction relatif à une moyenne est un intervalle, calculé à partir d'un échantillon aléatoire d'une population, pour lequel il peut être affirmé avec un niveau donné de confiance que la moyenne d'un nouvel *échantillon* aléatoire d'effectif spécifié sera comprise dans cet intervalle. Un intervalle de confiance relatif à une moyenne (tel que défini dans l'ISO 2602) est aussi un intervalle calculé à partir d'un échantillon aléatoire d'une population pour lequel un énoncé de confiance est possible; toutefois, l'énoncé se rapporte dans ce cas à la moyenne de la *population*.

5 Intervalles de prédiction relatifs à toutes les observations d'un nouvel échantillon d'une population de distribution normale dont l'écart-type est inconnu

5.1 Intervalles unilatéraux

Un intervalle de prédiction unilatéral se rapportant à une population de distribution normale dont l'écart-type est inconnu est de la forme $(\bar{x} - ks, b)$ ou $(a, \bar{x} - ks)$, où les valeurs de la moyenne de l'échantillon, $\bar{x}$, et de l'écart-type de l'échantillon, s, sont déterminées à partir d'un échantillon d'effectif n prélevé au hasard dans la population. Le coefficient d'intervalle de prédiction, k, dépend de n, de l'effectif du nouvel échantillon, de m, et du niveau de confiance, C; les valeurs de k sont présentées dans l'Annexe A.

EXEMPLE On sait par l'expérience qu'une loi normale est une approximation étroite des pressions causées dans les tubes de canons par des obus d'artillerie d'un type donné. Un échantillon de 20 unités a une pression moyenne de 562,3 MPa et l'écart-type de la pression est de 8,65 MPa. Un lot de 5 000 nouvelles unités au total doit être produit dans des conditions de fabrication identiques. Quelle pression de tube de canon ne sera, à un niveau de confiance de 95 %, dépassée par aucun des 5 000 obus tirés dans des conditions identiques?

Le Tableau A.2 donne des coefficients d'intervalle de prédiction au niveau de confiance de 95 %. Dans ce tableau, le coefficient d'intervalle de prédiction correspondant est $k = 5{,}251$. La limite supérieure d'un intervalle de prédiction unilatéral à un niveau de confiance de 95 % est donc

$$\bar{x} + ks = 562{,}3 + 5{,}251 \times 8{,}65 = 607{,}7 \text{ MPa}$$

Il peut donc y avoir confiance à 95 % qu'aucune des 5 000 unités du lot ne produira une pression dans le tube de canon supérieure à 607,7 MPa.

Cet exemple sert aussi à illustrer l'utilisation du Formulaire A.

5.2 Intervalles bilatéraux symétriques

Un intervalle de prédiction bilatéral symétrique pour une population de distribution normale dont l'écart-type est inconnu est de la forme $(\bar{x} - ks, \bar{x} + ks)$. Le coefficient d'intervalle de prédiction, k, dépend de n, de l'effectif m du nouvel échantillon et du niveau de confiance, C; les valeurs de k sont présentées dans l'Annexe B.

EXEMPLE On sait par l'expérience que les temps jusqu'à détonation des grenades à main dégoupillées d'un type particulier suivent une loi normale approximative. Un échantillon aléatoire d'effectif 30 est prélevé et testé et les temps jusqu'à détonation sont enregistrés. Le temps jusqu'à détonation moyen de l'échantillon est de 5,140 s et l'écart-type de l'échantillon est de 0,241 s. Un intervalle de prédiction bilatéral symétrique est exigé pour l'ensemble du lot suivant de 10 000 grenades à un niveau de confiance de 99 %.

Le Tableau B.4 donne les coefficients d'intervalle de prédiction au niveau de confiance 99 %. Dans ce tableau, les entrées n = 30 et m = 10 000 donnent la valeur k = 6,059. L'intervalle de prédiction symétrique est le suivant:

$$(\bar{x} - ks,\ \bar{x} + ks) = (5{,}140 - 6{,}059 \times 0{,}241,\ 5{,}140 + 6{,}059 \times 0{,}241) = (3{,}68,\ 6{,}60)$$

Il peut donc y avoir confiance à 99 % qu'aucune des 10 000 grenades suivantes n'aura un temps jusqu'à détonation hors de l'étendue comprise entre 3,68 s et 6,60 s.

5.3 Intervalles de prédiction relatifs à des populations non normales qui peuvent être transformées à la normalité

Pour les populations non normales qui peuvent être transformées à la normalité, les procédures relatives aux populations normales sont d'abord appliquées aux données transformées; l'intervalle de prédiction est ensuite établi en appliquant la transformation inverse aux limites de prédiction résultantes.

EXEMPLE Pour les données de l'exemple en 5.2, il est supposé que les temps jusqu'à détonation suivent approximativement une loi logarithmique normale, à savoir que les logarithmes des temps jusqu'à détonation suivent approximativement une loi normale. Les temps jusqu'à détonation de l'échantillon, $x_1, x_2, \ldots, x_n$, sont par conséquent transformés à la normalité en prenant leur logarithme naturel, soit $y_i = \ln x_i$ pour $i = 1, 2, \ldots, 30$. Il est supposé que la moyenne de l'échantillon pour les données transformées est $\bar{y}$ = 1,60 et que l'écart-type de l'échantillon est s_y = 0,05. Le coefficient d'intervalle de prédiction pour une confiance à 99 % qu'aucun des 10 000 temps jusqu'à détonation suivants ne sera hors d'un intervalle bilatéral demeure, bien entendu, inchangé à k = 6,059. L'intervalle de prédiction symétrique relatif aux données transformées est le suivant:

$$(\bar{y} - ks_y,\ \bar{y} + ks_y) = (1{,}60 - 6{,}059 \times 0{,}05,\ 1{,}60 + 6{,}059 \times 0{,}05) = (1{,}297,\ 1{,}903)$$

Les unités de mesure de y sont des logarithmes de secondes. La transformation inverse pour convertir ces unités en secondes est l'exponentiation. L'intervalle de prédiction à un niveau de confiance de 99 % pour le temps jusqu'à détonation de l'ensemble des 10 000 grenades suivantes est donc:

$$\left(\mathrm{e}^{1{,}297},\ \mathrm{e}^{1{,}903}\right) = (3{,}66,\ 6{,}71)\ \mathrm{s}$$

NOTE 1 Le même résultat serait obtenu en utilisant des logarithmes à n'importe quelle autre base, pour autant que l'antilogarithme à même base soit utilisé lors de la conversion aux unités d'origine.

NOTE 2 Quand un intervalle de prédiction bilatéral est déterminé conformément à 5.2 ou à 6.2, ses limites des populations de distribution normale sont symétriques par rapport (c'est-à-dire équidistants) à la médiane estimée de la population. Cette symétrie est perdue pour les populations de distribution non normale qui sont transformées à la normalité conformément à 5.3 ou à 6.3.

5.4 Détermination d'un effectif approprié, n, de l'échantillon initial, pour une valeur maximale donnée du coefficient d'intervalle de prédiction, k

Le niveau de confiance, l'effectif m du nouvel échantillon et la valeur approximative souhaitée du coefficient d'intervalle de prédiction sont parfois donnés et il est exigé de déterminer l'effectif n de l'échantillon initial. Le tableau correspondant au niveau de confiance donné et au type d'intervalle de prédiction est repéré (c'est-à-dire un des tableaux de l'Annexe A pour un intervalle unilatéral ou un des tableaux de l'Annexe B pour un intervalle bilatéral) et la colonne pour la valeur donnée de m est identifiée. Dans cette colonne, identifier la première valeur de k qui n'est pas supérieure au maximum donné. Sur cette ligne, la valeur de n donnée dans la première colonne à gauche du tableau est l'effectif requis de l'échantillon initial.

NOTE Si l'entrée en bas de cette colonne dépasse la valeur acceptable maximale de k, il n'y a pas d'effectif de l'échantillon initial assez grand pour satisfaire à l'exigence. Il convient donc d'envisager une réduction du niveau de confiance.

EXEMPLE Considérer une situation d'échantillonnage pour acceptation dans laquelle, avant l'utilisation de la présente partie de l'ISO 16269, la pratique était d'accepter des lots d'effectif 5 000 chaque fois que $\bar{x} + 4{,}75\, s \leqslant 0{,}1$, où x est la porosité en loi normale d'un composant fritté et $\bar{x}$ et s sont la moyenne de l'échantillon et l'écart-type de l'échantillon sur la base d'un échantillon d'effectif 30 prélevé au hasard dans une population de distribution normale. Il est supposé qu'il a été décidé de remplacer ce critère d'acceptation par un critère qui donnera une confiance à 95 % qu'aucun des articles du lot n'a $x > 0{,}1$. Le producteur déclare que ce critère d'acceptation le satisfera pour autant que le coefficient d'intervalle de prédiction ne soit pas supérieur à la valeur 4,75 à laquelle il est habitué et sous réserve que l'exigence relative à l'effectif de l'échantillon ne soit pas excessive.

À la troisième page du Tableau A.2, la colonne pour m égal à 5 000 est identifiée. Dans cette colonne, pour un échantillon d'effectif 40, k = 4,771, mais, pour un échantillon d'effectif 45, k = 4,717, ce qui est inférieur à 4,75. Le producteur accepte d'augmenter l'effectif de l'échantillon en le faisant passer à 45 avec k = 4,717 pour des lots futurs.

5.5 Détermination de l'intervalle de confiance correspondant à un intervalle de prédiction donné

Plutôt que de déterminer l'intervalle de prédiction correspondant à un niveau de confiance donné, il est parfois exigé de déterminer, à partir de l'échantillon initial, le niveau de confiance correspondant à un intervalle spécifié. Il peut s'agir d'un intervalle unilatéral $(\bar{x} - ks, b)$ ou $(a, \bar{x} + ks)$, ou d'un intervalle bilatéral $(\bar{x} - ks, \bar{x} + ks)$ qui est symétrique par rapport à la moyenne de l'échantillon.

La valeur de k correspondant à l'intervalle de prédiction souhaité est d'abord calculée. Il est ensuite possible d'établir le niveau de confiance correspondant à cet intervalle, par interpolation à partir des valeurs représentées dans les tableaux, comme spécifié en G.1.4.

6 Intervalles de prédiction pour toutes les observations d'un nouvel échantillon d'une population de distribution normale dont l'écart-type est connu

6.1 Intervalles unilatéraux

Un intervalle de prédiction unilatéral relatif à une population de distribution normale dont l'écart-type σ est connu est de la forme $(\bar{x} - k\sigma, b)$ ou $(a, \bar{x} + k\sigma)$. Le coefficient d'intervalle de prédiction, k, dépend de n, de l'effectif m du nouvel échantillon et du niveau de confiance, C; les valeurs de k sont présentées dans l'Annexe C.

EXEMPLE Des tubes en argile vitrifiée de 150 mm de diamètre fabriqués par un procédé donné ont des longueurs qui suivent une loi normale avec un écart-type de 4,49 mm. Il a été établi que la longueur moyenne d'un échantillon de 50 tubes est de 1 760,60 mm. À un niveau de confiance de 99 %, quelle longueur les 1 000 tubes suivants dépasseront-ils tous?

Dans le Tableau C.4, pour n = 50 et m = 1 000, le coefficient d'intervalle de prédiction correspondant est k = 4,306. La limite inférieure de prédiction pour l'ensemble des 1 000 longueurs suivantes est donc:

$$\bar{x} - k\sigma = 1\,760{,}60 - 4{,}306 \times 4{,}49 = 1\,741$$

Ainsi, il y a confiance à 99 % qu'aucune des longueurs des 1 000 tubes suivants ne sera inférieure à 1 741 mm.

Ce type d'information pourrait s'avérer utile si le fabricant envisageait d'offrir une garantie pour sa production. Par exemple, le fabricant pourrait en l'occurrence garantir avec une bonne assurance une longueur d'au moins 1 740 mm.

6.2 Intervalles bilatéraux symétriques

Un intervalle de prédiction bilatéral symétrique pour une population de distribution normale dont l'écart-type σ est connu est de la forme $(\bar{x} - k\sigma, \bar{x} + k\sigma)$. Le coefficient d'intervalle de prédiction, k, dépend de n, de l'effectif m du nouvel échantillon et du niveau de confiance, C; les valeurs de k sont présentées dans l'Annexe D.

EXEMPLE Il est supposé que, pour les données de l'exemple en 6.1, il est exigé de calculer un intervalle de prédiction bilatéral pour l'ensemble des 10 000 longueurs suivantes à un niveau de confiance de 95 %. Le tableau approprié pour un niveau de confiance de 95 % est le Tableau D.2. Dans ce dernier, pour n = 50 et m = 10 000, le coefficient d'intervalle de prédiction bilatéral est égal à 4,605. L'intervalle de prédiction est le suivant:

$$(\bar{x} - k\sigma,\ \bar{x} + k\sigma) = (1\,760{,}60 - 4{,}605 \times 4{,}49,\ 1\,760{,}60 + 4{,}605 \times 4{,}49) = (1\,739{,}9,\ 1\,781{,}3)$$

Il peut donc y avoir confiance à 95 % que les 10 000 tubes suivants ont tous des longueurs comprises entre 1 739,9 mm et 1 781,3 mm.

6.3 Intervalles de prédiction pour des populations non normales qui peuvent être transformées à la normalité

Pour les populations non normales qui peuvent être transformées à la normalité, les procédures pour déterminer un intervalle de prédiction lorsque l'écart-type de la population est connu sont similaires à celles utilisées dans les cas où l'écart-type de la population est inconnu et qui sont décrites en 5.3. Les procédures pour les populations normales sont d'abord appliquées aux données transformées; puis l'intervalle de prédiction est établi en appliquant la transformation inverse aux limites de prédiction résultantes.

EXEMPLE On sait que la longévité à la fatigue d'un composant structurel d'avion suit une loi approximativement logarithmique normale, à savoir que les logarithmes des temps jusqu'à rupture peuvent être considérés comme suivant une loi normale. On sait aussi par l'expérience que l'écart-type de $\log_{10}$(longévité) est approximativement égal à 0,11. Six échantillons du composant sont soumis à un essai de fatigue et le nombre de cycles de chargement jusqu'à la rupture est enregistré comme suit:

229 200; 277 900; 332 400; 369 700; 380 800; 406 300.

Deux nouveaux composants doivent être fabriqués selon la même spécification. Combien de cycles de chargement est-il possible d'appliquer avec une confiance à 99,9 % qu'aucun de ces deux composants ne connaîtra une rupture?

En prenant les logarithmes décimaux, la moyenne de $x = \log_{10}$(longévité) est $\bar{x} = 5{,}513\,86$. Dans le Tableau C.6, pour $n = 6$ et $m = 2$, le coefficient d'intervalle de prédiction unilatéral correspondant est $k = 3{,}554$. La limite de prédiction inférieure pour les deux composants suivants est donc:

$$\bar{x} - k\sigma = 5{,}513\,86 - 3{,}554 \times 0{,}11 = 5{,}122\,92$$

et, par les antilogarithmes, $10^{5{,}122\,92} = 132\,715$.

Ainsi, compte tenu du fait que l'écart-type de $\log_{10}$(longévité) n'est connu qu'à deux chiffres significatifs, il peut y avoir confiance à 99,9 % que les deux composants nouveaux supporteront avec succès au moins 130 000 cycles de chargement.

6.4 Détermination d'un effectif approprié, n, de l'échantillon initial pour une valeur donnée de k

La procédure est la même que celle décrite en 5.4, à l'exception du fait que l'Annexe C ou D est utilisée au lieu de l'Annexe A ou B.

6.5 Détermination du niveau de confiance correspondant à un intervalle de prédiction donné

Il est possible de calculer, à partir des Annexes C et D, le niveau de confiance correspondant soit à un intervalle unilatéral $(\bar{x} - k\sigma, b)$ ou $(a, \bar{x} + k\sigma)$, soit à un intervalle bilatéral $(\bar{x} - k\sigma, \bar{x} + k\sigma)$ qui est symétrique par rapport à la moyenne de l'échantillon.

La valeur de k correspondant à l'intervalle de prédiction souhaité est d'abord calculée. Il est ensuite possible d'établir le niveau de confiance pour cet intervalle par interpolation entre les valeurs représentées dans les Tableaux, comme décrit en G.1.4.

7 Intervalles de prédiction relatifs à la moyenne d'un nouvel échantillon d'une population de distribution normale

Il est possible d'appliquer un processus simple en deux étapes pour obtenir le coefficient d'intervalle de prédiction pour la moyenne d'un nouvel échantillon de m observations de la même population de distribution normale, en utilisant les mêmes tableaux. Le coefficient d'intervalle de prédiction correspondant à une observation future simple est d'abord identifié. Puis ce coefficient d'intervalle de prédiction est multiplié par $\sqrt{(n+m)/[m(n+1)]}$. Cette procédure s'applique aux intervalles unilatéraux et bilatéraux et aux cas où l'écart-type de la population est connu ou inconnu.

EXEMPLE Pour les données de l'exemple donné en 6.1, on suppose qu'il est exigé de fournir une limite inférieure de prédiction avec un niveau de confiance de 99 % sur la moyenne des longueurs des 1 000 tubes suivants. Dans le Tableau C.4, le coefficient d'intervalle de prédiction pour une taille d'échantillon initiale de 50 et une observation future simple est 2,350. Le coefficient d'intervalle de prédiction exigé est donc:

$$k = 2{,}350 \times \sqrt{(n+m)/[m(n+1)]} = 2{,}350 \times \sqrt{1\,050/51\,000} = 0{,}337\,2$$

Il s'ensuit que la limite inférieure de prédiction pour la longueur moyenne des 1 000 tubes suivants est:

$$\bar{x} - k\sigma = 1\,760{,}60 - 0{,}337\,2 \times 4{,}49 = 1\,759 \text{ mm}$$

Cet exemple sert à illustrer l'utilisation du Formulaire B.

8 Intervalles de prédiction non paramétriques

8.1 Généralités

Lorsque la variable est continue mais la loi de la population est inconnue, il convient d'utiliser des méthodes non paramétriques pour produire des intervalles de prédiction. Ces méthodes sont fondées sur les statistiques d'ordre $x_{[1]}, x_{[2]}, \ldots, x_{[n]}$. En général, les intervalles de prédiction non paramétriques unilatéraux sont de la forme $(x_{[i]}, b)$ ou $(a, x_{[i]})$ où $1 \leqslant i \leqslant n$, alors que les intervalles de prédiction bilatéraux non paramétriques sont de la forme $(x_{[i]}, x_{[j]})$ où $1 \leqslant i \leqslant j \leqslant n$. La présente partie de l'ISO 16269 présente des procédures pour les intervalles unilatéraux de la forme $(x_{[1]}, b)$ ou $(a, x_{[n]})$ et pour les intervalles bilatéraux de la forme $(x_{[1]}, x_{[n]})$.

Le problème que posent de tels intervalles est de déterminer quelle doit être l'importance de l'effectif de l'échantillon initial pour que l'on puisse avoir la confiance requise que l'intervalle de prédiction contiendra au moins $m - r$ des m valeurs suivantes. Les Annexes E et F sont données à cette fin.

8.2 Intervalles unilatéraux

Les Tableaux E.1 à E.6 présentent des effectifs n d'échantillons initiaux à partir desquels il est possible d'avoir un niveau de confiance, C, que l'intervalle de prédiction non paramétrique unilatéral $(x_{[1]}, b)$ [ou alternativement $(a, x_{[n]})$] inclura au moins $m - r$ valeurs d'un nouvel échantillon de m valeurs de la même population, pour une étendue donnée des valeurs de C, m et r.

EXEMPLE Une limite inférieure de prédiction non paramétrique est exigée pour la résistance à la flexion de tubes en argile vitrifiée, afin qu'il puisse y avoir confiance à 90 % que 10 tubes au maximum de chaque nouveau lot de 200 tubes auront une résistance inférieure. Quel effectif de l'échantillon initial est-il exigé?

Le Tableau E.1 donne les effectifs des échantillons initiaux pour un niveau de confiance de 90 %. Dans ce tableau, pour $m = 200$ et $r = 10$, l'effectif d'échantillon correspondant est $n = 46$. Un échantillon de 46 tubes est prélevé au hasard et soumis à un essai de résistance à la flexion. La résistance la plus faible est établie à 6,4 kN·m. Ainsi, il peut y avoir confiance à 90 % que, pour des tubes fabriqués dans des conditions identiques à celles de l'échantillon initial, 10 tubes au maximum dans chaque lot de 200 tubes auront une résistance à la flexion inférieure à 6,4 kN·m.

8.3 Intervalles bilatéraux

Les Tableaux F.1 à F.6 présentent des effectifs n d'échantillons initiaux à partir desquels il est possible d'avoir un niveau de confiance, C, que l'intervalle de prédiction non paramétrique bilatéral $(x_{[1]}, x_{[n]})$ inclura au moins $m - r$ valeurs d'un nouvel échantillon de m valeurs de la même population, pour une étendue donnée des valeurs de C, m et r.

EXEMPLE Un fournisseur propose des batteries d'automobile en lots de 100, et souhaite donner une forme de garantie à ses détaillants concernant l'étendue des valeurs de la tension, x, dans chaque lot. Comme il n'est pas certain de la loi que suivent les tensions, il décide d'utiliser une approche non paramétrique. Quel effectif de l'échantillon initial lui serait-il nécessaire pour être confiant à 90 % qu'une batterie au maximum dans chaque lot a une tension non comprise dans l'étendue des tensions de son échantillon?

Le Tableau F.1 donne les effectifs d'échantillons initiaux pour un niveau de confiance de 90 %. Dans ce tableau, pour $m = 100$ et $r = 1$, l'effectif de l'échantillon initial est $n = 410$. Le fournisseur soumet à l'essai 410 batteries et établit que la tension la plus faible est $x_{[1]} = 11{,}81$ V et la tension la plus forte est $x_{[410]} = 12{,}33$ V. Il garantit par conséquent qu'une batterie au maximum par lot a une tension non comprise entre 11,81 V et 12,33 V.

Cet exemple sert aussi à illustrer l'utilisation du Formulaire C. Noter que si le fournisseur souhaitait avoir une confiance à 90 % qu'*aucune* batterie dans des lots de 100 n'a une tension hors de ces limites, les limites devraient être fondées sur un échantillon de 1 850 batteries, c'est-à-dire plus de quatre fois plus important que l'échantillon choisi.