

Essais d'aptitude des laboratoires : Propositions de limites pour équilibrer les risques de fausse alerte et de défaut de vraies alertes

Louis-Jean Hollebecq
Directeur scientifique et technique

Table of content

1	Introduction	2
2	Symboles et abréviations	3
3	Bases pour la détermination des limites de signaux.....	4
3.1	Introduction.....	4
3.2	Détermination d'un niveau de risque approprié pour les risques nominaux α et β	5
3.3	Détermination d'un niveau de confiance approprié pour la détermination des limites d'alerte correspondant à ce risque nominal	5
3.4	Forme de la distribution	6
3.5	Principes de base utilisés pour calculer les alertes	6
3.6	Principes de base sur l'impact des valeurs aberrantes et sur la méthode utilisée pour le réduire.....	7
3.7	Principes de base utilisés pour déterminer les paramètres de référence pour les EA.....	7
3.8	Principes de base de l'utilisation de la méthode Monte-Carlo	7
4	Résultats d'évaluation du biais	8
4.1	Introduction.....	8
4.2	Résultats des déterminations.....	9
4.3	Conclusions.....	11
5	Résultats pour l'évaluation de la répétabilité.....	12
5.1	Introduction.....	12
5.2	Résultats des déterminations.....	12
5.3	Conclusions.....	16
6	Résultats des évaluations utilisant des méthodes non paramétriques.....	16
6.1	Introduction.....	16
6.2	Résultats pour le premier principe de base.....	18
6.3	Résultats pour le deuxième principe de base	19
6.4	Adaptation des niveaux de risque et d'IC aux cas où des méthodes non paramétriques doivent être utilisées.....	19
6.5	Utilisation de cette méthode pour les résultats d'essais non numériques	20
6.6	Conclusions pour les limites déterminées à l'aide de méthodes non paramétriques	20
7	Conclusions	21
8	Références	21

Résumé :

Les EA reposent sur des scores qui ne doivent pas dépasser certaines limites (généralement 2 et 3 pour le biais) afin de vérifier si les participants appartiennent à la population principale. Ces limites sont toujours conventionnelles et les risques α théoriques associés ne sont pas toujours les mêmes (les limites 2 et 3 correspondent à 2,275 % et 0,135 %, tandis que la norme ISO 5725-2 considère des risques de 1 % et 5 %). La pratique habituelle consistant à utiliser 2 niveaux d'alerte permet de distinguer les performances douteuses et les mauvaises. Cependant, les probabilités de ne pas déclarer les résultats des participants comme des valeurs aberrantes (risque β) ne sont pas prises en compte par ces limites. Dans cette étude, nous avons défini une zone « de doute » comme celle où les risques α et β sont faibles et équilibrés plutôt que d'ignorer le risque β . Cela évite les situations habituelles où β est très élevé, c'est-à-dire un EA avec une puissance très faible. Nous avons déterminé les limites correspondantes pour évaluer le biais et la répétabilité avec $\alpha=\beta=1\%$ à un niveau de confiance de 90 %. Pour le biais, ces « bandes de doute » sont proches des bandes habituelles lorsque $n=110$, plus larges pour des valeurs inférieures de n et vice-versa. Nous avons également déterminé des limites à l'aide de méthodes non paramétriques, qui sont alors exprimées sous forme de rangs plutôt que de scores. Sans surprise, cette méthode est moins efficace et moins puissante et ne devrait être utilisée que lorsque les méthodes paramétriques ne peuvent pas être utilisées.

1 Introduction

La réalisation d'essais d'aptitude des laboratoires nécessite de fixer des limites pour le déclenchement d'alertes pour les scores calculés afin d'évaluer les performances des participants. Ces limites sont généralement liées à un risque théorique α de déclenchement d'une alerte pour un participant qui fait partie de la population répondant aux exigences de la méthode d'essai. De facto, les résultats d'un participant qui remplit effectivement les exigences de la méthode d'essai mais qui, par hasard, se trouve dans l'une des extrémités de la distribution seront considérés comme « aberrants » même s'ils ne le sont pas d'un point de vue théorique.

On distingue deux traditions différentes concernant cette question :

- ✚ La norme ISO 5725-2 [1] fournit des tableaux comprenant des valeurs de risque α de 1 % et 5 % ;
- ✚ Les normes ISO 13528 [2] et ISO 17043 [3] recommandent des limites de 2 et 3 pour les scores z concernant le biais, ce qui renvoie implicitement à une distribution gaussienne avec des valeurs bilatérales pour le risque α égales à 0,135 % et 2,275 % respectivement. On peut également remarquer que, dans d'autres parties de la norme ISO 13528 [2] (à savoir son paragraphe 10.6 concernant un score combiné de biais et de répétabilité), des exemples sont fournis avec des valeurs α de 1 % et 5 %.

On peut en conclure que le choix des valeurs α est toujours conventionnel. On choisit généralement 0,135 % ou 1 % pour les limites d'action et 2,275 % ou 5 % pour les limites d'alerte, mais d'autres limites peuvent être pertinentes. Par exemple, il peut être judicieux d'adopter une limite de 20 % lorsque le risque de ne pas détecter un dysfonctionnement est très critique. Dans tous les cas, le fournisseur d'EA doit indiquer clairement les valeurs α qu'il utilise et les justifier lorsqu'elles ne sont pas celles habituellement adoptées.

Dans tous les cas, aucune considération n'est accordée au risque β correspondant de ne pas déclencher d'alerte pour un participant qui ne fait pas partie de la population répondant aux exigences de la méthode d'essai, c'est-à-dire une véritable valeur aberrante. Plusieurs causes peuvent conduire à de telles situations :

- ✚ Lorsque l'écart du participant par rapport aux exigences de la méthode d'essai n'a pas d'impact très significatif sur ses résultats, le biais induit peut être trop faible pour être détecté lors d'un EA ;

- ✚ Lorsque le nombre de participants ou de répétitions est faible, les effets de l'estimation des valeurs de référence pour l'EA peuvent permettre d'accepter de nombreuses valeurs aberrantes réelles.

Le risque α caractérise donc l'efficacité de l'EA, tandis que le risque β caractérise sa puissance. Cette deuxième question a été étudiée de manière approfondie dans [4] et [5]. L'une des principales conclusions de ces études était que le risque α réel est toujours bien inférieur aux valeurs théoriques, tandis que le risque β est généralement très élevé, pouvant atteindre près de 100 % lorsque les conditions de l'EA ne sont pas bonnes (c'est-à-dire lorsque 1- le nombre de répétitions est trop faible pour annuler l'effet de l'écart-type interne des laboratoires participants ou 2- lorsque le nombre de participants est trop faible pour obtenir une évaluation correcte des participants).

La pratique courante actuelle consistant à utiliser deux niveaux d'alerte (généralement des signaux d'alerte et des signaux d'action) permet de distinguer les cas où les performances sont douteuses de ceux où elles sont susceptibles d'être mauvaises. D'autre part, la volonté de maintenir les alertes pour les performances douteuses peut s'expliquer par la volonté d'alerter les participants même dans les cas douteux, c'est-à-dire de garder à l'esprit l'existence du risque β et d'éviter trop de situations où les résultats du participant sont en fait aberrants sans déclencher d'alerte, sachant que le risque β diminue évidemment lorsque le risque α augmente.

L'idée vient alors que nous pourrions définir une zone « douteuse » comme celle où ni le risque α ni le risque β est faible, ce qui nous permet d'équilibrer ces risques, plutôt que d'ignorer le risque β comme c'est généralement le cas, ce qui conduit à un déséquilibre massif entre eux dans la pratique ($\alpha \ll \beta$) et, par conséquent, à un manque de puissance de l'EA.

Pour y parvenir, les étapes suivantes doivent être suivies :

1. Déterminer un niveau de risque égal et approprié pour α et β , que nous pouvons appeler le « risque nominal » ;
2. Déterminer un niveau de confiance approprié pour la détermination des limites d'alerte correspondant à ce risque nominal ;
3. Déterminer ces limites comme les limites pouvant prendre un score calculé lors d'un EA pour un participant dont le score réel correspond au risque nominal, en fonction de ce qui précède et du nombre de participants.

L'objectif de cette étude est de détailler ces étapes pour trois situations différentes :

- ✚ Dans le cas d'une distribution normale, correspondant à l'évaluation habituelle du biais (généralement les scores z) et des incertitudes (généralement les scores ζ , comme détaillé dans la norme ISO 13528 [2]) ;
- ✚ Dans le cas d'une distribution basée sur le χ^2 , correspondant à l'évaluation de la répétabilité, généralement les scores z_r , comme décrit dans [6] ;
- ✚ Dans le cas d'une distribution inconnue, en utilisant des statistiques non paramétriques.

2 Symboles et abréviations

Les symboles utilisés dans ce document sont listés dans le Tableau 1.

Tableau 1. Liste des symboles utilisés dans de document.

Symbole	Désignation et commentaires
$2u$	Incertitude élargie avec $k = 2$
k	Coefficient d'élargissement pour le calcul de l'IC

Symbole	Désignation et commentaires
Limit-	Limite inférieure des « zones de doute », c'est-à-dire de l'IC sur les scores d'un participant situé exactement à la limite nominale
Limit+	Limite supérieure des « zones de doute », c'est-à-dire de l'IC sur les scores d'un participant situé exactement à la limite nominale
n	Nombre de participants à un EA
r	Nombre de répétitions par un même participant lors d'un EA
s_{ri}	Estimation de σ_{ri}
s_{rref}	Estimation de σ_r pris comme référence lors d'un EA
X_{pt}	Valeur de référence utilisée pour évaluer le biais lors d'un EA
Z-score	Paramètre qui caractérise le biais d'un participant, tel que défini dans la norme ISO 13528
z-score	Estimation du score Z lors d'un EA
ZR-score	Paramètre qui caractérise la répétabilité d'un participant, défini comme le rapport s_{ri}/s_{rref}
zr-score	Estimation du score zr d'un participant lors d'un EA
α	Risque de déclencher une alerte pour un participant qui ne le mérite pas
β	Risque de ne pas déclencher une alerte pour un participant qui le mérite
σ_L	Écart type interlaboratoires
σ_{pt}	Valeur de référence pour les écarts types des biais des participants lors d'un EA
σ_r	Écart type de répétabilité
σ_{ri}	Écart type de répétabilité du participant <i>i</i>

Abréviations :

-  CIL : comparaison inter-laboratoires
-  EA : essais d'aptitude
-  ET : écart-type
-  IC : intervalle de confiance bilatéral. Par exemple, IC95% signifie l'intervalle de confiance bilatéral [2,5%;97,5%]
-  MMC : méthode de Monte-Carlo

3 Bases pour la détermination des limites de signaux

3.1 Introduction

En cohérence avec les indications du § 1, le calcul des limites des demandes d'alerte doit tenir compte de la distribution des scores calculés pour un participant correspondant exactement au risque nominal, et prendre en considération les centiles de cette distribution afin de déterminer un IC approprié pour l'estimation de ce risque nominal.

Cela implique de traiter les questions suivantes :

-  La détermination d'un niveau approprié pour les risques nominaux α et β , afin de déterminer un score Z nominal de référence d'un participant hypothétique correspondant exactement au risque nominal ;

- ✚ La détermination d'un niveau de confiance approprié pour la détermination des limites d'alerte correspondant à ce risque nominal ;
- ✚ La prise en compte de la forme de la distribution des résultats des tests ;
- ✚ La méthode utilisée pour déterminer les limites d'alerte ;
- ✚ L'impact des valeurs aberrantes et de la méthode utilisée pour le réduire ;
- ✚ Les bases utilisées pour déterminer les paramètres de référence pour l'EA.

Toutes ces questions sont détaillées ci-après.

Comme le calcul des limites implique certaines questions qui peuvent difficilement être traitées de manière adéquate avec des outils théoriques (typiquement les algorithmes robustes pour déterminer les paramètres de référence), la méthode Monte-Carlo (MMC) a été utilisée pour déterminer nos propositions de limites de signal, voir par exemple [7] pour obtenir des explications à ce sujet.

3.2 Détermination d'un niveau de risque approprié pour les risques nominaux α et β

Comme rappelé au § 1, le choix des valeurs α est toujours conventionnel. On choisit généralement 0,135 % ou 1 % pour les limites d'action et 2,275 % ou 5 % pour les limites d'alerte, mais d'autres limites peuvent être pertinentes.

Une valeur appropriée pour les risques α et β devrait alors être telle que la zone « de doute » se situe lorsque les risques α et β ne sont pas atteints, avec une probabilité spécifiée. Il est alors évident que ce niveau devrait avoir un rapport avec les valeurs traditionnelles choisies pour déterminer les signaux d'alerte et les signaux d'action.

De notre côté, nous avons choisi un risque nominal de 1 % bilatéral (c'est-à-dire 0,5 % unilatéral), ce qui conduit, pour les distributions gaussiennes, à une limite centrale d'alerte de 2,576, qui peut être considérée comme une valeur moyenne entre les valeurs traditionnelles de 2 et 3 pour les signaux d'alerte et les signaux d'action respectivement.

Un tel choix nous permet d'espérer que les limites d'alerte calculées à partir de l'IC seront réparties à peu près à la même distance de cette valeur nominale, c'est-à-dire qu'elles ne seront pas loin des valeurs traditionnelles pour les limites d'alerte (soit 2 et 3) pour un certain nombre de participants, ce qui n'est pas loin des situations habituelles rencontrées lors des EA.

Évidemment, d'autres valeurs pourraient avoir du sens, mais nous verrons plus loin que ce choix était approprié.

3.3 Détermination d'un niveau de confiance approprié pour la détermination des limites d'alerte correspondant à ce risque nominal

Pour chaque nombre de participants, l'incertitude sur les paramètres utilisés pour calculer les scores entraîne une incertitude sur les valeurs de ces scores calculés. Cette incertitude diminue évidemment lorsque le nombre de participants augmente. Il s'agit alors de déterminer un niveau de confiance pour la détermination des limites d'alerte qui :

- ✚ Soit pertinent, c'est-à-dire corresponde à un risque raisonnable au regard des pratiques habituelles ;
- ✚ Ne soit pas trop éloigné des pratiques actuelles, c'est-à-dire déclenche un nombre raisonnable d'alertes, ni trop peu, ni trop.

Nous avons constaté que l'IC90 % (c'est-à-dire un niveau de confiance bilatéral de 90 %) est pertinent au regard de ces deux exigences.

3.4 Forme de la distribution

La plupart des EA visent à évaluer le biais des résultats des participants, pour lesquels une distribution gaussienne est implicitement supposée dans les normes de référence. La norme ISO 13528 [2] stipule que la normalité de la distribution n'est pas une exigence, mais recommande de vérifier que la distribution est symétrique et, si ce n'est pas le cas, recommande d'utiliser un changement de variable pour la rendre symétrique.

Cependant, certains types d'évaluation nécessitent manifestement de prendre en compte d'autres types de distributions. En particulier, l'évaluation de la répétabilité des participants nécessite de prendre en compte des distributions basées sur χ^2 (voir [6]) liées à la distribution des estimations d'un écart-type.

Nous pouvons également envisager des cas où aucune hypothèse correcte ne peut être formulée concernant la forme de la distribution. Nous avons alors essayé de traiter ces situations au moyen de méthodes statistiques non paramétriques, en gardant à l'esprit que, de toute évidence, ces méthodes ne peuvent jamais être aussi puissantes que les statistiques paramétriques.

Cette étude a donc examiné trois cas :

- ✚ Les résultats des tests suivant une distribution gaussienne ;
- ✚ Les écarts types des résultats des tests, suivant des distributions basées sur χ^2 , voir [6] ;
- ✚ Les données suivant une distribution inconnue.

3.5 Principes de base utilisés pour calculer les alertes

Même si cela n'est pas explicitement traité dans les normes de référence (ISO 5725-2 [1], ISO 13528 [2] and ISO 17043 [3]), au moins deux principes de base sont utilisés pour calculer les alertes :

Le premier, utilisé pour évaluer le biais (généralement les scores z), considère que les valeurs aberrantes sont celles situées aux extrémités de la distribution. Les alertes sont alors basées sur une vérification visant à déterminer si les résultats du participant se situent à l'intérieur ou au-delà d'une distance donnée par rapport à la valeur de référence ;

Le second, utilisé pour vérifier la répétabilité (généralement les ratios de Cochran et les scores k de Mandel), considère que tous les participants doivent présenter le même ET de répétabilité (hypothèse d'homoscédasticité). Les alertes sont alors basées sur une vérification visant à déterminer si les résultats du participant sont significativement différents de la valeur de référence ou pas.

Le premier type d'évaluation demande de déterminer deux paramètres :

- ✚ Une valeur de référence centrale (typiquement le X_{pt} de la norme ISO 13528 [2]) ;
- ✚ Une valeur de référence pour la distance acceptable par rapport à la valeur de référence (typiquement le σ_{pt} de la norme ISO 13528 [2]).

Le deuxième type d'évaluation suppose la détermination d'un seul paramètre : la valeur de référence centrale (généralement σ_r pour l'évaluation de la répétabilité). Il convient de noter que si elle était utilisée pour le biais, cette option conduirait à considérer tous les participants qui ne sont pas contenus dans l'IC sur la valeur de référence comme des valeurs aberrantes, c'est-à-dire la plupart des participants. Il s'agirait alors d'une méthode inutilisable dans la pratique, car elle déclencherait beaucoup trop d'alertes.

Cette étude a donc examiné :

- ✚ Le premier type pour les évaluations utilisant des distributions gaussiennes ;
- ✚ Le deuxième type pour les évaluations d'ET ;

✚ Les deux types pour les distributions inconnues.

3.6 Principes de base sur l'impact des valeurs aberrantes et sur la méthode utilisée pour le réduire

Toutes les méthodes de référence pour les CIL (ISO 5725-2 [1], ISO 13528 [2] and ISO 17043 [3]) mettent en garde contre l'impact des valeurs aberrantes sur les résultats des comparaisons car, lors des CIL, l'organisateur a peu de contrôle sur la qualité des données. En particulier, lorsque certains chiffres semblent étranges, il n'est pas possible de vérifier si une raison technique ou pratique peut les expliquer et permettre de décider, sur une base technique, que ces chiffres ne font pas partie de la population principale des résultats (c'est-à-dire la population des résultats d'essais obtenus conformément aux exigences de la méthode d'essai). Plusieurs méthodes statistiques robustes sont décrites dans ces normes pour faire face à la présence de valeurs aberrantes.

De plus, lorsqu'un résultat d'essai presque aberrant est présent dans la série de résultats d'essais (ce qui est, par construction, toujours le cas dans nos séries MMC), cela affecte la détermination des paramètres et, par conséquent, les résultats de l'évaluation. Cet effet est particulièrement fort lorsque l'EA implique un faible nombre de participants.

Pour tenir compte de tous ces éléments, nos calculs ont été effectués à l'aide des méthodes robustes Algo A et Algo S (telles que décrites dans les normes ISO 5725-2 [1] and ISO 13528 [2]) afin de calculer les paramètres de référence pour les distributions, qui sont les plus couramment utilisés par les fournisseurs d'EA.

3.7 Principes de base utilisés pour déterminer les paramètres de référence pour les EA

Même si, pour des raisons pratiques, le consensus des participants est généralement utilisé pour déterminer les paramètres de référence pour les EA, il existe plusieurs autres bonnes solutions (voir la norme ISO 13528 [2]) qui, selon les cas, peuvent être mieux adaptées à la situation que la méthode habituelle du consensus des participants.

Il est évident qu'aucune méthode statistique ne peut être utilisée pour déterminer les limites lorsque la valeur de référence n'est pas déterminée par la méthode du « consensus ». Par exemple, pour l'évaluation du biais selon la norme ISO 13528 [2], lorsque la valeur centrale X_{pt} est fixée par formulation et que l'écart acceptable σ_{pt} provient d'une source externe, certaines hypothèses doivent être formulées concernant la précision avec laquelle ces paramètres peuvent être déterminés. Ces hypothèses sont liées aux principes de base de la méthode d'essai plutôt qu'à des questions statistiques, et ne peuvent jamais être applicables à tous les EA, c'est-à-dire devenir des règles générales pour la réalisation d'EA.

C'est pourquoi nous avons décidé de concentrer notre étude sur la méthode du consensus. De fait, la méthode du consensus est généralement moins efficace que les autres et, par conséquent, l'utilisation des limites que nous avons calculées dans cette étude est susceptible de conduire à un IC avec une confiance supérieure à notre valeur nominale de 90 %. Par conséquent, dans la plupart des cas, elles feront sens quelle que soit la manière dont les valeurs de référence sont calculées, même si elles ont été calculées uniquement avec la méthode du consensus.

3.8 Principes de base de l'utilisation de la méthode Monte-Carlo

L'utilisation de la MMC nécessite un modèle qui corresponde raisonnablement aux situations rencontrées dans le monde réel. Cela a été facile pour cette étude, pour laquelle tous les modèles sont disponibles dans les normes de référence.

L'utilisation des méthodes Monte-Carlo nécessite également l'utilisation de valeurs d'entrée aléatoires. Lorsque plusieurs valeurs aléatoires sont nécessaires pour produire un résultat Monte-Carlo et que des corrélations entre

elles s'appliquent dans la vie réelle, ces corrélations doivent être intégrées dans les valeurs aléatoires d'entrée pour les calculs. Dans notre cas, nous n'avons dû produire qu'une seule valeur aléatoire, soit des biais, soit un écart-type aléatoire, ce qui n'a donc pas posé de problème.

Pour garantir la validité des conclusions, les séries aléatoires doivent être suffisamment nombreuses, en fonction de nombreux facteurs. Dans notre étude, nous avons calculé des séries de 10^7 à 10^8 résultats pour chaque situation. Chacune d'entre elles a été divisée en 50 à 1000 sous-groupes, ce qui nous a permis de vérifier la répétabilité des paramètres calculés et du pourcentage d'alertes au sein de ces sous-groupes et de calculer un IC associé (avec $k = 2$).

4 Résultats d'évaluation du biais

4.1 Introduction

En cohérence avec le § 3, les déterminations ont été effectuées comme suit :

- ✚ Une série de nombre de participants allant de 5 à 250 a été prise en compte ;
- ✚ Les résultats des participants ont été obtenus à l'aide d'une distribution gaussienne ;
- ✚ Dans chaque série de participants, l'un d'entre eux a été fixé à la limite nominale d'alerte (c'est-à-dire $Z = 2,576$, correspondant à des risques nominaux α et β de 1 % bilatéraux) au lieu d'être aléatoire ;
- ✚ L'Algo A a été utilisé pour déterminer la valeur de référence et l'écart type de référence de la distribution, puis utilisé pour calculer le score z estimé du participant dont la valeur réelle est $Z = 2,576$;
- ✚ Les centiles 5 % et 95 % de ces scores z estimés ont ensuite été calculés pour déterminer l'IC90 % correspondant.

La Figure 1 montre un exemple de distribution d'une sous-série de 400 estimations de scores z pour un participant dont le score Z réel est de 2,576, avec un nombre de participants de 25. Les centiles 5 % et 95 % correspondants sont respectivement 1,711 et 3,743 dans ce cas, ce qui n'est pas très éloigné des valeurs 1,673 et 3,927 déterminées avec 100 répétitions de ces courbes.

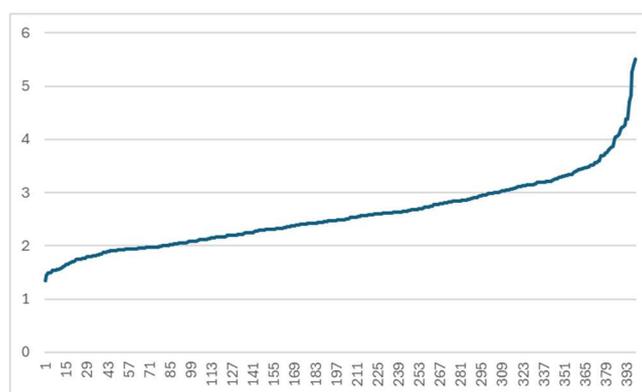


Figure 1: Exemple de distribution d'une sous-série de 400 estimations de scores z pour un participant dont le score Z réel est de 2,576, avec un nombre de participants égal à 25. Les centiles 5 % et 95 % correspondants sont respectivement 1,711 and 3,743 dans ce cas.

Il convient de noter que ces limites s'appliquent à la valeur apparente de σ_{pt} , qui est généralement supérieure à σ_L car une contribution supplémentaire de l'écart-type interne du laboratoire ne peut jamais être évitée, voir [4] et [5]. Cet effet réduit le risque α et augmente le risque β . Les références [4] et [5] ont également démontré que lorsque le rapport λ défini comme $\lambda = \sigma_r / (r \cdot \sigma_L)$ (où σ_r est l'ET des répétitions au sein d'un même participant, r

est le nombre de répétitions au sein de chaque participant et σ_l est l'ET inter-laboratoires) est inférieur à 0,17, cet effet devient négligeable et les risques α et β ne dépendent que du nombre de participants.

4.2 Résultats des déterminations

Le Tableau 2 présente les résultats de la détermination des limites pour les alertes correspondant à des risques nominaux α et β égaux à 1 % (bilatéraux), correspondant à des scores Z réels égaux à 2,576. Bien entendu, ces valeurs sont opposées et inversées avec des signes négatifs pour les scores Z réels égaux à -2,576 (par exemple, [-4,96;-1,24] pour $n = 10$).

Tableau 2. Résultats de la détermination des limites pour les alertes correspondant à des risques nominaux α et β égaux à 0,5%.

n	Résultats				Propositions	
	Limit-	2u	Limit+	2u	Limit-	Limit+
3	0,6743	0,0000	13,468	0,021	0,67	13,5
4	0,7943	0,0002	6,6109	0,0064	0,79	6,6
5	0,8194	0,0004	8,6169	0,0085	0,82	8,6
6	0,9184	0,0004	5,9168	0,0047	0,92	5,9
7	0,9952	0,0004	6,2009	0,0053	0,995	6,2
8	1,1128	0,0004	5,3178	0,0041	1,11	5,3
9	1,1680	0,0005	5,6239	0,0047	1,17	5,6
10	1,2391	0,0005	4,9617	0,0035	1,24	4,96
11	1,2773	0,0006	5,0483	0,0038	1,28	5,05
12	1,3395	0,0006	4,6708	0,0034	1,34	4,67
13	1,3685	0,0006	4,7614	0,0035	1,36	4,76
14	1,4186	0,0006	4,4737	0,0032	1,42	4,475
15	1,4408	0,0006	4,5188	0,0031	1,44	4,52
16	1,4837	0,0007	4,3083	0,0029	1,48	4,31
17	1,5015	0,0007	4,3511	0,0032	1,5	4,35
18	1,5392	0,0007	4,1857	0,0029	1,54	4,185
19	1,5539	0,0007	4,2120	0,0030	1,55	4,21
20	1,5866	0,0007	4,0773	0,0027	1,59	4,08
21	1,5994	0,0007	4,0984	0,0027	1,6	4,1
22	1,6278	0,0007	3,9897	0,0026	1,63	3,99
23	1,6381	0,0007	4,0054	0,0027	1,64	4,005
24	1,6638	0,0007	3,9138	0,0025	1,66	3,915
25	1,6731	0,0008	3,9266	0,0027	1,67	3,925
26	1,6954	0,0008	3,8504	0,0025	1,695	3,85
27	1,7031	0,0007	3,8587	0,0025	1,7	3,86
28	1,7240	0,0008	3,7917	0,0024	1,72	3,79
29	1,7305	0,0008	3,7970	0,0024	1,73	3,8
30	1,7493	0,0008	3,7384	0,0024	1,75	3,74
31	1,7556	0,0008	3,7461	0,0024	1,755	3,745
32	1,7726	0,0008	3,6956	0,0024	1,77	3,695
33	1,7788	0,0008	3,7039	0,0023	1,78	3,705
34	1,7941	0,0008	3,6538	0,0023	1,795	3,655
35	1,7992	0,0008	3,6599	0,0023	1,8	3,66

n	Résultats				Propositions	
	Limit-	2u	Limit+	2u	Limit-	Limit+
36	1,8139	0,0008	3,6169	0,0021	1,81	3,62
37	1,8179	0,0009	3,6239	0,0023	1,82	3,625
38	1,8321	0,0008	3,5832	0,0022	1,83	3,58
39	1,8357	0,0008	3,5870	0,0022	1,84	3,59
40	1,8494	0,0009	3,5529	0,0022	1,85	3,55
45	1,8830	0,0009	3,5006	0,0022	1,88	3,5
50	1,9186	0,0009	3,4331	0,0021	1,92	3,43
55	1,9428	0,0009	3,3965	0,0021	1,94	3,395
60	1,9714	0,0010	3,3456	0,0019	1,97	3,345
65	1,9893	0,0009	3,3181	0,0020	1,99	3,32
70	2,0122	0,0009	3,2810	0,0019	2,01	3,28
75	2,0268	0,0010	3,2599	0,0019	2,03	3,26
80	2,0459	0,0010	3,2296	0,0018	2,045	3,23
85	2,0594	0,0010	3,2113	0,0020	2,06	3,21
90	2,0724	0,0010	3,1894	0,0019	2,07	3,19
95	2,0833	0,0010	3,1726	0,0018	2,08	3,17
100	2,0962	0,0010	3,1527	0,0019	2,095	3,15
105	2,1056	0,0010	3,1401	0,0019	2,105	3,14
110	2,1169	0,0010	3,1232	0,0017	2,12	3,125
115	2,1254	0,0010	3,1129	0,0018	2,125	3,11
120	2,1357	0,0010	3,0971	0,0017	2,135	3,1
125	2,1425	0,0010	3,0895	0,0016	2,14	3,09
130	2,1517	0,0010	3,0746	0,0017	2,15	3,075
135	2,1574	0,0010	3,0663	0,0017	2,16	3,065
140	2,1657	0,0010	3,0556	0,0018	2,165	3,055
145	2,1714	0,0010	3,0468	0,0017	2,17	3,045
150	2,1789	0,0010	3,0369	0,0018	2,18	3,035
160	2,1911	0,0011	3,0205	0,0017	2,19	3,02
170	2,2015	0,0010	3,0070	0,0017	2,2	3,01
180	2,2115	0,0011	2,9948	0,0015	2,21	2,995
190	2,2212	0,0010	2,9801	0,0017	2,22	2,98
200	2,2286	0,0011	2,9712	0,0017	2,23	2,97
210	2,2370	0,0011	2,9588	0,0017	2,24	2,96

n	Résultats				Propositions	
	Limit-	2u	Limit+	2u	Limit-	Limit+
220	2,2446	0,0011	2,9504	0,0017	2,245	2,95
230	2,2519	0,0012	2,9399	0,0016	2,25	2,94

n	Résultats				Propositions	
	Limit-	2u	Limit+	2u	Limit-	Limit+
240	2,2579	0,0010	2,9335	0,0016	2,26	2,93
250	2,2645	0,0011	2,9244	0,0016	2,265	2,925

Ces chiffres montrent que les limites classiques pour les scores z , à savoir 2 et 3, ne sont pas très éloignées de la situation où α et β sont égaux à 1 % bilatéral avec un IC90 % lorsque $n = 110$.

La Figure 2 montre les valeurs « Limit+ » pour les alertes correspondant à des risques nominaux α et β égaux à 1 % bilatéral pour $3 \leq n \leq 25$. On constate sur cette figure que les valeurs impaires de n présentent des valeurs limites plus élevées que les valeurs paires voisines de n . Le même phénomène se produit également pour les valeurs « Limit- », mais avec moins d'importance. Sans surprise, l'importance du phénomène diminue lorsque n augmente.

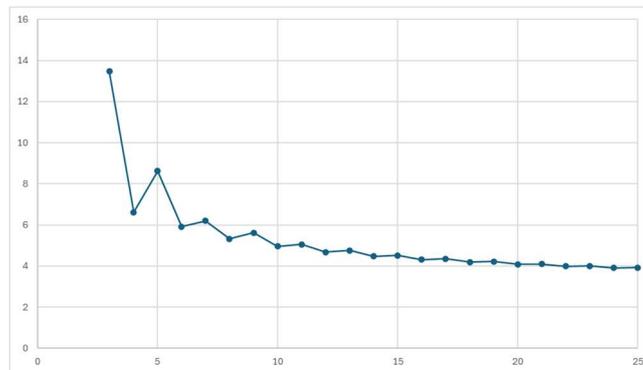
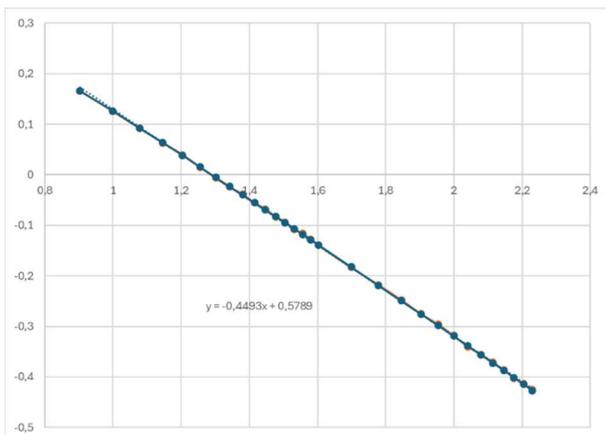
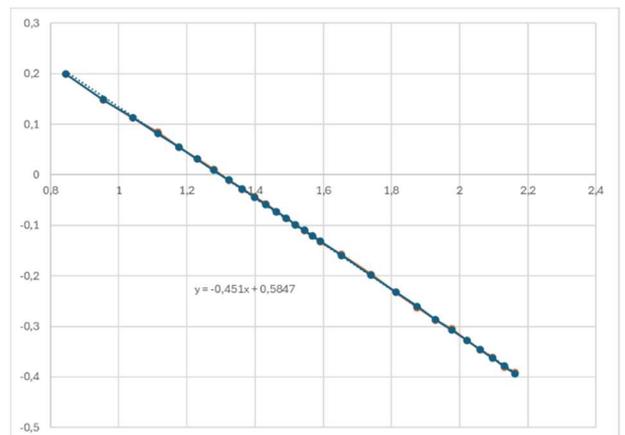


Figure 2: Limit+ pour les alertes correspondant à des risques nominaux α et β égaux à 0,5% pour $3 \leq n \leq 25$.

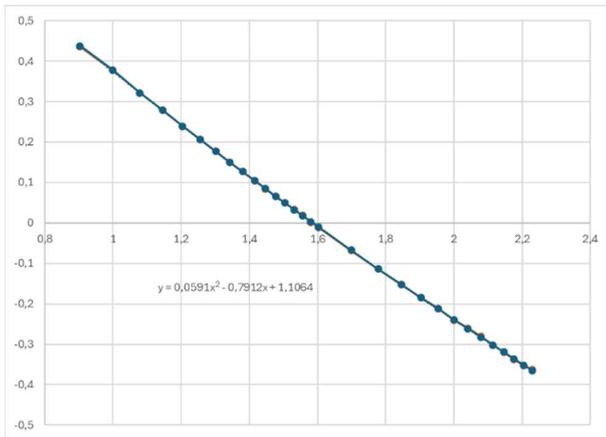
Pour cette raison, nous avons séparé les nombres pairs et les nombres impairs lors de la recherche d'équations empiriques nous permettant de calculer les valeurs intermédiaires des limites. Pour ce faire, nous avons représenté graphiquement les valeurs logarithmiques des écarts entre les alertes et la valeur nominale (c'est-à-dire 2,576) en fonction du logarithme du nombre de participants, voir Figure 3.



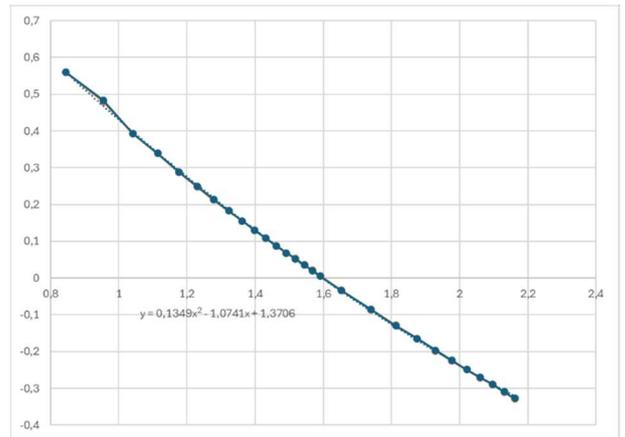
Limites inférieures pour les nombres pairs de participants



Limites inférieures pour les nombres impairs de participants



Limites supérieures pour les nombres pairs de participants



Limites supérieures pour les nombres impairs de participants

Figure 3: Logarithmes des écarts entre les limites d'alerte et la valeur nominale (2,576) en fonction du logarithme du nombre de participants pour des risques nominaux α et β égaux à 0,5% et pour $7 \leq n \leq 250$.

Ces figures montrent que des régressions linéaires permettent de calculer avec une précision suffisante les valeurs des limites inférieures. Pour les limites supérieures, il faut utiliser un polynôme de degré 2 si l'on veut obtenir une précision supérieure à 0,01, même si la courbe n'est pas loin d'être droite.

L'Equation (1) suivante permet de calculer les limites d'alerte pour $10 \leq n \leq 250$ à 0,02 près pour la limite+ pour les valeurs impaires de n et à 0,01 près pour les autres limites.

$$Lim = 2,576 \pm 10^a \quad (1)$$

où

\pm est « - » pour les limites inférieures et « + » pour les limites supérieures,
 $a = -0,45 \cdot \log(n) + 0,58$ pour la limite inférieure lorsque n est pair,
 $a = -0,45 \cdot \log(n) + 0,585$ pour la limite inférieure lorsque n est impair,
 $a = 0,059 \cdot (\log(n))^2 - 0,791 \cdot \log(n) + 1,106$ pour la limite supérieure lorsque n est pair,
 $a = 0,135 \cdot (\log(n))^2 - 1,075 \cdot \log(n) + 1,37$ pour la limite supérieure lorsque n est impair,
 et n est le nombre de résultats utilisés pour calculer a .

4.3 Conclusions

Les limites classiques pour les scores z , c'est-à-dire 2 et 3, ne sont pas éloignées de la situation où α et β sont égaux à 1 % bilatéral avec un IC90 % lorsque $n = 110$.

L'IC90 % autour de la valeur nominale 2,576 est plus élevé pour les valeurs impaires de n que pour les valeurs paires voisines de n .

Des formules empiriques ont pu être déterminées pour le calcul des limites, en fonction du nombre de participants.

Ces limites s'appliquent à la valeur apparente de σ_{pt} , qui est généralement supérieure à σ_L car une contribution supplémentaire de l'écart-type interne du laboratoire ne peut jamais être évitée. Cet effet réduit le risque α et

augmente le risque β . Lorsque le rapport λ défini comme $\lambda = \sigma_r / (n_r \cdot \sigma_L)$ est inférieur à 0,17, cet effet devient négligeable et aucune différence entre les risques α et β ne s'applique alors.

Lorsque le rapport λ est nettement supérieur à 0,17, une solution pour équilibrer les risques α et β consisterait à utiliser $\sigma_{pt} = \sqrt{s_{pt}^2 - s_r^2 / r}$ (où s_{pt} est l'écart type de la valeur moyenne des participants et s_r est l'écart type moyen des répétitions au sein d'un participant (qui peut être calculé avec l'Algo S). Cependant, l'IC sur ces σ_{pt} impliquant une variance imbriquée, est généralement beaucoup plus grand que sur s_{pt} (voir [8]) et, par conséquent, nos propositions du Tableau 2 ne sont pas valables (trop étroites) pour une telle utilisation.

5 Résultats pour l'évaluation de la répétabilité

5.1 Introduction

En cohérence avec le § 3, les déterminations ont été effectuées comme suit :

- ✚ Des séries de 3 à 250 participants et de 2 à 25 répétitions par participant ont été prises en compte ;
- ✚ Les résultats des participants ont été obtenus à l'aide d'une distribution $\sqrt{\chi_{r-1}^2 / (r-1)}$, qui décrit les estimations de l'écart-type de répétabilité, voir [6] ;
- ✚ Comme proposé dans les normes de référence habituelles, le calcul des limites a été effectué en utilisant l'hypothèse d'homoscédasticité, voir § 3.5 et [6] ;
- ✚ Les scores zr , c'est-à-dire les rapports sr_i / sr_{ref} (où sr_i est l'ET de répétabilité du participant i et sr_{ref} est l'ET de répétabilité de référence) tels que proposés dans [6] et qui sont similaires à ceux proposés dans la norme ISO 5725-2 [1], ont été utilisés pour caractériser les résultats des participants ;
- ✚ Dans chaque série de participants, l'un d'entre eux a été fixé à la limite nominale d'alerte en fonction de r (soit $ZR = \sqrt{\chi_{0,995,r-1}^2 / (r-1)}$), correspondant à des risques nominaux α et β de 1 % bilatéraux) au lieu d'être aléatoire ;
- ✚ Seules les limites supérieures d'alerte ont été prises en compte, car les limites inférieures ne sont généralement pas pertinentes pour l'alerte, voir [6]. Par conséquent, les risques nominaux α et β deviennent unilatéraux et sont égaux à 0,5 % ;
- ✚ L'Algo S a été utilisé pour déterminer l'écart type de référence de la distribution, puis utilisé pour calculer les scores zr estimés des participants avec une valeur réelle $ZR = \sqrt{\chi_{0,995,r-1}^2 / (r-1)}$;
- ✚ Les centiles 5 % et 95 % de ces scores zr estimés ont ensuite été calculés afin de déterminer l'IC90 % correspondant (de la même manière que pour l'évaluation du biais, voir Figure 1).

Contrairement au cas de l'évaluation du biais (voir § 4.1), grâce à l'utilisation d'une hypothèse d'homoscédasticité, aucun écart-type parasite supplémentaire n'est à craindre dans la détermination des limites, de sorte que les limites calculées ici représentent véritablement celles pour lesquelles les risques nominaux α et β sont égaux à 0,5 % (seule l'extrémité supérieure de la distribution est prise en compte).

5.2 Résultats des déterminations

Le Tableau 3 présente les valeurs ZR correspondant aux risques nominaux α et β égaux à 0,5 %, en fonction de r , nombre de répétitions par participant.

Tableau 3. Valeurs ZR correspondant aux risques nominaux α and β égaux à 0,5%.

r	2	3	4	5	6	8	10	12	16	20	25
Nominal Limits	2,807	2,302	2,069	1,927	1,830	1,702	1,619	1,560	1,479	1,425	1,378

Le Tableau 4 présente les résultats de la détermination des limites pour les alertes correspondant à des risques nominaux α et β égaux à 0,5 % (unilatéral, côté supérieur), correspondant à des scores ZR réels égaux aux limites nominales du Tableau 3.

Tableau 4. Résultats de la détermination des limites pour les alertes correspondant à des risques nominaux α et β égaux à 0,5 %.

n	r	Résultats				Propositions	
		Limit-	2u	Limit+	2u	Limit-	Limit+
3	2	1,1901	0,0002	7,7845	0,0083	1,19	7,8
3	3	1,1936	0,0002	3,8816	0,0022	1,194	3,88
3	4	1,1845	0,0001	2,9915	0,0014	1,184	2,99
3	5	1,1741	0,0001	2,5828	0,0009	1,174	2,58
3	6	1,1658	0,0001	2,3441	0,0007	1,166	2,34
3	8	1,1522	0,0001	2,0681	0,0005	1,152	2,07
3	10	1,1409	0,0001	1,9075	0,0004	1,141	1,91
3	12	1,1309	0,0001	1,7908	0,0004	1,131	1,79
3	16	1,1146	0,0001	1,6288	0,0003	1,115	1,63
3	20	1,1032	0,0001	1,5289	0,0002	1,103	1,53
3	25	1,0924	0,0001	1,4459	0,0002	1,092	1,45
4	2	1,2985	0,0002	5,9487	0,0054	1,299	5,95
4	3	1,2859	0,0002	3,4591	0,0017	1,286	3,46
4	4	1,2669	0,0002	2,7808	0,0011	1,267	2,78
4	5	1,2493	0,0002	2,4475	0,0008	1,249	2,45
4	6	1,2355	0,0001	2,2448	0,0006	1,236	2,245
4	8	1,2139	0,0001	2,0036	0,0004	1,214	2,00
4	10	1,1971	0,0001	1,8603	0,0004	1,197	1,86
4	12	1,1823	0,0001	1,7551	0,0003	1,182	1,755
4	16	1,1593	0,0001	1,6087	0,0002	1,159	1,61
4	20	1,1432	0,0001	1,5170	0,0002	1,143	1,52
4	25	1,1286	0,0001	1,4403	0,0002	1,128	1,44
5	2	1,4443	0,0005	6,9233	0,0081	1,444	6,9
5	3	1,3892	0,0005	3,6544	0,0032	1,39	3,655
5	4	1,3530	0,0004	2,8712	0,0020	1,353	2,87
5	5	1,3253	0,0003	2,5038	0,0013	1,326	2,5
5	6	1,3036	0,0003	2,2867	0,0011	1,304	2,29
5	8	1,2727	0,0003	2,0305	0,0007	1,27	2,03
5	10	1,2493	0,0002	1,8795	0,0006	1,249	1,88
5	12	1,2284	0,0002	1,7697	0,0005	1,228	1,77
5	16	1,1953	0,0002	1,6184	0,0003	1,195	1,62
5	20	1,1736	0,0001	1,5252	0,0003	1,174	1,53
5	25	1,1550	0,0001	1,4473	0,0003	1,155	1,45
6	2	1,5768	0,0007	5,9903	0,0057	1,58	6,0
6	3	1,4744	0,0007	3,4356	0,0027	1,47	3,44
6	4	1,4197	0,0005	2,7591	0,0015	1,42	2,76
6	5	1,3816	0,0005	2,4298	0,0012	1,38	2,43
6	6	1,3546	0,0004	2,2302	0,0009	1,355	2,23
6	8	1,3156	0,0003	1,9934	0,0007	1,32	1,99
6	10	1,2873	0,0003	1,8513	0,0005	1,29	1,85

n	r	Résultats				Propositions	
		Limit-	2u	Limit+	2u	Limit-	Limit+
6	12	1,2609	0,0002	1,7476	0,0005	1,26	1,75
6	16	1,2205	0,0002	1,6050	0,0003	1,22	1,605
6	20	1,1955	0,0002	1,5160	0,0003	1,20	1,52
6	25	1,1739	0,0001	1,4422	0,0002	1,17	1,44
8	2	1,7734	0,0008	5,7634	0,0057	1,77	5,76
8	3	1,6058	0,0007	3,3491	0,0025	1,61	3,35
8	4	1,5249	0,0005	2,7045	0,0015	1,525	2,70
8	5	1,4715	0,0005	2,3895	0,0011	1,47	2,39
8	6	1,4340	0,0004	2,1989	0,0009	1,43	2,20
8	8	1,3812	0,0003	1,9700	0,0007	1,38	1,97
8	10	1,3439	0,0003	1,8329	0,0005	1,34	1,83
8	12	1,3119	0,0003	1,7329	0,0005	1,31	1,73
8	16	1,2610	0,0002	1,5949	0,0003	1,26	1,595
8	20	1,2289	0,0002	1,5094	0,0003	1,23	1,51
8	25	1,2022	0,0001	1,4378	0,0003	1,20	1,44
10	2	1,9113	0,0009	5,5468	0,0050	1,91	5,55
10	3	1,6961	0,0008	3,2654	0,0023	1,7	3,27
10	4	1,5951	0,0006	2,6559	0,0015	1,595	2,66
10	5	1,5299	0,0005	2,3549	0,0010	1,53	2,35
10	6	1,4849	0,0004	2,1712	0,0009	1,485	2,17
10	8	1,4229	0,0003	1,9501	0,0006	1,42	1,95
10	10	1,3799	0,0003	1,8178	0,0005	1,38	1,82
10	12	1,3440	0,0003	1,7195	0,0004	1,34	1,72
10	16	1,2884	0,0002	1,5860	0,0003	1,29	1,59
10	20	1,2525	0,0002	1,5027	0,0003	1,25	1,50
10	25	1,2218	0,0001	1,4334	0,0002	1,22	1,43
13	2	2,0547	0,0011	5,3939	0,0056	2,05	5,4
13	3	1,7855	0,0009	3,1969	0,0024	1,79	3,2
13	4	1,6653	0,0007	2,6106	0,0014	1,67	2,61
13	5	1,5892	0,0006	2,3203	0,0010	1,59	2,32
13	6	1,5362	0,0004	2,1441	0,0008	1,54	2,14
13	8	1,4650	0,0004	1,9303	0,0007	1,465	1,93
13	10	1,4162	0,0003	1,8010	0,0005	1,42	1,80
13	12	1,3760	0,0002	1,7064	0,0004	1,38	1,71
13	16	1,3159	0,0002	1,5760	0,0003	1,32	1,58
13	20	1,2763	0,0002	1,4954	0,0003	1,28	1,50
13	25	1,2423	0,0002	1,4280	0,0002	1,24	1,43
16	2	2,1658	0,0011	5,0829	0,0045	2,16	5,1
16	3	1,8509	0,0009	3,1059	0,0022	1,85	3,105
16	4	1,7140	0,0006	2,5581	0,0014	1,71	2,56

n	r	Résultats				Propositions	
		Limit-	2u	Limit+	2u	Limit-	Limit+
16	5	1,6290	0,0006	2,2826	0,0010	1,63	2,28
16	6	1,5711	0,0005	2,1139	0,0008	1,57	2,11
16	8	1,4925	0,0004	1,9092	0,0006	1,49	1,91
16	10	1,4399	0,0003	1,7844	0,0005	1,44	1,78
16	12	1,3971	0,0003	1,6924	0,0004	1,40	1,69
16	16	1,3338	0,0002	1,5667	0,0003	1,33	1,57
16	20	1,2923	0,0002	1,4885	0,0002	1,29	1,49
16	25	1,2563	0,0002	1,4229	0,0002	1,26	1,42
20	2	2,2726	0,0012	4,8840	0,0046	2,27	4,9
20	3	1,9119	0,0010	3,0377	0,0022	1,91	3,04
20	4	1,7601	0,0007	2,5151	0,0013	1,76	2,515
20	5	1,6670	0,0005	2,2514	0,0009	1,67	2,25
20	6	1,6040	0,0005	2,0880	0,0008	1,60	2,09
20	8	1,5195	0,0004	1,8916	0,0006	1,52	1,89
20	10	1,4621	0,0003	1,7693	0,0005	1,46	1,77
20	12	1,4170	0,0003	1,6808	0,0004	1,42	1,68
20	16	1,3503	0,0002	1,5582	0,0003	1,35	1,56
20	20	1,3070	0,0002	1,4816	0,0002	1,31	1,48
20	25	1,2693	0,0002	1,4177	0,0002	1,27	1,42
25	2	2,3670	0,0012	4,7362	0,0042	2,37	4,735
25	3	1,9666	0,0013	2,9788	0,0030	1,97	2,98
25	4	1,8008	0,0010	2,4798	0,0019	1,80	2,48
25	5	1,7000	0,0007	2,2244	0,0013	1,70	2,22
25	6	1,6330	0,0008	2,0675	0,0011	1,63	2,07
25	8	1,5419	0,0006	1,8753	0,0007	1,54	1,875
25	10	1,4822	0,0005	1,7572	0,0007	1,48	1,76
25	12	1,4344	0,0004	1,6699	0,0005	1,43	1,67
25	16	1,3649	0,0003	1,5505	0,0004	1,375	1,55
25	20	1,3195	0,0003	1,4754	0,0003	1,32	1,48
25	25	1,2805	0,0002	1,4130	0,0003	1,28	1,41
32	2	2,4681	0,0020	4,5297	0,0054	2,47	4,5
32	3	2,0223	0,0017	2,9112	0,0028	2,02	2,91
32	4	1,8406	0,0011	2,4372	0,0018	1,84	2,44
32	5	1,7335	0,0009	2,1944	0,0014	1,73	2,19
32	6	1,6605	0,0007	2,0451	0,0012	1,66	2,045
32	8	1,5641	0,0006	1,8579	0,0008	1,56	1,86
32	10	1,5008	0,0004	1,7423	0,0006	1,50	1,74
32	12	1,4509	0,0004	1,6582	0,0005	1,45	1,66
32	16	1,3783	0,0003	1,5417	0,0005	1,38	1,54
32	20	1,3315	0,0003	1,4692	0,0003	1,33	1,47
32	25	1,2913	0,0003	1,4082	0,0003	1,29	1,41
40	2	2,5508	0,0020	4,3943	0,0056	2,55	4,4
40	3	2,0653	0,0017	2,8600	0,0029	2,065	2,86
40	4	1,8736	0,0010	2,4060	0,0017	1,87	2,41
40	5	1,7594	0,0009	2,1700	0,0014	1,76	2,17
40	6	1,6825	0,0008	2,0245	0,0010	1,68	2,03
40	8	1,5820	0,0005	1,8436	0,0008	1,58	1,84
40	10	1,5152	0,0005	1,7317	0,0006	1,52	1,73
40	12	1,4631	0,0005	1,6491	0,0005	1,46	1,65
40	16	1,3891	0,0003	1,5349	0,0004	1,39	1,535
40	20	1,3409	0,0003	1,4634	0,0003	1,34	1,46

n	r	Résultats				Propositions	
		Limit-	2u	Limit+	2u	Limit-	Limit+
40	25	1,2994	0,0002	1,4037	0,0003	1,30	1,40
50	2	2,6262	0,0023	4,2680	0,0055	2,63	4,3
50	3	2,1049	0,0018	2,8153	0,0027	2,11	2,815
50	4	1,9014	0,0012	2,3780	0,0015	1,90	2,38
50	5	1,7817	0,0008	2,1497	0,0013	1,78	2,15
50	6	1,7018	0,0007	2,0075	0,0011	1,70	2,01
50	8	1,5969	0,0005	1,8307	0,0009	1,60	1,83
50	10	1,5284	0,0005	1,7213	0,0007	1,53	1,72
50	12	1,4750	0,0004	1,6400	0,0005	1,48	1,64
50	16	1,3986	0,0003	1,5285	0,0004	1,40	1,53
50	20	1,3490	0,0003	1,4584	0,0003	1,35	1,46
50	25	1,3066	0,0002	1,3998	0,0003	1,31	1,40
63	2	2,6975	0,0023	4,1640	0,0050	2,70	4,2
63	3	2,1410	0,0017	2,7747	0,0025	2,14	2,78
63	4	1,9281	0,0012	2,3530	0,0017	1,93	2,35
63	5	1,8030	0,0009	2,1306	0,0013	1,80	2,13
63	6	1,7197	0,0007	1,9915	0,0009	1,72	1,99
63	8	1,6110	0,0006	1,8193	0,0008	1,61	1,82
63	10	1,5403	0,0005	1,7112	0,0007	1,54	1,71
63	12	1,4854	0,0004	1,6329	0,0005	1,49	1,63
63	16	1,4067	0,0003	1,5226	0,0004	1,41	1,52
63	20	1,3563	0,0003	1,4540	0,0004	1,36	1,45
63	25	1,3131	0,0003	1,3956	0,0003	1,31	1,40
80	2	2,7648	0,0022	4,0659	0,0048	2,77	4,1
80	3	2,1736	0,0016	2,7355	0,0026	2,18	2,74
80	4	1,9517	0,0012	2,3280	0,0015	1,95	2,33
80	5	1,8222	0,0010	2,1123	0,0012	1,82	2,11
80	6	1,7363	0,0009	1,9775	0,0009	1,74	1,98
80	8	1,6237	0,0006	1,8077	0,0007	1,62	1,81
80	10	1,5511	0,0005	1,7033	0,0007	1,55	1,70
80	12	1,4943	0,0005	1,6251	0,0005	1,49	1,625
80	16	1,4143	0,0004	1,5169	0,0004	1,41	1,52
80	20	1,3629	0,0003	1,4496	0,0003	1,36	1,45
80	25	1,3188	0,0002	1,3921	0,0003	1,32	1,39
100	2	2,8187	0,0026	3,9818	0,0045	2,82	4,0
100	3	2,2036	0,0025	2,7046	0,0028	2,21	2,705
100	4	1,9728	0,0018	2,3073	0,0027	1,97	2,31
100	5	1,8373	0,0015	2,0966	0,0015	1,84	2,10
100	6	1,7495	0,0014	1,9652	0,0014	1,75	1,965
100	8	1,6345	0,0010	1,7997	0,0011	1,635	1,80
100	10	1,5597	0,0008	1,6961	0,0008	1,56	1,70
100	12	1,5021	0,0005	1,6183	0,0007	1,50	1,62
100	16	1,4203	0,0005	1,5123	0,0005	1,42	1,51
100	20	1,3683	0,0004	1,4456	0,0004	1,37	1,45
100	25	1,3236	0,0003	1,3890	0,0004	1,32	1,39
125	2	2,8739	0,0043	3,9109	0,0064	2,88	3,9
125	3	2,2272	0,0024	2,6754	0,0042	2,23	2,675
125	4	1,9889	0,0017	2,2928	0,0026	1,99	2,29
125	5	1,8521	0,0011	2,0826	0,0016	1,85	2,08
125	6	1,7616	0,0012	1,9551	0,0013	1,76	1,955
125	8	1,6443	0,0008	1,7897	0,0011	1,64	1,79

<i>n</i>	<i>r</i>	Résultats				Propositions	
		Limit-	2 <i>u</i>	Limit+	2 <i>u</i>	Limit-	Limit+
125	10	1,5677	0,0007	1,6900	0,0008	1,57	1,69
125	12	1,5083	0,0005	1,6130	0,0007	1,51	1,61
125	16	1,4259	0,0005	1,5076	0,0006	1,43	1,51
125	20	1,3730	0,0004	1,4421	0,0005	1,37	1,44
125	25	1,3278	0,0003	1,3862	0,0004	1,33	1,39
125	2	2,8699	0,0034	3,9060	0,0059	2,87	3,9
125	3	2,2282	0,0022	2,6734	0,0037	2,23	2,67
125	4	1,9885	0,0017	2,2913	0,0022	1,99	2,29
125	5	1,8515	0,0012	2,0828	0,0017	1,85	2,08
125	6	1,7613	0,0012	1,9542	0,0013	1,76	1,95
125	8	1,6440	0,0009	1,7914	0,0012	1,64	1,79
125	10	1,5669	0,0007	1,6888	0,0008	1,57	1,69
125	12	1,5087	0,0006	1,6134	0,0008	1,51	1,61
125	16	1,4265	0,0005	1,5083	0,0006	1,43	1,51
125	20	1,3730	0,0004	1,4421	0,0005	1,37	1,44
125	25	1,3279	0,0004	1,3863	0,0004	1,33	1,39
160	2	2,9259	0,0031	3,8356	0,0059	2,93	3,84
160	3	2,2510	0,0026	2,6432	0,0033	2,25	2,64
160	4	2,0074	0,0019	2,2718	0,0021	2,01	2,27
160	5	1,8650	0,0014	2,0718	0,0016	1,865	2,07
160	6	1,7739	0,0010	1,9435	0,0013	1,77	1,94
160	8	1,6523	0,0009	1,7827	0,0010	1,65	1,78
160	10	1,5755	0,0007	1,6823	0,0007	1,58	1,68
160	12	1,5151	0,0005	1,6074	0,0008	1,52	1,61
160	16	1,4313	0,0005	1,5036	0,0005	1,43	1,50

<i>n</i>	<i>r</i>	Résultats				Propositions	
		Limit-	2 <i>u</i>	Limit+	2 <i>u</i>	Limit-	Limit+
160	20	1,3775	0,0004	1,4383	0,0005	1,38	1,44
160	25	1,3313	0,0004	1,3833	0,0004	1,33	1,38
200	2	2,9669	0,0043	3,7847	0,0060	2,97	3,78
200	3	2,2736	0,0026	2,6274	0,0033	2,27	2,63
200	4	2,0217	0,0019	2,2572	0,0021	2,02	2,26
200	5	1,8778	0,0011	2,0608	0,0017	1,88	2,06
200	6	1,7823	0,0010	1,9341	0,0013	1,78	1,93
200	8	1,6607	0,0007	1,7761	0,0010	1,66	1,78
200	10	1,5810	0,0008	1,6763	0,0008	1,58	1,68
200	12	1,5207	0,0006	1,6036	0,0007	1,52	1,60
200	16	1,4358	0,0004	1,5002	0,0006	1,44	1,50
200	20	1,3812	0,0005	1,4356	0,0005	1,38	1,44
200	25	1,3345	0,0004	1,3806	0,0004	1,33	1,38
250	2	3,0066	0,0039	3,7339	0,0054	3,01	3,73
250	3	2,2910	0,0027	2,6066	0,0028	2,29	2,61
250	4	2,0331	0,0018	2,2460	0,0020	2,03	2,25
250	5	1,8867	0,0018	2,0508	0,0018	1,89	2,05
250	6	1,7902	0,0009	1,9272	0,0012	1,79	1,93
250	8	1,6670	0,0008	1,7710	0,0010	1,67	1,77
250	10	1,5865	0,0009	1,6727	0,0007	1,59	1,67
250	12	1,5253	0,0008	1,5985	0,0007	1,525	1,60
250	16	1,4394	0,0005	1,4970	0,0005	1,44	1,50
250	20	1,3847	0,0005	1,4332	0,0004	1,39	1,43
250	25	1,3376	0,0004	1,3788	0,0004	1,34	1,38

Ces résultats sont présentés dans la Figure 4, qui représente :

-  Les écarts entre les limites inférieures et les valeurs nominales du Tableau 3 (en bleu) ;
-  Les écarts entre les limites supérieures et les valeurs nominales du Tableau 3 (en orange) ;

pour les alertes correspondant à des risques nominaux α et β égaux à 0,5 % et $2 \leq r \leq 25$ en fonction de $\log(r)$, pour $n = 3 - 4 - 5 - 6 - 8 - 10 - 13 - 16 - 20 - 25 - 32 - 40 - 50 - 63 - 80 - 100 - 125 - 160 - 200$ et 250. Chaque ligne représente une valeur de n . Les limites inférieure et supérieure se rapprochent de plus en plus lorsque n augmente (c'est-à-dire que les courbes extérieures correspondent à $n = 3$ tandis que les courbes les plus proches de 0 correspondent à $n = 250$).

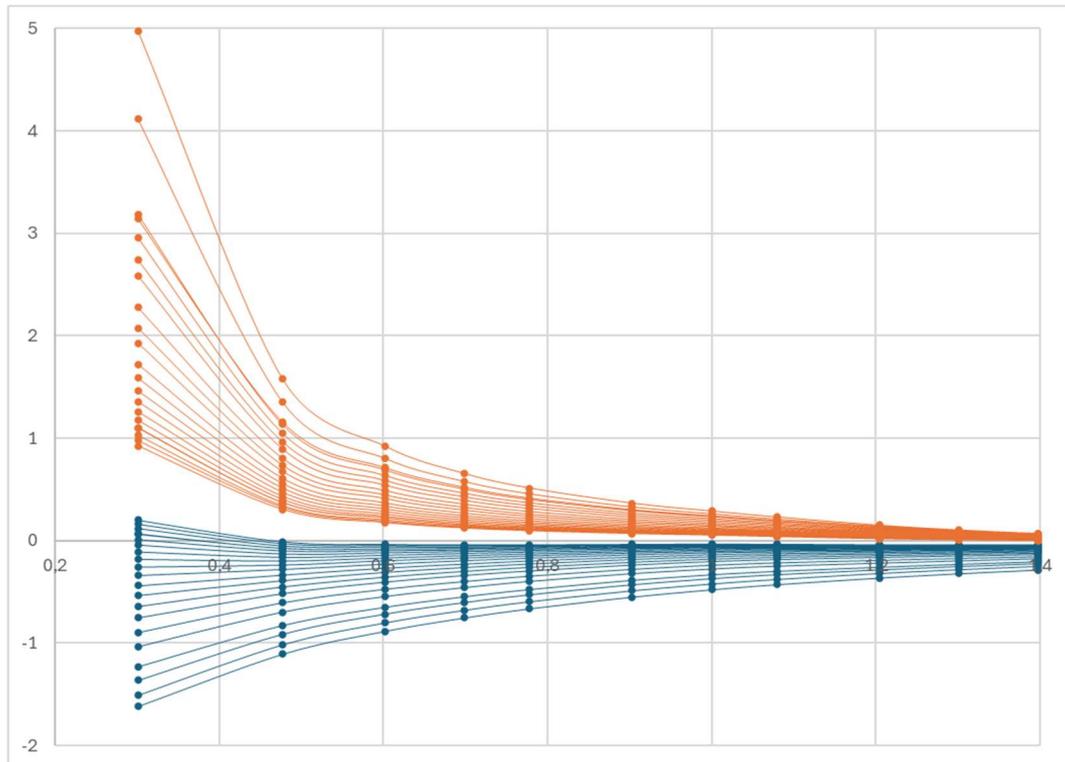


Figure 4: Écart entre les limites inférieures (en bleu) et supérieures (en orange) et les valeurs nominales pour les alertes correspondant à des risques nominaux α et β égaux à 0,5 % et $2 \leq r \leq 25$ en fonction de $\log(r)$, pour $n = 3 - 4 - 5 - 6 - 8 - 10 - 13 - 16 - 20 - 25 - 32 - 40 - 50 - 63 - 80 - 100 - 125 - 160 - 200 - 250$. (Chaque ligne représente une valeur de n , les limites inférieure et supérieure se rapprochent de plus en plus lorsque n augmente)

On peut voir sur cette figure que, pour des valeurs élevées de n et des valeurs faibles de r , l'IC90 % n'inclut pas la valeur nominale de ZR (certains points de certaines courbes bleues se trouvent au-dessus de 0). Par exemple, pour $n = 250$ et $r = 2$, la valeur nominale de ZR est 2,807 tandis que l'IC90 % est [2,97 ; 3,78]. Cela signifie que, dans ces cas, l'évaluation de la répétabilité à l'aide de notre proposition de [6] est probablement un peu trop sévère. Dans ces cas, une proposition non fondée sur l'hypothèse d'homoscédasticité pourrait également être pertinente, en alternative à la proposition actuelle. Cependant, la principale préoccupation reste le choix du niveau de risque nominal, qui est toujours conventionnel (voir § 1) et qui conduit à des différences assez importantes dans la détermination des limites (par exemple, choisir un risque nominal de 0,135 % correspondant à la limite habituelle $z = 3$ pour le biais conduirait à une limite ZR nominale de 3,205 au lieu de 2,807).

En dehors de ces cas extrêmes, les courbes des limites enveloppent assez bien les valeurs nominales du ZR, représentées par la ligne d'ordonnée « 0 » dans la Figure 4.

5.3 Conclusions

Les limites déterminées dans cette étude sont pertinentes pour l'évaluation de la répétabilité.

6 Résultats des évaluations utilisant des méthodes non paramétriques

6.1 Introduction

La plupart des EA peuvent être traités à l'aide de méthodes paramétriques, car dans la plupart des cas, on peut raisonnablement supposer qu'une loi de distribution représente de manière adéquate la distribution des résultats des essais (généralement la distribution gaussienne pour les biais et une dérivation de la loi χ^2 pour l'écart-type).

Lorsque les résultats d'essais ne suivent manifestement pas une distribution gaussienne, une transformation logarithmique ou (pour les proportions) une transformation $\log(p/(1-p))$ (où p est la proportion) permet généralement d'obtenir des résultats transformés qui suivent une distribution gaussienne.

Cependant, il peut arriver que de telles transformations soient difficiles à réaliser en raison de valeurs égales à 0 ou qu'il ne soit pas évident de trouver une transformation adéquate des résultats. Voir par exemple dans la Figure 5 une hypothétique distribution atypique des résultats d'essais.

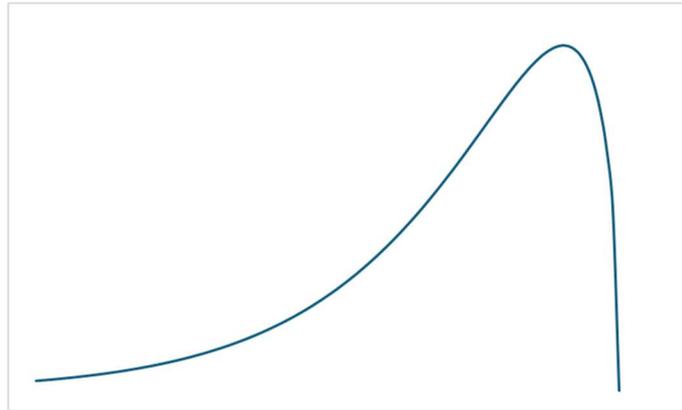


Figure 5: Exemple de distribution atypique de résultats d'essais.

De telles situations peuvent résulter de particularités de la méthode d'essai. Par exemple, la loi de distribution illustrée à la Figure 5 peut apparaître lorsque les écarts par rapport à la méthode d'essai sont plus susceptibles de produire des résultats d'essai trop faibles que des résultats d'essai trop élevés. Le fournisseur d'EA doit en tenir compte pour déterminer les limites d'alerte appropriées.

Il existe également des situations où aucune loi de distribution n'existe, par exemple pour les essais dont les résultats sont exprimés en catégories, même lorsque ces catégories sont exprimées en chiffres (par exemple, catégories 1 à 5), car dans ces cas, ces chiffres doivent être considérés comme des noms de catégories et non comme des valeurs numériques (voir l'avertissement de la norme ISO 13528 [2]).

Dans de tels cas, des méthodes non paramétriques généralement basées sur les rangs peuvent être envisagées. Cependant, la norme ISO 13528 [2] stipule que les méthodes qui déclarent un nombre fixe de valeurs aberrantes ne doivent pas être utilisées, ce qui rend implicitement toutes les méthodes non paramétriques non recommandées. En effet, les méthodes non paramétriques utilisent généralement des distributions binomiales pour déterminer les probabilités associées à un rang de valeur classée. L'application de cette méthode aux valeurs extrêmes permet de déterminer leur probabilité d'occurrence et, par conséquent, de déterminer si elles doivent être considérées comme des valeurs aberrantes ou non. Un rang peut alors être associé au risque nominal et de chaque côté d'un IC90 %.

Il est évident que nous ne pouvons pas nous attendre à ce que les méthodes non paramétriques soient aussi efficaces que les méthodes paramétriques, elles ne doivent donc être utilisées que lorsque les autres méthodes ne peuvent pas être appliquées.

Dans notre étude, nous avons pris en compte deux principes de base pour calculer les alertes, comme indiqué au § 3.5 :

- ✚ Le premier considère que les valeurs aberrantes sont celles situées aux extrémités de la distribution (option généralement choisie pour l'évaluation du biais, voir § 4.1) ;
- ✚ Le second considère que tous les participants présentent la même valeur (option généralement choisie pour l'évaluation de l'écart-type, voir § 5.1). Si elle est utilisée pour le biais, cette option conduirait à déclarer la plupart des participants comme valeurs aberrantes (voir § 3.5).

Dans le premier principe de base, nous avons ensuite utilisé des risques α et β de 1 % bilatéraux (soit 0,5 % de chaque côté de la distribution). Dans ces conditions, la loi binomiale peut être approximée par la loi de Poisson, avec $p = 0,005$ et n égal au nombre de participants ($n.p$ est généralement inférieur à 1). Les calculs consistent alors à calculer les rangs pour lesquels les résultats de la loi de Poisson cumulative sont inférieurs à 5 % ou supérieurs à 95 % afin de déterminer l'IC90 % correspondant, en fonction de n .

Pour le deuxième principe de base, nous avons considéré que tous les participants sont supposés avoir la même valeur, c'est-à-dire que chacun d'entre eux a une probabilité égale à 0,5 d'être supérieur ou inférieur à la valeur de référence. Les calculs consistent alors à calculer les rangs pour lesquels les résultats de la loi binomiale cumulative sont inférieurs à 5 % ou supérieurs à 95 % afin de déterminer l'IC90 % correspondant, en fonction de n .

Dans les deux cas, l'utilisation du MMC n'est pas nécessaire.

Évidemment, le même nombre de participants reçoivent des signaux à chaque extrémité de la distribution, de sorte que les valeurs n critiques pour les alertes déclenchent en fait un signal pour 2 participants supplémentaires : un à l'extrémité inférieure de la distribution et un à l'extrémité supérieure de la distribution.

Une autre situation délicate qui peut se présenter lors de l'utilisation de cette méthode est celle où, en raison de l'arrondi, les mêmes valeurs de résultats d'essais doivent être considérés comme aberrants et non aberrants. Par exemple, si les 5 résultats les plus élevés sont égaux et que 2 d'entre eux doivent être considérés comme aberrants, que faut-il décider pour ces 5 résultats égaux les plus élevés ? Une option consiste à comparer le nombre de résultats égaux et le nombre de signaux à déclarer parmi ces résultats afin de décider si tous ou aucun d'entre eux doivent déclencher un signal. Par exemple, dans le cas présent, comparez 2 résultats « aberrants » et 3 résultats « non aberrants », puis décidez que les 5 ne sont pas aberrants. En cas d'équilibre (par exemple, lorsque 2 résultats « aberrants » et 2 résultats « non aberrants » doivent être comparés), il convient de considérer que toutes les valeurs ne sont pas aberrantes en raison des règles d'arrondi utilisées pour calculer les valeurs des Tableau 5, Tableau 6 and Tableau 7.

6.2 Résultats pour le premier principe de base

Le Tableau 5 présente les résultats de la détermination du nombre de participants à chaque extrémité de la distribution qui devraient recevoir un signal d'alerte ou un signal d'action, correspondant aux risques nominaux α et β appartenant aux 1 % (bilatéral) extrêmes de la distribution.

Tableau 5. Résultats de détermination du nombre de participants à chaque extrémité de la distribution qui devraient recevoir un signal d'alerte ou un signal d'action, correspondant aux risques nominaux α et β appartenant aux 1 % (bilatéral) extrêmes de la distribution.

Signaux	$2 \leq n \leq 10$	$11 \leq n \leq 71$	$72 \leq n \leq 163$	$164 \leq n \leq 273$	$274 \leq n \leq 394$	$395 \leq n \leq 460$	$461 \leq n \leq 522$	$523 \leq n \leq 657$
Alerte	0	1	2	3	4	5	5	6
Action	0	0	0	0	0	0	1	1

Par exemple, lors d'un EA impliquant 100 participants, les 2 résultats les plus bas et les 2 résultats les plus élevés déclenchent un signal d'alerte, mais aucun signal d'action n'est déclenché.

Sans surprise, nous pouvons constater que cette méthode n'est pas très efficace, car il faut 461 participants pour déclencher un seul signal d'action. Elle ne peut donc jamais être utilisée dans la pratique.

6.3 Résultats pour le deuxième principe de base

Le Tableau 6 présente les résultats de la détermination du nombre de participants à chaque extrémité de la distribution qui devraient recevoir un signal d'alerte ou un signal d'action, correspondant à des risques nominaux α et β égaux à 1 % (bilatéral) différents de la valeur de référence.

Tableau 6. Résultats de détermination du nombre de participants à chaque extrémité de la distribution qui devraient recevoir un signal d'alerte ou un signal d'action, correspondant à des risques nominaux α et β égaux à 1 % (bilatéral) différents de la valeur de référence.

	$2 \leq n \leq 4$	$5 \leq n \leq 7$	$8 \leq n \leq 10$	$11 \leq n \leq 12$	$13 \leq n \leq 15$	$16 \leq n \leq 17$	$18 \leq n \leq 20$	$21 \leq n \leq 22$
Alertes	All	All	All	All	All	All	All	All
Actions	0	1	2	3	4	5	6	7

Seuls les résultats pour $n \leq 22$ sont fournis dans ce tableau, mais nous avons pu vérifier que tous les participants reçoivent une alerte pour toutes les valeurs de n habituellement rencontrées lors des EA. Il s'agit donc d'une méthode qui ne peut jamais être utilisée dans la pratique.

6.4 Adaptation des niveaux de risque et d'IC aux cas où des méthodes non paramétriques doivent être utilisées

Pour faire face aux situations décrites aux § Erreur ! Source du renvoi introuvable. et § 6.3, nous avons déterminé une autre série de limites en utilisant des risques nominaux α et β appartenant aux de 10 % (bilatéral) extrémaux de la distribution (au lieu de 1 %) et un IC80 % (au lieu de 90 %). Le Tableau 7 présente les résultats de ces déterminations.

Tableau 7. Résultats de la détermination du nombre de participants à chaque extrémité de la distribution qui devraient recevoir un signal d'alerte ou un signal d'action, correspondant aux risques nominaux α et β appartenant aux queues 10 % (bilatéral) extrémaux de la distribution.

	Alerte	Action		Alerte	Action		Alerte	Action
$n \leq 2$	0	0	$64 \leq n \leq 77$	6	1	$141 \leq n \leq 156$	11	4
$3 \leq n \leq 10$	1	0	$78 \leq n \leq 93$	7	2	$157 \leq n \leq 159$	12	4
$11 \leq n \leq 22$	2	0	$94 \leq n \leq 106$	8	2	$160 \leq n \leq 172$	12	5
$23 \leq n \leq 34$	3	0	$107 \leq n \leq 108$	8	3	$173 \leq n \leq 185$	13	5
$35 \leq n \leq 46$	4	0	$109 \leq n \leq 124$	9	3	$186 \leq n \leq 206$	13	6
$47 \leq n \leq 48$	4	1	$125 \leq n \leq 133$	10	3	$207 \leq n \leq 210$	14	6
$49 \leq n \leq 63$	5	1	$134 \leq n \leq 140$	10	4			

Ces résultats montrent que cette méthode peut être utilisée lorsque le nombre de participants est suffisamment important. Lorsque le nombre de participants est faible, une grande quantité des signaux d'alerte sont déclenchés, mais aucun signal d'action ne l'est, ce qui reflète le manque de puissance de la méthode, qui permet difficilement de déterminer si un résultat est correct ou non.

6.5 Utilisation de cette méthode pour les résultats d'essais non numériques

Les résultats d'essais non numériques (c'est-à-dire les résultats d'essais exprimés en termes de catégories) peuvent être classés en trois types :

1. Catégories ordonnées (par exemple, le caractère liquoreux ou non d'un vin) ;
2. Catégories non ordonnées (relativement la propriété qu'elles sont censées représenter), voir un exemple fourni au § E15 de la norme ISO 13528 [2] ;
3. Résultats binaires (par exemple, réussite/échec), qui peuvent être considérés comme des résultats de type 1 avec seulement deux catégories.

Pour le 2^e type, la norme ISO 13528 [2] propose d'établir un classement en fonction de la fréquence d'occurrence dans les résultats des essais (par exemple, une série de résultats d'essais tels que A : 5 – B : 13 – C : 3 – D : 27, les catégories peuvent alors être classées comme suit : D – B – A – C).

La méthode exposée au § 6.4 peut alors être appliquée aux résultats des essais classés. Il est évident que des contrôles unilatéraux doivent être utilisés pour les types 2 et 3, car il n'est pas logique de considérer les résultats des essais les plus fréquents comme des valeurs aberrantes. Les chiffres du § 6.4 peuvent alors être utilisés sur un seul côté de la distribution, avec des risques α et β associés de 5 % au lieu de 10 %.

Par exemple, pour la série de tests non ordonnée mentionnée ci-dessus (c'est-à-dire A : 5 – B : 13 – C : 3 – D : 27), la procédure suivante devrait être mise en œuvre :

1. Le nombre total de résultats d'essais est de 48, ce qui devrait déclencher 4 signaux d'alerte et 1 signal d'action (voir Tableau 7) ;
2. Le signal d'action doit être attribué à un résultat d'essai « C » ;
3. Cependant, il y a 3 résultats d'essai « C ». Conformément aux dispositions du § 6.1 (1 est inférieur ou égal à la moitié de 3), aucun signal d'action ne doit être déclenché. Le signal d'action doit alors être converti en signal d'alerte ;
4. Au total, 4 + 1 = 5 signaux d'alerte doivent donc être déclenchés ;
5. Les 3 résultats de test « C » déclenchent donc un signal d'alerte ;
6. Il reste donc 2 signaux d'alerte à attribuer aux résultats d'essai « A », qui suivent les résultats « C » dans le classement ;
7. Compte tenu des dispositions du § 6.1 (2 est inférieur ou égal à la moitié de 5), aucun signal d'alerte ne doit être déclenché pour les résultats d'essai « A » ;
8. En conclusion générale, les résultats d'essai « A », « B » et « D » peuvent être acceptés, tandis que les résultats d'essai « C » déclenchent un signal d'alerte.

6.6 Conclusions pour les limites déterminées à l'aide de méthodes non paramétriques

Les méthodes non paramétriques peuvent également être utilisées pour déterminer les limites d'alerte, qui sont alors exprimées en termes de rangs plutôt qu'en termes de scores.

Sans surprise, ces méthodes sont moins efficaces et moins puissantes que les méthodes paramétriques. C'est pourquoi elles ne sont pas recommandées par les normes de référence. Elles ne doivent être utilisées que lorsque les méthodes paramétriques ne peuvent pas être utilisées, car aucune loi de distribution ne peut être raisonnablement supposée ou n'existe, à condition qu'un nombre suffisant de participations soit disponible.

7 Conclusions

Nous avons pu déterminer des limites pour les alertes en fonction du nombre de participations pour l'évaluation du biais ainsi que pour l'évaluation de la répétabilité.

Pour le biais, les limites classiques 2 et 3 pour les scores z ne sont très éloignés de la situation où les risques α et β sont égaux à 1 % bilatéral avec un IC90 % lorsque $n = 110$. Des valeurs inférieures de n nécessitent une « bande de doute » plus large autour de la valeur nominale 2,576, tandis que des valeurs supérieures de n permettent d'en réduire l'ampleur. Des formules empiriques ont pu être déterminées pour le calcul des limites, en fonction du nombre de participants. Ces limites s'appliquent à la valeur apparente de σ_{pt} , qui est généralement supérieure à σ_L , ce qui réduit le risque α et augmente le risque β . Lorsque le rapport λ défini comme $\lambda = \sigma_r / (r \cdot \sigma_L)$ est inférieur à 0,17, cet effet devient négligeable et aucune différence entre les risques α et β ne s'applique alors.

Pour les évaluations de répétabilité, lorsque les valeurs de n sont élevées et celles de r faibles, l'IC90 % n'inclut pas la valeur nominale de ZR . Cela signifie que, dans ces cas, l'évaluation de la répétabilité à l'aide de la proposition de [6] est probablement un peu trop stricte. Dans ces cas, une proposition qui ne serait pas fondée sur l'hypothèse d'homoscédasticité pourrait également être pertinente, en alternative à la proposition de ce document. Cependant, la principale préoccupation reste le choix du niveau de risque nominal, qui est toujours conventionnel (voir § 1) et qui conduit à des différences assez importantes dans la détermination des limites. En dehors de ces cas extrêmes, les limites déterminées englobent assez bien les valeurs nominales de ZR .

Des méthodes non paramétriques peuvent également être utilisées pour déterminer les limites d'alerte, qui sont alors exprimées en termes de rangs plutôt qu'en termes de scores. Sans surprise, ces méthodes sont moins efficaces et moins puissantes que les méthodes paramétriques. C'est pourquoi elles ne devraient être utilisées que lorsque les méthodes paramétriques ne peuvent pas être utilisées, car aucune loi de distribution ne peut être raisonnablement supposée ou même exister, à condition qu'un nombre suffisant de participations soit disponible.

8 Références

- [1] ISO 5725-2:2019, Exactitude (justesse et fidélité) des résultats et méthodes de mesure - Partie 2: Méthode de base pour la détermination de la répétabilité et de la reproductibilité d'une méthode de mesure normalisée
- [2] ISO 13528:2022, Méthodes statistiques utilisées dans les essais d'aptitude par comparaison interlaboratoires
- [3] ISO/IEC 17043:2023, Évaluation de la conformité — Exigences générales concernant la compétence des organisateurs d'essais d'aptitude
- [4] L.J. Hollebecq, "Risque β lors des essais d'aptitude selon le nombre de participants", Publications techniques CompaLab, Décembre 2022, DOI: <https://www.compalab.org/medias/files/publication-interne-risque-beta-fr.pdf>
- [5] L.J. Hollebecq, "Beta risk in proficiency testing in relation with the number of participants", ACTA IMEKO, Vol. 12, Nr 3, 1-9, September 2023 DOI: <https://doi.org/10.21014/actaimeko.v12i3.1433>
- [6] L.J. Hollebecq, "Essais d'aptitude pour la répétabilité", Publications techniques CompaLab, Juillet 2025, DOI: <https://www.compalab.org/medias/files/publication-ea-et-fr.pdf>
- [7] David Luengo, Luca Martino, Mónica Bugallo, Víctor Elvira and Simo Särkkä, "A survey of Monte Carlo methods for parameter estimation", EURASIP Journal on Advances in Signal Processing, Article 25, May 2020 DOI: <https://doi.org/10.1186/s13634-020-00675-6>
- [8] L.J. Hollebecq, "Intervalles de confiance sur écarts-types emboîtés", Publications techniques CompaLab, Janvier 2025, DOI: <https://www.compalab.org/medias/files/publication-interne-estimation-et-composes-fr.pdf>