

---

---

**Méthodes statistiques utilisées dans  
les essais d'aptitude par comparaison  
interlaboratoires**

*Statistical methods for use in proficiency testing by interlaboratory  
comparison*

iTeh STANDARD PREVIEW  
(standards.iteh.ai)

ISO 13528:2022

<https://standards.iteh.ai/catalog/standards/sist/686650e2-6a8a-4cc6-ba16-3c53b91c086e/iso-13528-2022>



iTeh STANDARD PREVIEW  
(standards.iteh.ai)

ISO 13528:2022

<https://standards.iteh.ai/catalog/standards/sist/686650e2-6a8a-4cc6-ba16-3c53b91c086e/iso-13528-2022>



**DOCUMENT PROTÉGÉ PAR COPYRIGHT**

© ISO 2022

Tous droits réservés. Sauf prescription différente ou nécessité dans le contexte de sa mise en œuvre, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie, ou la diffusion sur l'internet ou sur un intranet, sans autorisation écrite préalable. Une autorisation peut être demandée à l'ISO à l'adresse ci-après ou au comité membre de l'ISO dans le pays du demandeur.

ISO copyright office  
Case postale 401 • Ch. de Blandonnet 8  
CH-1214 Vernier, Genève  
Tél.: +41 22 749 01 11  
E-mail: [copyright@iso.org](mailto:copyright@iso.org)  
Web: [www.iso.org](http://www.iso.org)

Publié en Suisse

## Sommaire

Page

Avant-propos .....	v
<b>0 Introduction</b> .....	<b>vi</b>
<b>1 Domaine d'application</b> .....	<b>1</b>
<b>2 Références normatives</b> .....	<b>1</b>
<b>3 Termes et définitions</b> .....	<b>1</b>
<b>4 Principes généraux</b> .....	<b>5</b>
4.1 Exigences générales relatives aux méthodes statistiques .....	5
4.2 Modèle de base .....	5
4.3 Approches générales pour l'évaluation des performances .....	6
<b>5 Lignes directrices pour le plan statistique des programmes d'essais d'aptitude</b> .....	<b>6</b>
5.1 Présentation du plan statistique de programmes d'essais d'aptitude .....	6
5.2 Base d'un plan statistique .....	6
5.3 Considérations sur la distribution statistique des résultats .....	7
5.4 Considérations sur les petits nombres de participants .....	8
5.5 Lignes directrices pour choisir le format de rendu .....	9
5.5.1 Exigences générales relatives au format de rendu .....	9
5.5.2 Consignation des mesurages répétés .....	9
5.5.3 Consignation sous la forme «inférieur à» ou «supérieur à» une limite (données tronquées) .....	10
5.5.4 Nombre de chiffres significatifs .....	10
<b>6 Lignes directrices pour la revue initiale des entités soumises à l'essai d'aptitude et des résultats</b> .....	<b>11</b>
6.1 Homogénéité et stabilité des entités soumises à l'essai d'aptitude .....	11
6.2 Considérations sur les différentes méthodes de mesure .....	12
6.3 Élimination des valeurs aberrantes .....	13
6.4 Examen visuel des données .....	13
6.5 Méthodes statistiques robustes .....	13
6.6 Techniques de détection des valeurs aberrantes pour des résultats individuels .....	14
<b>7 Détermination de la valeur assignée et de son incertitude-type</b> .....	<b>15</b>
7.1 Choix de la méthode de détermination de la valeur assignée .....	15
7.2 Détermination de l'incertitude de la valeur assignée .....	16
7.3 Formulation .....	17
7.4 Matériau de référence certifié .....	18
7.5 Résultats provenant d'un seul laboratoire .....	18
7.6 Valeur consensuelle déterminée par des laboratoires experts .....	19
7.7 Valeur consensuelle déterminée à partir des résultats des participants .....	20
7.8 Comparaison de la valeur assignée avec une valeur de référence indépendante .....	21
<b>8 Détermination des critères d'évaluation des performances</b> .....	<b>22</b>
8.1 Approches pour la détermination des critères d'évaluation .....	22
8.2 Détermination par perception d'experts .....	23
8.3 Détermination à partir de l'expérience acquise lors des campagnes précédentes d'un programme d'essais d'aptitude .....	23
8.4 Détermination par l'utilisation d'un modèle général .....	24
8.5 Utilisation des écarts-types de répétabilité et de reproductibilité obtenus lors d'une étude collaborative de la fidélité d'une méthode de mesure .....	24
8.6 Détermination à partir de données obtenues lors d'une même campagne d'un programme d'essais d'aptitude .....	25
8.7 Surveillance de la concordance interlaboratoires .....	26
<b>9 Calcul des statistiques de performances</b> .....	<b>26</b>
9.1 Considérations générales pour la détermination des performances .....	26
9.2 Limitation de l'incertitude de la valeur assignée .....	27

9.3	Estimations des écarts (erreur de mesure).....	28
9.4	Scores $z$ .....	29
9.5	Scores $z'$ .....	30
9.6	Scores zêta ( $\zeta$ ).....	31
9.7	Scores $E_n$ .....	32
9.8	Évaluation des incertitudes d'essais des participants.....	33
9.9	Scores composés de performance.....	34
<b>10</b>	<b>Méthodes graphiques de description des scores de performance</b> .....	<b>35</b>
10.1	Application de méthodes graphiques.....	35
10.2	Histogrammes de résultats ou de scores de performance.....	35
10.3	Diagrammes de densité par la méthode du noyau.....	36
10.4	Diagrammes en bâtons des scores de performance normalisés.....	38
10.5	Graphique de Youden.....	38
10.6	Graphiques d'écarts-types de répétabilité.....	39
10.7	Échantillons fractionnés.....	40
10.8	Méthodes graphiques de combinaison des scores de performance sur plusieurs campagnes d'un programme d'essais d'aptitude.....	41
<b>11</b>	<b>Plan et analyse de programmes d'essais d'aptitude qualitatifs (y compris les propriétés nominales et ordinales)</b> .....	<b>42</b>
11.1	Types de données qualitatives.....	42
11.2	Plan statistique.....	42
11.3	Valeurs assignées pour les programmes d'essais d'aptitude.....	43
11.4	Évaluation des performances et attribution de scores pour les programmes d'essais d'aptitude qualitatifs.....	45
	<b>Annexe A (normative) Symboles</b> .....	<b>47</b>
	<b>Annexe B (informative) Homogénéité et stabilité des entités soumises à l'essai d'aptitude</b> .....	<b>49</b>
	<b>Annexe C (informative) Analyse robuste</b> .....	<b>57</b>
	<b>Annexe D (informative) Recommandations supplémentaires concernant les procédures statistiques</b> .....	<b>68</b>
	<b>Annexe E (informative) Exemples illustratifs</b> .....	<b>73</b>
	<b>Annexe F (Informative) Exemple de code informatique pour la création de la représentation graphique et l'analyse de rééchantillonnage (bootstrapping) des résultats des essais d'aptitude</b> .....	<b>97</b>
	<b>Bibliographie</b> .....	<b>98</b>

## 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 (IEC) 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/IEC, 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/IEC, Partie 2 (voir [www.iso.org/directives](http://www.iso.org/directives)).

L'attention est attirée 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 dans la liste des déclarations de brevets reçues par l'ISO (voir [www.iso.org/brevets](http://www.iso.org/brevets)).

Les appellations commerciales éventuellement mentionnées dans le présent document sont données pour information, par souci de commodité, à l'intention des utilisateurs et ne sauraient constituer un engagement.

Pour une explication de la signification des termes et expressions spécifiques de l'ISO liés à l'évaluation de la conformité, ou pour toute information au sujet de l'adhésion de l'ISO aux principes de l'OMC concernant les obstacles techniques au commerce (OTC), voir le lien suivant: [www.iso.org/iso/fr/foreword.html](http://www.iso.org/iso/fr/foreword.html).

Le comité chargé de l'élaboration du présent document est l'ISO/TC 69, *Application des méthodes statistiques*, sous-comité SC 6, *Méthodes et résultats de mesure*.

Cette troisième édition de l'ISO 13528 annule et remplace la deuxième édition (ISO 13528:2015), qui a fait l'objet d'une révision mineure. Les modifications sont les suivantes:

- ajout de notes en [10.1](#), [10.4.3](#) et [10.5.3](#) pour attirer l'attention sur les méthodes graphiques supplémentaires pouvant contribuer à satisfaire aux dispositions de [10.1](#);
- correction des [Formules B.4](#) et [B.8](#) afin d'utiliser  $s_t^2$  au lieu de  $w_t^2$ ;
- correction de la [Formule B.16](#) de sorte que le terme intégré à la racine carrée soit toujours une valeur non négative;
- dans le [Tableau C.2](#), correction en 0,399 4 du facteur de correction associé à  $p = 2$ ;
- ajout de références documentaires supplémentaires dans la Bibliographie comme source des valeurs du [Tableau C.2](#), qui sont citées en référence dans les Notes 1 et 2 de [C.5.2.1](#);
- harmonisation des styles de police (italique ou police Roman) utilisés dans les formules, sur l'ensemble du document.

## 0 Introduction

### 0.1 Objectifs des essais d'aptitude

Les essais d'aptitude impliquent l'utilisation de comparaisons interlaboratoires pour déterminer les performances de participants (qui peuvent être des laboratoires, des organismes de contrôle ou des individus) pour des essais spécifiés ou des mesures et ainsi surveiller la continuité des performances des participants. Il existe un certain nombre d'objectifs pour les essais d'aptitude, tels que décrits dans l'Introduction de l'ISO/IEC 17043. Ces objectifs comprennent l'évaluation des performances des laboratoires, l'identification de problèmes dans les laboratoires, la détermination de l'efficacité et la comparabilité des méthodes d'essai ou de mesure, l'amélioration de la confiance des clients des laboratoires, la validation des incertitudes revendiquées et la formation des laboratoires participants. Le plan et les techniques statistiques appliqués doivent être appropriés à l'objectif ou aux objectifs définis.

### 0.2 Justification de l'évaluation par scores dans les programmes d'essais d'aptitude

Diverses stratégies d'évaluation par scores sont disponibles et utilisées pour les essais d'aptitude. Bien que les calculs détaillés diffèrent, la plupart des programmes d'essais d'aptitude comparent l'écart du participant par rapport à une valeur assignée avec un critère numérique qui est utilisé pour déterminer si cet écart est préoccupant ou non. Les stratégies utilisées pour les valeurs assignées et pour le choix d'un critère d'évaluation des écarts des participants sont donc critiques. En particulier, il est important de déterminer s'il convient que la valeur assignée et le critère d'évaluation des écarts soient indépendants des résultats des participants ou s'il convient qu'ils soient obtenus à partir des résultats soumis. Dans le présent document, les deux stratégies sont prévues. Toutefois, l'attention est attirée sur la discussion figurant aux [Articles 7](#) et [8](#) concernant les avantages et les inconvénients associés au choix de valeurs assignées ou de critères d'évaluation des écarts qui ne sont pas obtenus à partir des résultats des participants. On verra que, en général, le choix de valeurs assignées et de critères d'évaluation indépendamment des résultats des participants offre des avantages. Cela est notamment le cas pour le critère utilisé pour évaluer les écarts par rapport à la valeur assignée – tel que l'écart-type d'aptitude ou une tolérance sur l'erreur de mesure – pour lequel un choix cohérent fondé sur la pertinence pour une utilisation finale particulière des résultats de mesure est particulièrement utile.

### 0.3 ISO 13528 et ISO/IEC 17043

Le présent document soutient la mise en œuvre de l'ISO/IEC 17043, notamment les exigences relatives au plan statistique, à la validation des entités soumises à l'essai d'aptitude, à la revue des résultats et au rendu des statistiques résumées. L'Annexe B de l'ISO/IEC 17043:2010 fournit une description succincte des méthodes statistiques générales utilisées dans les programmes d'essais d'aptitude. Le présent document est destiné à être complémentaire de l'ISO/IEC 17043, en fournissant des recommandations détaillées ne figurant pas dans ce document sur des méthodes statistiques particulières pour les essais d'aptitude.

La définition de l'essai d'aptitude donnée dans l'ISO/IEC 17043 est reprise dans le présent document, avec des notes décrivant les différents types d'essais d'aptitude et l'étendue des conceptions pouvant être utilisées. Le présent document ne peut pas couvrir spécifiquement l'ensemble des objectifs, des conceptions, des matrices et des mesurandes. Les techniques présentées dans le présent document sont destinées à être appliquées de façon étendue, notamment pour des programmes d'essais d'aptitude nouvellement établis. Il est attendu que les techniques statistiques utilisées dans le cadre d'un programme d'essais d'aptitude particulier évoluent au fur et à mesure que le programme gagne en maturité; et les scores, les critères d'évaluation et les méthodes graphiques sont affinés afin de mieux répondre aux besoins spécifiques d'un groupe cible de participants, d'organismes d'accréditation et d'autorités réglementaires.

Le présent document incorpore des recommandations publiées pour les essais d'aptitude de laboratoires d'analyses chimiques<sup>[32]</sup>, mais inclut également une gamme plus étendue de procédures à utiliser avec des méthodes de mesure et d'identifications qualitatives valides. La révision du présent document contient la plupart des méthodes statistiques et recommandations issues de la première édition, complétées autant que nécessaire par les documents précédemment cités en référence et par le

domaine d'application étendu de l'ISO/IEC 17043. L'ISO/IEC 17043 comprend des essais d'aptitude pour les personnes et les organismes de contrôle, incluant l'Annexe B de l'ISO/IEC 17043:2010 qui comprend des éléments d'appréciation pour les résultats qualitatifs.

Le présent document contient des techniques statistiques qui sont conformes à d'autres Normes internationales, notamment à celles du TC69/SC6, et en particulier à la série de normes ISO 5725 portant sur l'Exactitude: justesse et fidélité. Les techniques sont également destinées à refléter d'autres Normes internationales, le cas échéant, et à être en cohérence avec le Guide ISO/IEC 98-3 (GUM) et avec le Guide ISO/IEC 99 (VIM).

#### 0.4 Expertise en statistique

L'ISO/IEC 17043 exige que, pour être compétent, un organisateur d'essais d'aptitude doit avoir accès à l'expertise statistique et doit autoriser du personnel spécifique à mener l'analyse statistique. Ni l'ISO/IEC 17043 ni le présent document ne peuvent définir plus en détail en quoi consiste cette expertise nécessaire. Pour certaines applications, un niveau élevé en statistiques est utile, mais généralement les besoins en expertise peuvent être assurés par des personnes ayant une expertise technique dans d'autres domaines, connaissant les concepts et les techniques statistiques de base et ayant une expérience ou une formation dans les techniques courantes applicables à l'analyse des données issues de programmes d'essais d'aptitude. Si un consultant est responsable du plan et/ou de l'analyse statistique, il est très important que cette personne ait une expérience dans le domaine des comparaisons interlaboratoires, même si elle a un niveau élevé dans le domaine des statistiques. La formation classique en statistiques avancées n'inclut pas l'acquisition de connaissances dans le domaine des comparaisons interlaboratoires, et les causes uniques d'erreur de mesure se produisant durant les essais d'aptitude peuvent sembler obscures. Les recommandations données dans le présent document ne peuvent pas fournir toute l'expertise nécessaire pour prendre en compte toutes les applications, et ne peuvent pas remplacer l'expérience acquise lors de la pratique des comparaisons interlaboratoires.

#### 0.5 Logiciels

Les logiciels nécessaires pour l'analyse statistique des données d'essais d'aptitude peuvent varier de façon importante, allant de l'arithmétique simple d'un tableur pour les petits programmes d'essais d'aptitude utilisant des valeurs de référence connues, jusqu'aux logiciels statistiques sophistiqués utilisés pour les méthodes statistiques faisant appel à des calculs itératifs ou d'autres méthodes numériques avancées. La plupart des techniques mentionnées dans le présent document peuvent être mises en œuvre par des tableurs classiques, éventuellement avec des routines personnalisées pour un programme d'essais d'aptitude ou une analyse spécial(e); certaines techniques nécessitent des logiciels qui sont gratuits. Dans tous les cas, il est attendu des utilisateurs qu'ils vérifient la validité et l'exactitude de leurs calculs, en particulier lorsque des programmes spéciaux ont été entrés par l'utilisateur. Toutefois, même lorsque les techniques traitées dans le présent document sont appropriées et correctement mises en œuvre par des logiciels adéquats, elles ne peuvent pas être appliquées sans la surveillance d'une personne disposant d'une expertise technique et statistique suffisante pour cerner la nature des applications et des hypothèses statistiques, et pour identifier et rechercher des anomalies susceptibles de se produire lors de toute campagne d'un programme d'essais d'aptitude.





# Méthodes statistiques utilisées dans les essais d'aptitude par comparaison interlaboratoires

## 1 Domaine d'application

Le présent document fournit des descriptions détaillées de méthodes statistiques à utiliser par les organisateurs d'essais d'aptitude pour concevoir des programmes d'essais d'aptitude et pour analyser les données obtenues à partir de tels programmes. Le présent document donne des recommandations sur l'interprétation des données d'essais d'aptitude par les participants à de tels programmes d'essais d'aptitude et par les organismes d'accréditation.

Les procédures du présent document peuvent être appliquées pour démontrer que les résultats de mesure obtenus par des laboratoires, des organismes de contrôle et des personnes remplissent les critères spécifiés pour des performances acceptables.

Le présent document est applicable à des essais d'aptitude dans lesquels les résultats consignés sont des observations quantitatives ou qualitatives sur des entités soumises à l'essai.

**NOTE** Les procédures énoncées dans le présent document peuvent être également appliquées pour l'évaluation d'un avis d'expert lorsque les avis ou les jugements sont consignés sous une forme qui peut être comparée de manière objective avec une valeur de référence indépendante ou avec une statistique consensuelle. Par exemple, lorsque des entités soumises à l'essai d'aptitude sont classées en catégories connues par contrôle – ou lorsque l'on cherche à déterminer par contrôle si des entités soumises à l'essai d'aptitude proviennent ou non de la même source d'origine – et lorsque les résultats de la classification sont comparés de manière objective, les dispositions du présent document qui se rapportent à des propriétés (qualitatives) nominales peuvent être appliquées.

## 2 Références normatives

Les documents suivants sont cités dans le texte de sorte qu'ils constituent, pour tout ou partie de leur contenu, des exigences 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: Termes statistiques généraux et termes utilisés en calcul des probabilités*

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

ISO 5725-1, *Exactitude (justesse et fidélité) des résultats et méthodes de mesure — Partie 1: Principes généraux et définitions*

ISO/IEC 17043, *Évaluation de la conformité — Exigences générales concernant les essais d'aptitude*

Guide ISO 30, *Matériaux de référence — Termes et définitions choisis*

Guide ISO/IEC 99, *Vocabulaire international de métrologie — Concepts fondamentaux et généraux et termes associés (VIM)*

## 3 Termes et définitions

Pour les besoins du présent document, les termes et définitions de l'ISO 3534-1, l'ISO 3534-2, l'ISO 5725-1, l'ISO/IEC 17043, le Guide ISO/IEC 99, le Guide ISO 30 ainsi que les suivants, s'appliquent. En cas de différences entre ces références concernant l'utilisation des termes, les définitions de l'ISO 3534-1 et de l'ISO 3534-2 s'appliquent. Les symboles mathématiques sont énumérés à l'[Annexe A](#).

L'ISO et l'IEC tiennent à jour des bases de données terminologiques destinées à être utilisées en normalisation, consultables aux adresses suivantes:

- ISO Online browsing platform: disponible à l'adresse <https://www.iso.org/obp>;
- IEC Electropedia: disponible à l'adresse <https://www.electropedia.org/>.

**3.1**  
**comparaison interlaboratoires**  
organisation, exécution et évaluation de mesurages ou d'essais sur la même entité ou sur des entités similaires par deux laboratoires ou plus selon des conditions prédéterminées

**3.2**  
**essai d'aptitude**  
évaluation de la performance d'un participant par rapport à des critères préétablis au moyen d'une *comparaison interlaboratoires* (3.1)

Note 1 à l'article: Pour les besoins du présent document, le terme «essai d'aptitude» est considéré dans son sens le plus large et il inclut, sans s'y limiter:

- les programmes quantitatifs, dans lesquels l'objectif est de quantifier un ou plusieurs mesurandes pour chaque entité soumise à l'essai d'aptitude;
- les programmes qualitatifs, dans lesquels l'objectif est d'identifier ou de décrire une ou plusieurs caractéristiques qualitatives de l'entité soumise à l'essai d'aptitude;
- les programmes séquentiels, dans lesquels une ou plusieurs entités soumises à l'essai d'aptitude sont distribuées séquentiellement pour procéder à l'essai ou au mesurage, et reviennent par intervalles à l'organisateur d'essais d'aptitude;
- les programmes simultanés, dans lesquels les entités soumises à l'essai d'aptitude sont réparties en vue d'essais ou de mesurages réalisés en même temps au cours d'une période de temps définie;
- les exercices de situation unique, dans lesquels les entités soumises à l'essai d'aptitude sont fournies à une seule occasion;
- les programmes continus, dans lesquels les entités soumises à l'essai d'aptitude sont fournies à intervalles réguliers;
- les échantillonnages, dans lesquels des échantillons sont prélevés en vue d'une analyse ultérieure et l'objectif du programme d'essais d'aptitude inclut l'évaluation de l'exécution de l'échantillonnage; et
- les interprétations de données, dans lesquelles des ensembles de données ou d'autres informations sont fournis et les informations sont traitées pour en effectuer une interprétation (ou obtenir un autre résultat).

**3.3**  
**valeur assignée**  
valeur attribuée à une propriété particulière d'une entité soumise à l'essai d'aptitude

**3.4**  
**écart-type pour l'évaluation de l'aptitude**  
mesure de la dispersion utilisée dans l'évaluation des résultats d'un *essai d'aptitude* (3.2)

Note 1 à l'article: Cela peut être interprété comme l'écart-type des résultats pour une population par rapport à une population hypothétique de laboratoires œuvrant en totale conformité aux exigences.

Note 2 à l'article: L'écart-type pour l'évaluation de l'aptitude ne s'applique qu'aux résultats sur une échelle linéaire ou sur une échelle d'intervalle.

Note 3 à l'article: Les programmes d'essais d'aptitude n'évaluent pas tous la performance en se fondant sur la dispersion des résultats.

[SOURCE: ISO/IEC 17043:2010, modifiée — Dans la définition «fondée sur les informations disponibles» a été supprimé. La Note 1 à l'article a été rajoutée, et les Notes 2 et 3 légèrement modifiées.]

**3.5****erreur de mesure**

différence entre la valeur mesurée d'une grandeur et une valeur de référence

[SOURCE: Guide ISO/IEC 99:2007, modifié — Les notes ont été supprimées.]

**3.6****erreur maximale tolérée**

valeur extrême de l'*erreur de mesure* (3.5), par rapport à une valeur de référence connue, qui est tolérée par les spécifications ou règlements pour un mesurage, un instrument de mesure ou un système de mesure donné

[SOURCE: Guide ISO/IEC 99:2007, modifié — Les notes ont été supprimées.]

**3.7****score z**

mesure normalisée de la performance, calculée à partir du résultat du participant, de la *valeur assignée* (3.3) et de l'*écart-type pour l'évaluation de l'aptitude* (3.4)

Note 1 à l'article: Une variante commune du score z, souvent appelée z' (communément prononcé z-prime), est formée en combinant l'incertitude de la valeur assignée avec l'écart-type pour l'évaluation de l'aptitude avant de calculer le score z.

**3.8****score zêta**

mesure normalisée de la performance, calculée à partir du résultat du participant, de la *valeur assignée* (3.3) et des incertitudes-types composées associées au résultat et à la *valeur assignée* (3.3)

**3.9****proportion du score limite admissible**

mesure normalisée de la performance, calculée à partir du résultat du participant, de la *valeur assignée* (3.3) et du critère d'*erreur de mesure* (3.5) lors d'un essai d'aptitude

Note 1 à l'article: Pour des résultats uniques, la performance peut être exprimée par l'écart par rapport à la valeur assignée ( $D$  ou  $D\%$ ).

**3.10****signal d'action**

indication de la nécessité d'une action à la suite du résultat d'un essai d'aptitude

EXEMPLE Par convention, un score z supérieur à 2 est considéré comme une indication de la nécessité de rechercher les causes possibles; un score z de 3 ou supérieur à 3 est, par convention, considéré comme un signal d'action indiquant la nécessité d'une action corrective.

**3.11****valeur consensuelle**

valeur obtenue à partir de l'ensemble des résultats lors d'une *comparaison interlaboratoires* (3.1)

Note 1 à l'article: L'expression «valeur consensuelle» est généralement utilisée pour décrire des estimations de position et de dispersion obtenues à partir des résultats des participants à une campagne du programme d'essais d'aptitude, mais peut aussi être utilisée pour désigner des valeurs obtenues à partir des résultats d'un sous-ensemble spécifié de ces résultats ou, par exemple, d'un certain nombre de laboratoires experts.

**3.12****valeur aberrante**

élément d'un ensemble de valeurs qui est incohérent avec les autres éléments de cet ensemble

Note 1 à l'article: Une valeur aberrante peut apparaître par hasard dans la population attendue, provenir d'une population différente ou résulter d'un enregistrement incorrect ou d'une autre faute.

Note 2 à l'article: De nombreux programmes d'essais d'aptitude utilisent le terme «valeur aberrante» pour désigner un résultat qui génère un signal d'action. Il ne s'agit pas là de l'utilisation prévue du terme. Bien que les valeurs aberrantes génèrent habituellement des signaux d'action, il est possible d'avoir des signaux d'action issus de résultats qui ne sont pas des valeurs aberrantes.

[SOURCE: ISO 5725-1:1994, modifiée — Les notes à l'article ont été ajoutées.]

### 3.13

#### **participant**

laboratoire, organisme ou particulier, recevant les entités soumises à l'essai d'aptitude et soumettant ses résultats à l'organisateur d'essais d'aptitude (3.2)

### 3.14

#### **entité soumise à l'essai d'aptitude**

échantillon, produit, artéfact, matériau de référence, élément d'un matériel, étalon, ensemble de données ou autres informations utilisées pour évaluer la performance d'un participant (3.13) dans des essais d'aptitude (3.2)

Note 1 à l'article: Dans la plupart des cas, les entités soumises à l'essai d'aptitude sont conformes à la définition de «matériau de référence» (3.17) donnée dans le Guide ISO 30.

### 3.15

#### **organisateur d'essais d'aptitude**

organisme ayant la responsabilité de toutes les tâches d'élaboration et d'exécution d'un programme d'essais d'aptitude (3.2)

### 3.16

#### **programme d'essais d'aptitude**

essai d'aptitude (3.2) conçu et exécuté en une ou plusieurs campagnes dans un domaine spécifié d'essai, de mesurage, d'étalonnage ou d'inspection

Note 1 à l'article: Un programme d'essais d'aptitude peut recouvrir un type particulier d'essai, d'étalonnage, d'inspection ou un certain nombre d'essais, d'étalonnages ou d'inspections sur des entités soumises à l'essai d'aptitude.

### 3.17

#### **matériau de référence**

##### **MR**

matériau, suffisamment homogène et stable quant à une ou plusieurs propriétés spécifiées, qui a été préparé pour être adapté à son utilisation prévue dans un processus de mesure

Note 1 à l'article: MR est un terme générique.

Note 2 à l'article: Les propriétés peuvent être quantitatives ou qualitatives, par exemple l'identité de substances ou d'espèces.

Note 3 à l'article: Les utilisations prévues peuvent être l'étalonnage d'un système de mesure, l'évaluation d'une méthode de mesure, l'assignation de valeurs à d'autres matériaux et le contrôle de la qualité.

[SOURCE: Guide ISO 30:2015, modifié — La Note 4 a été supprimée.]

### 3.18

#### **matériau de référence certifié**

##### **MRC**

matériau de référence (MR) (3.17) caractérisé par une procédure métrologiquement valide applicable à une ou plusieurs propriétés spécifiées et accompagné d'un certificat de MR qui indique la valeur de la propriété spécifiée, son incertitude associée, et une expression de la traçabilité métrologique

Note 1 à l'article: Le concept de valeur inclut une propriété nominale ou un attribut qualitatif tels que l'identité ou la séquence. Les incertitudes concernant ces propriétés peuvent être exprimées par des probabilités ou des niveaux de confiance.

[SOURCE: Guide ISO 30:2015, modifié — Les Notes 2, 3 et 4 ont été supprimées.]

## 4 Principes généraux

### 4.1 Exigences générales relatives aux méthodes statistiques

**4.1.1** Les méthodes statistiques utilisées doivent être adaptées à l'objectif et statistiquement correctes. Toutes les hypothèses statistiques sur lesquelles les méthodes ou le plan sont fondés doivent être indiquées dans le plan ou dans une description écrite du programme d'essais d'aptitude et il doit être démontré qu'elles sont raisonnables.

NOTE Une méthode statistiquement valide a une solide base théorique, a des performances connues dans les conditions d'utilisation prévues et repose sur des hypothèses ou des conditions dont on peut démontrer qu'elles s'appliquent suffisamment bien aux données pour l'objectif considéré.

**4.1.2** Le plan statistique et les méthodes d'analyse statistiques des données doivent être en cohérence avec les objectifs annoncés pour le programme d'essais d'aptitude.

**4.1.3** L'organisateur d'essais d'aptitude doit fournir aux participants une description des méthodes de calcul utilisées, une explication de l'interprétation générale des résultats et un énoncé des limites associées à l'interprétation. Cette description doit figurer dans chaque rapport relatif à chaque campagne du programme d'essais d'aptitude ou dans un récapitulatif séparé des procédures mis à la disposition des participants.

**4.1.4** L'organisateur d'essais d'aptitude doit s'assurer que tous les logiciels ont fait l'objet d'une validation adéquate.

### 4.2 Modèle de base

**4.2.1** Pour les résultats quantitatifs de programmes d'essais d'aptitude dans lesquels un seul résultat est consigné pour une entité donnée soumise à l'essai d'aptitude, le modèle de base est donné dans la [Formule \(1\)](#):

$$x_i = \mu + \varepsilon_i \quad (1)$$

où

$x_i$  désigne le résultat d'essai d'aptitude du participant  $i$ ;

$\mu$  désigne la valeur vraie du mesurande;

$\varepsilon_i$  désigne l'erreur de mesure du participant  $i$ , distribuée selon un modèle pertinent.

NOTE 1 Les modèles courants pour  $\varepsilon$  comprennent: la distribution normale  $\varepsilon_i \sim N(0, \sigma^2)$  avec une moyenne de 0 et une variance constante ou différente pour chaque laboratoire; ou plus couramment, une distribution «normale contaminée par des valeurs aberrantes» consistant en un mélange d'une loi normale avec une distribution plus étendue représentant la population de résultats erronés.

NOTE 2 La base de l'évaluation de la performance avec les scores  $z$  et  $\sigma_{pt}$  est que dans une population «idéalisée» de laboratoires compétents, l'écart-type des résultats des essais d'aptitude serait inférieur ou égal à  $\sigma_{pt}$ .

NOTE 3 Ce modèle diffère du modèle de base de l'ISO 5725, en ce qu'il ne contient pas le terme de biais du laboratoire  $B_j$ . Cela est dû au fait qu'il n'est pas possible de faire la distinction entre les termes de biais du laboratoire et d'erreur résiduelle lorsqu'une seule observation est consignée. Toutefois, lorsque les résultats d'un participant issus de plusieurs campagnes ou entités soumises à un programme d'essais d'aptitude sont pris en compte, il peut être utile d'inclure un terme distinct pour le biais du laboratoire.

4.2.2 Pour des résultats ordinaux ou qualitatifs, d'autres modèles peuvent être appropriés ou il peut n'exister aucun modèle statistique.

### 4.3 Approches générales pour l'évaluation des performances

4.3.1 Il existe trois approches générales différentes pour l'évaluation des performances dans un programme d'essais d'aptitude. Ces approches sont utilisées pour atteindre des buts différents dans le cadre du programme d'essais d'aptitude. Les approches sont énumérées ci-dessous:

- a) performances évaluées par comparaison à des critères déduits de manière externe;
- b) performances évaluées par comparaison à d'autres participants;
- c) performances évaluées par comparaison à l'incertitude de mesure revendiquée.

4.3.2 Les approches générales peuvent être appliquées différemment pour déterminer la valeur assignée et pour déterminer les critères d'évaluation des performances, par exemple, lorsque la valeur assignée est la moyenne robuste des résultats des participants et que l'évaluation des performances est déduite de  $\sigma_{pt}$  ou de  $\delta_E$ , où  $\delta_E$  est une tolérance prédéfinie sur l'erreur de mesure et  $\sigma_{pt} = \delta_E/3$ ; de la même manière; dans certaines situations, la valeur assignée peut être une valeur de référence, mais  $\sigma_{pt}$  peut être un écart-type robuste de résultats de participants. Dans l'approche c) utilisant l'incertitude de mesure, la valeur assignée est généralement une valeur de référence appropriée.

## 5 Lignes directrices pour le plan statistique des programmes d'essais d'aptitude

### 5.1 Présentation du plan statistique de programmes d'essais d'aptitude

Les essais d'aptitude portent sur l'évaluation des performances des participants et, en tant que tels, ils ne traitent pas spécifiquement de biais ni de fidélité (bien que ceux-ci puissent être évalués avec des plans spécifiques). Les performances des participants sont évaluées par le biais de l'évaluation statistique de leurs résultats à la suite des mesurages ou des interprétations qu'ils effectuent sur les entités soumises à l'essai d'aptitude. Les performances sont souvent exprimées sous forme de scores de performance permettant une interprétation cohérente sur une gamme de mesurandes et pouvant permettre la comparaison, sur la même base, des résultats relatifs à différents mesurandes. Les scores de performance sont généralement obtenus en comparant la différence entre le résultat consigné par le participant et une valeur assignée avec une dispersion admissible ou avec une estimation de l'incertitude de mesure de la différence. L'examen des scores de performance sur de nombreuses campagnes d'un programme d'essais d'aptitude peut fournir des informations permettant de savoir si des laboratoires démontrent la présence ou non d'effets systématiques constants «biais» ou d'une fidélité médiocre à long terme.

Les [Articles 5 à 10](#) suivants donnent des recommandations relatives à la conception des programmes d'essais d'aptitude quantitatifs et aux traitements statistiques des résultats, y compris le calcul et l'interprétation de divers scores de performance. Les éléments d'appréciation relatifs aux programmes d'essais d'aptitude qualitatifs (y compris les programmes ordinaux) sont donnés à [l'Article 11](#).

### 5.2 Base d'un plan statistique

5.2.1 Conformément à l'ISO/IEC 17043:2010, 4.4.4.1, le plan statistique «doit être élaboré pour répondre aux objectifs du programme d'essais d'aptitude, sur la base de la nature des données (quantitatives ou qualitatives, y compris ordinales et nominales), des hypothèses statistiques, de la nature des erreurs et du nombre attendu de résultats». Par conséquent, des programmes d'essais

d'aptitude avec des objectifs différents et des sources d'erreurs différentes peuvent avoir des plans différents.

Des considérations relatives au plan pour des objectifs courants sont énumérées ci-après. D'autres objectifs sont possibles.

**EXEMPLE 1** Pour un programme d'essais d'aptitude comparant le résultat d'un participant avec une valeur de référence prédéterminée, dans des limites spécifiées avant le début de la campagne, le plan doit inclure une méthode permettant d'obtenir une valeur de référence définie de manière externe, une méthode fixant les valeurs limites et une méthode d'attribution de scores.

**EXEMPLE 2** Pour un programme d'essais d'aptitude comparant les résultats d'un participant avec des résultats combinés provenant d'un groupe issu de la même campagne, dans des limites spécifiées avant le début de la campagne, le plan doit prendre en considération la façon dont la valeur assignée sera déterminée à partir des résultats combinés ainsi que les méthodes de détermination des valeurs limites et d'attribution de scores.

**EXEMPLE 3** Pour un programme d'essais d'aptitude comparant les résultats d'un participant avec des résultats combinés provenant d'un groupe au cours de la même campagne, dans des limites déterminées par la variabilité des résultats des participants, le plan doit prendre en compte le calcul d'une valeur assignée et une mesure appropriée de la dispersion ainsi que la méthode d'attribution de scores.

**EXEMPLE 4** Pour un programme d'essais d'aptitude comparant les résultats d'un participant avec la valeur assignée, en utilisant les incertitudes de mesure du participant, le plan doit prendre en compte la façon dont la valeur assignée et son incertitude doivent être obtenues et la façon dont les incertitudes de mesure des participants doivent être utilisées dans l'attribution de scores.

**EXEMPLE 5** Pour un programme d'essais d'aptitude ayant pour objectif de comparer les performances de différentes méthodes de mesure, le plan doit prendre en compte les statistiques résumées pertinentes et les méthodes permettant de les calculer.

**5.2.2** Divers types de données sont utilisés dans les essais d'aptitude, y compris des données quantitatives, nominales (catégoriques) et ordinales. Parmi les variables quantitatives, certains résultats peuvent se situer sur une échelle d'intervalle, ou sur une échelle relative ou linéaire. Pour certaines mesures sur une échelle quantitative, seul un ensemble discret et discontinu de valeurs peut être réalisé (par exemple, dilutions séquentielles); toutefois, dans de nombreux cas, ces résultats peuvent être traités par des techniques applicables à des variables quantitatives continues.

**NOTE 1** Pour des valeurs quantitatives, une échelle d'intervalle est une échelle sur laquelle les intervalles (différences) sont significatifs, mais pas les ratios, telle que l'échelle de température Celsius. Une échelle linéaire est une échelle sur laquelle les intervalles et les ratios sont tous deux significatifs, telle que l'échelle de température Kelvin, ou la plupart des unités courantes de longueur.

**NOTE 2** Pour des valeurs qualitatives, une échelle nominale a des valeurs distinctes pour lesquelles l'ordre n'est pas significatif, telle que les noms d'espèces bactériennes. Les valeurs sur une échelle ordinale ont un ordre significatif, mais les différences ne sont pas significatives; par exemple, une échelle telle que «grand, moyen, petit» peut être ordonnée, mais les différences entre les valeurs ne sont pas définies autrement qu'en termes du nombre de valeurs intermédiaires.

**5.2.3** Les programmes d'essais d'aptitude peuvent être utilisés à d'autres fins que celles mentionnées ci-dessus, comme indiqué en 0.1 et dans l'ISO/IEC 17043. Le plan doit être approprié pour tous les objectifs définis pour le programme d'essais d'aptitude particulier.

## 5.3 Considérations sur la distribution statistique des résultats

**5.3.1** L'ISO/IEC 17043:2010, 4.4.4.2, exige une cohérence entre les techniques d'analyse statistiques et les hypothèses statistiques concernant les données. Pour les essais d'aptitude, les techniques d'analyse les plus courantes présument qu'un ensemble de résultats fourni par des participants compétents sera approximativement distribué selon une loi normale, ou au moins unimodale et raisonnablement symétrique (après transformation si nécessaire). Il existe une hypothèse supplémentaire habituelle selon laquelle la distribution des résultats de mesurages déterminés par des participants compétents est mélangée avec (ou «contaminée par») des résultats issus d'une population de valeurs erronées pouvant