

L'interprétation et la communication des scores en neuropsychologie

Conférence de consensus



Table des matières

01 Introduction

02 Une conférence de consensus

03 L'interprétation et la communication des scores en neuropsychologie : une tour de Babel ?

04 La logique de la comparaison normative : seuil de décision et risque d'erreur

05 Quels outils statistiques choisir ?
Quel seuil de décision fixer ?

06 Au-delà du score seuil : considérations lors de l'interprétation

07 Des descripteurs sans ambiguïtés

08 Conclusion

01

02

03

04

05

06

07

08

Introduction



02

Une conférence de consensus ?

Jacques Grégoire, Université de Louvain



01

02

03

04

05

06

07

08



Objectifs d'une conférence de consensus

- Méthode permettant de déterminer des règles de bonne pratique dans un domaine précis d'une discipline donnée.
- Elle est utilisée lorsque certaines pratiques professionnelles sont insatisfaisantes (hétérogènes, floues ou controversées).
- Son objectif est de dégager un consensus sur des pratiques professionnelles recommandées au regard des connaissances scientifiques les plus récentes.



Les acteurs d'une conférence de consensus

- **Les promoteurs:** les scientifiques et les professionnels qui identifient les problèmes posés sur le terrain, qui formulent des questions clés et lancent le projet d'une conférence de consensus pour y répondre.
- **Les experts:** spécialistes du domaine chargés de faire le point sur les connaissances scientifiques actuelles à propos des questions posées. Ils rassemblent la documentation, identifient les points faisant débat et formulent des propositions.
- **Le jury :** comprend des membres issus des différents groupes concernés par les questions de pratique débattues (praticiens, chercheurs, patients, juristes...). Le président organise le travail du jury.



Les étapes d'une conférence de consensus

1. Formulation des questions par les promoteurs = cadre pour le travail des experts et du jury.
2. Rassemblement de la documentation scientifique et rédaction de documents de référence par les experts.
3. Organisation de la conférence proprement dite:
 - Présentation des synthèses préparées par les experts devant le jury et un public de personnes concernées par les questions de pratique abordées.
 - Echanges entre les experts, le jury et le public (questions, commentaires, suggestions).



Les étapes d'une conférence de consensus

4. Travail du jury: Prises de position par rapport aux propositions présentées lors de la conférence => recherche d'un consensus et formulation de recommandations.
5. Diffusion des recommandations dans la communauté des praticiens accompagnées des documents rédigés par les experts.





Les recommandations

- = Règles de bonne pratique, chacune formulée de manière brève dans un langage accessible par tous praticiens.
- Généralement accompagnées d'un court commentaire.
- Les argumentaires et les informations techniques (tableaux statistiques, modalités de calcul...) ne font pas partie des recommandations, mais sont présentés dans les documents annexes préparés par les experts.
- La méthode de conférence de consensus donne leur légitimité aux recommandations, car celles-ci prennent en compte les connaissances scientifiques récentes et les avis exprimés par un panel représentatif de chercheurs, praticiens et autres citoyens concernés.



03

L'interprétation et la communication des scores en neuropsychologie : une tour de Babel ?

Sylvie Willems, Université de Liège



01

02

03

04

05

06

07

08



Interpréter et communiquer nos résultats



Score
Brut



Restitution

Normes

Choix d'un seuil
de « normalité »

Choix des mots
« descripteurs »



Interpréter et communiquer nos résultats



Score
Brut

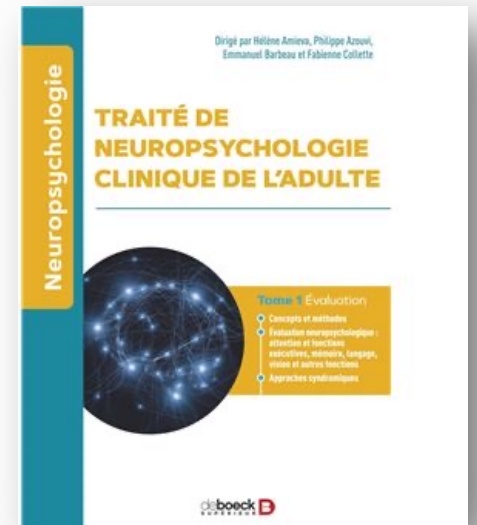


Restitution

Normes

Choix d'un seuil
de « normalité »

Choix des mots
« descripteurs »



Multiplés principes d'étiquetage

Suisse : Consensus ASNP (2014, 2018)

Tableau 1: Classification et interprétation clinique des résultats aux tests neuropsychologiques

Percentiles	Valeurs T	Valeurs z	QI	Description des valeurs statistiques (classification)	Interprétation clinique
>98 - 100	>70 - 80	>2.0 - 3.0	>130 - 145	très supérieur à la norme	
≥95 - ≤98	≥66.4 - ≤70	≥1.6 - ≤2.0	≥125 - ≤130	clairement supérieur à la norme	
>84 - <95	>60 - <66.4	>1.0 - <1.6	>115 - <125	à la limite supérieure de la norme	
≥16 - ≤84	≥40 - ≤60	≥-1.0 - ≤1.0	≥85 - ≤115	dans la norme	sans particularité
>5 - <16	>33.6 - <40	>-1.6 - <-1.0	>75 - <85	à la limite inférieure de la norme	possiblement déficitaire
≥2 - ≤5	≥30 - ≤33.6	≥-2.0 - ≤-1.6	≥70 - ≤75	clairement inférieur à la norme	déficitaire
0 - <2	20 - <30	-3.0 - <-2.0	55 - <70	très inférieur à la norme	sévèrement déficitaire

Multiplés principes d'étiquetage

USA : American Academy of Clinical Neuropsychology consensus conference (Guilmette et al., 2020)

Table 1. Recommended test score labels based on standard scores and percentiles for tests with normal distributions.

Standard Score	Percentile	Score Label
>130	≥98	Exceptionally high score
120–129	91–97	Above average score
110–119	75–90	High average score
90–109	25–74	Average score
80–89	9–24	Low average score
70–79	2–8	Below average score
<70	<2	Exceptionally low score

Divergence entre neuropsychologues

Guilmette et al. (2008) – 110 neuropsychologues américains

Nombre d'étiquettes différentes associées à un même score : **moyenne = 14 (range 9 à 23)**

Divergence entre neuropsychologues pour ...

- La terminologie
- Les seuils

The Clinical Neuropsychologist
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854040601064559

 Psychology Press
Taylor & Francis Group

ASSIGNING QUALITATIVE DESCRIPTIONS TO TEST SCORES IN NEUROPSYCHOLOGY: FORENSIC IMPLICATIONS

Thomas J. Guilmette¹, Leigh D. Hagan², and Anthony J. Giuliano³

¹Providence College, Brown Medical School, Southern New England Rehabilitation Center, Providence, RI, ²Private Practice, Chesterfield, VA, and ³Harvard Medical School, Cambridge, MA, Beth Israel-Deaconess Medical Center, Boston, MA

Divergence entre neuropsychologues

Leclef et al. (2018) – 547 psychologues spécialisés en neuropsychologie

- La majorité des professionnels présentent les scores accompagnés de descripteurs dans leurs rapports.
- Plus de la moitié ne fournissent pas la classification utilisée pour décrire les performances mesurées aux tests.
- Grande variabilité dans les seuils utilisés.

PRATIQUES CLINIQUES

L'interprétation des scores en neuropsychologie : la tour de Babel ?



Pierre LECLEF

Psychologue spécialisé en neuropsychologie, Cabinet de Neuropsychologie de Nevers (58)



Amélie PONCHEL

Psychologue spécialisée en neuropsychologie et docteur en psychologie.
Unité de NeuroPsychiatrie Comportementale, Hôpital La Pitié-Salpêtrière, Paris (75).
Vice-présidente et membre du conseil d'administration de l'OFPN



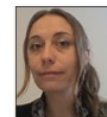
Sophie CHANCENOTTE

Psychologue spécialisée en neuropsychologie, CRTLA et pédopsychiatrie CHU Dijon, activité libérale (21)



Sandra MAREY

Psychologue spécialisée en neuropsychologie, activité libérale et CHU Hôpital d'Enfants Pédiatrie, Dijon (21)



Mathilde MUNEUX

Psychologue spécialisée en neuropsychologie, activité libérale et ADAPÉI (04)

mathildemuneux@gmail.com

Mots-clés

- Neuropsychologie clinique
- Psychométrie
- Statistiques
- Enquête nationale

Résumé

Pour répondre au plus juste à la plainte du patient, l'évaluation neuropsychologique prend une importance croissante, que ce soit dans le domaine sanitaire, médico-social ou libéral. Pour être rigoureuse et objective, cette évaluation nécessite l'utilisation de tests standardisés. Cependant, il existe un nombre important de tests dont le choix varie en fonction de la population, des processus cognitifs ciblés, de la pathologie concernée et des étalonnages (i.e., transformations des scores bruts en notes étalonnées). Cette diversité pose la question de l'homogénéité de l'interprétation des résultats aux tests au sein de la profession. Dresser un état des lieux de ces pratiques est donc l'objectif de cet article. Cette démarche s'appuie sur une enquête menée à l'échelle nationale grâce au soutien de l'Organisation Française des Psychologues spécialisés en Neuropsychologie (OFPN). Les résultats montrent une hétérogénéité importante dans l'interprétation des données psychométriques en fonction du secteur d'intervention du psychologue (type de structure, type de population rencontrée) et du nombre d'années d'exercice. Cet article se propose alors de discuter non seulement des conséquences de cette hétérogénéité sur les réponses apportées au patient mais aussi des actions à mener pour la promotion d'une pratique consensuelle, scientifique et rigoureuse de la psychométrie en neuropsychologie clinique.

Divergence au sein d'un rapport

- Changement de termes au fil du rapport (goût pour le synonyme).
- Changements de seuil d'un test à l'autre (en fonction de l'étalonnage utilisé – percentile, score Z, score T, ...).

PRATIQUES CLINIQUES

L'interprétation des scores en neuropsychologie : la tour de Babel ?



Pierre LECLEF

Psychologue spécialisé en neuropsychologie, Cabinet de Neuropsychologie de Nevers (58)



Amélie PONCHEL

Psychologue spécialisée en neuropsychologie et docteur en psychologie.

Unité de NeuroPsychiatrie Comportementale, Hôpital La Pitié-Salpêtrière, Paris (75).
Vice-présidente et membre du conseil d'administration de l'OFPN



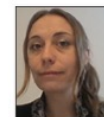
Sophie CHANCENOTTE

Psychologue spécialisée en neuropsychologie, CRTLA et pédopsychiatrie CHU Dijon, activité libérale (21)



Sandra MAREY

Psychologue spécialisée en neuropsychologie, activité libérale et CHU Hôpital d'Enfants Pédiatre, Dijon (21)



Mathilde MUNEUX

Psychologue spécialisée en neuropsychologie, activité libérale et ADAPÉI (04)

mathildemuneux@gmail.com

Mots-clés

- Neuropsychologie clinique
- Psychométrie
- Statistiques
- Enquête nationale

Résumé

Pour répondre au plus juste à la plainte du patient, l'évaluation neuropsychologique prend une importance croissante, que ce soit dans le domaine sanitaire, médico-social ou libéral. Pour être rigoureuse et objective, cette évaluation nécessite l'utilisation de tests standardisés. Cependant, il existe un nombre important de tests dont le choix varie en fonction de la population, des processus cognitifs ciblés, de la pathologie concernée et des étalonnages (i.e., transformations des scores bruts en notes étalonnées). Cette diversité pose la question de l'homogénéité de l'interprétation des résultats aux tests au sein de la profession. Dresser un état des lieux de ces pratiques est donc l'objectif de cet article. Cette démarche s'appuie sur une enquête menée à l'échelle nationale grâce au soutien de l'Organisation Française des Psychologues spécialisés en Neuropsychologie (OFPN). Les résultats montrent une hétérogénéité importante dans l'interprétation des données psychométriques en fonction du secteur d'intervention du psychologue (type de structure, type de population rencontrée) et du nombre d'années d'exercice. Cet article se propose alors de discuter non seulement des conséquences de cette hétérogénéité sur les réponses apportées au patient mais aussi des actions à mener pour la promotion d'une pratique consensuelle, scientifique et rigoureuse de la psychométrie en neuropsychologie clinique.



Des descripteurs ambigus

Termes axés sur :

- L'**incertitude**
- La **signification clinique** (défiance présumée)
- La **position** vis-à-vis d'une moyenne
- L'**estimation prémorbide**
- La **fréquence**

Exemples :

- « Performance limite »
- « Performance déficitaire », « performance faible »
- « Score inférieur à la moyenne »
- « Score inférieur au niveau attendu »
- « Score dans la norme »





Notre démarche

**Identification des
Promoteurs puis des
experts**



**Réunion des experts et
choix de la méthode**

Procédure

**Définition de la question
Production d'un document
par un sous-groupe**



**Révision du document
écrit par les experts**

Conférence de consensus



Recommandations





Procédure

Etape 1

Fin 2022

Une nécessité de consensus?

Xavier Seron

Sylvie Willems

Catherine Belin (GRECO)

Hélène Amieva (SNLF)

Philippe Azouvi (SNLF)

Identification des 'Promoteurs'
(SNLF, OFPN, GRECO)

Puis identification de 13 'Experts'
-> **Groupe Label**





Procédure

1. Choix d'une conférence de consensus (Présidence : Jacques Grégoire)
2. Définition de la question
3. Désignation d'un **groupe de travail restreint**
(Patrick Fery ; Hichem Slama ; Béatrice Degraeve ; Sylvie Willems; François Radiguer)

Objectif :

- **Spécification des contours de la question**
- **Production d'un document de position à réviser**

2 Etape

Mars 2023

« Groupe Label »

George Michael
Hélène Amieva
François Radiguer
Vincent Verdon
Philippe Azouvi
Christine Moroni
Martine Roussel

Amélie Ponchel
Sylvie Willems
Hichem Slama
Béatrice Degraeve
Patrick Fery
(Jacques Grégoire)





Procédure - Processus de création

Etape 3

Mars – Novembre 2023

« Groupe Label » restreint



4 Etape

Novembre 2023 – Avril 2024

« Groupe Label » élargi

Spécification de la question



Production du document de position des 'experts'





Procédure

Etape 5

Choix d'un Jury représentatif

- Chercheurs en neuropsychologie et psychométrie
- Clinicien.ne.s psychologues spécialisé.e.s en neuropsychologie
- Médecins spécialisé.e.s en neuropsychologie
- Experts de vécu

Thierry Lecerf ; Bruno Lenne ; Pierre-Yves Jonin ; Marie Geurten ; Philippe Allain; Emilie Favre ; Delphine Fleurion ; Giulia Dormal ; Valérie Vanderaspoilden; Catherine Belin ; Pascale Pradat-Diehl ; Olivier Godefroy ; Caroline Massot ; Cécile Donze ; Audrey Prégaldien ; Estelle Gernez





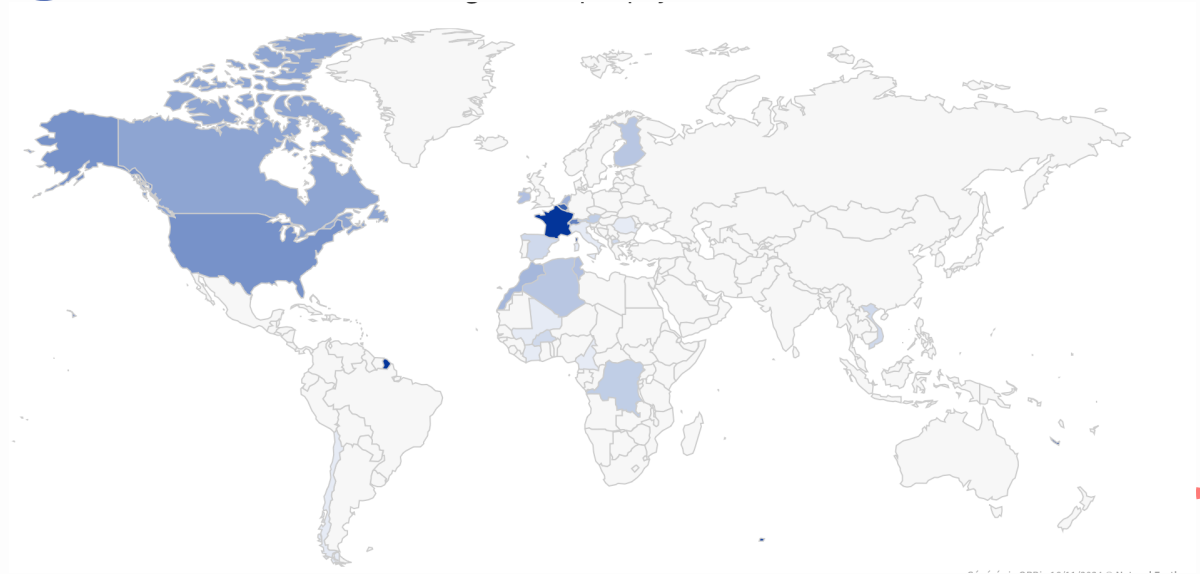
Procédure

Etape 5

Diffusion du document de position

9 septembre 2024

(> 1400 téléchargements)

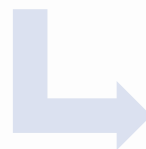




Définition de la question

1 **Question principale =**
Quels qualificatifs employer pour définir un score ?

2 **Question associée =**
Quel seuil pour qualifier un score de « hors norme » ?



Outils statistiques (Score Z, percentile...) et certaines propriétés des normes





Exclusion de questions pourtant fondamentales

- 1** Qualité et représentativité des normes
- 2** Standardisation de l'administration
- 3** Qualité psychométrique de l'outil (ce y compris, la sensibilité et spécificité)
- 4** Scores composites
- 5** Interprétation clinique





Analyse statistique « pré-clinique »

« Le résultat est-il hors-norme ? »

Choix du seuil et de la statistique

Considérations des variables affectant l'interprétation statistique

- Bilatéral ou unilatéral?
- Nombre de tâches?
- Hypothèses a priori?
- ...



Interprétation « clinique » post-statistique

Par exemple, la performance est-elle indicatrice d'une déficience ?

Considération du niveau de performance présumé

Intégration de l'ensemble des données dans une interprétation clinique cohérente

- Ensemble des épreuves
- Données médicales
- Données anamnestiques
- Observations
- ...





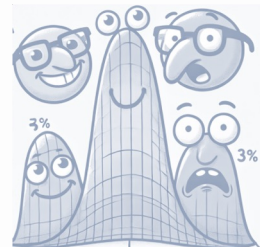
Questions ou commentaires ?



04

La logique de la comparaison normative : seuil de décision et risque d'erreur

Hichem Slama (Hôpital Universitaire de Bruxelles –
H.U.B, Université libre de Bruxelles – ULB)



01

02

03

04

05

06

07

08

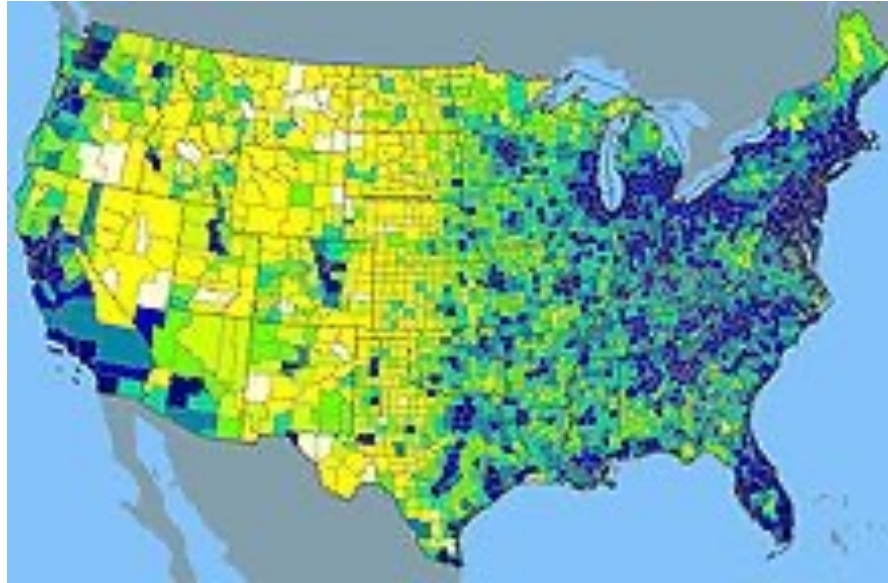
Répondre à une question

Prendre une décision



Va-t-il gagner ?

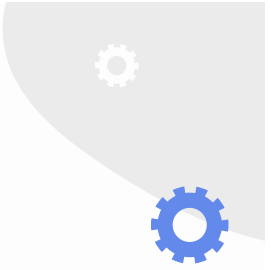




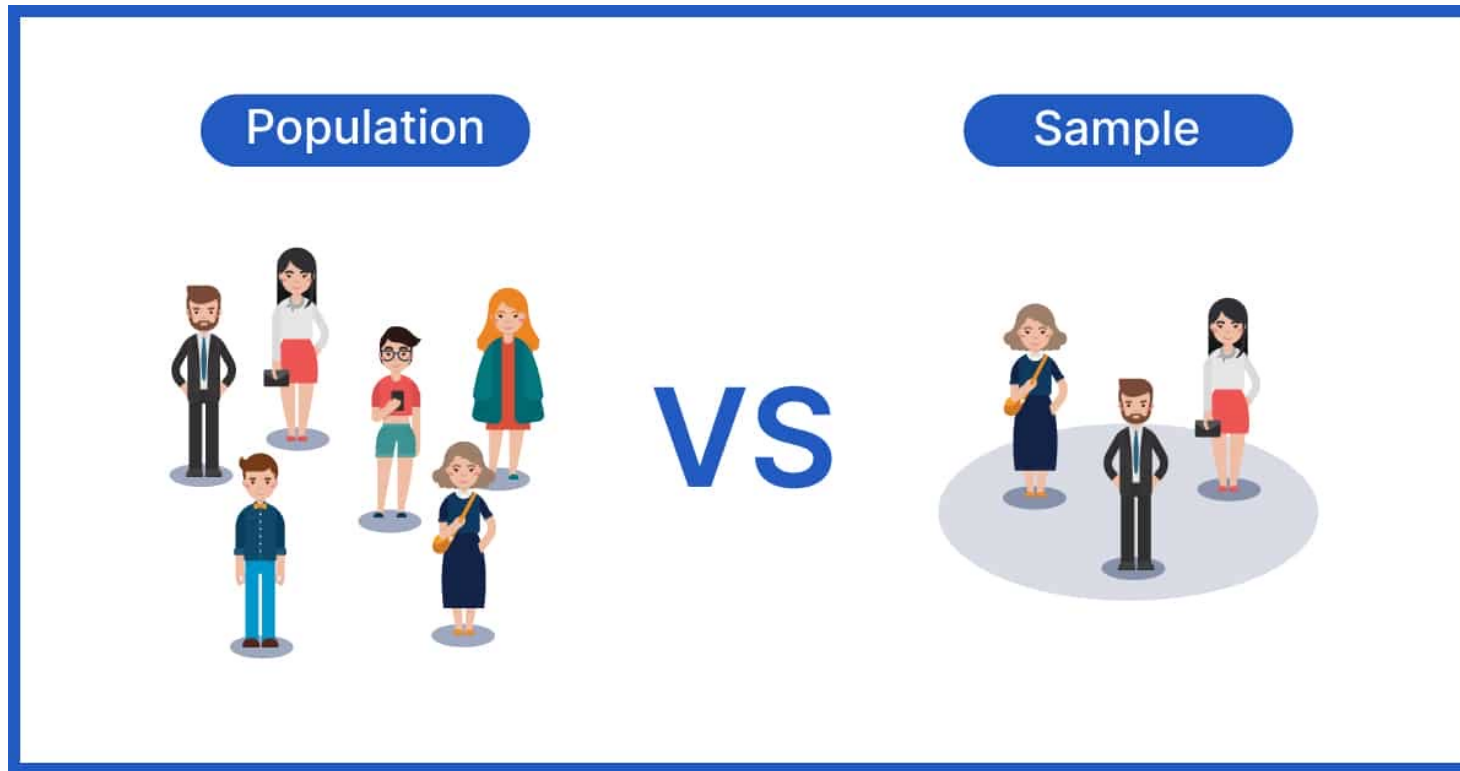
334,9 millions
de personnes



Sondage



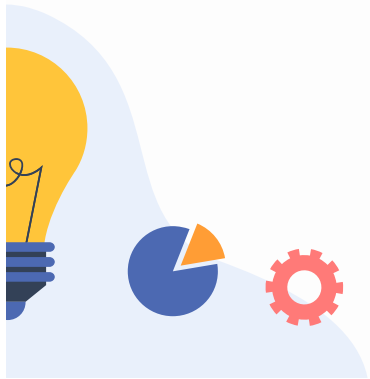
Distribution d'échantillonnage





Le concept clef qui sous-tend tous les tests statistiques est la distribution d'échantillonnage d'une statistique

Sans distribution d'échantillonnage, nous n'aurions pas de statistiques





Qualité de la réponse dépend:

- De la qualité de l'outil de mesure
 - Pas notre propos dans cette conférence (qualités psychométriques: validité, fidélité, sensibilité, spécificité)
- De la nature de l'échantillon
 - Pas notre propos dans cette conférence (représentativité de l'échantillon, critères d'inclusion/exclusion,...)
- De la qualité de la prise de données
 - Pas notre propos dans cette conférence (qualité de la construction normative)
- De la taille de l'échantillon
 - Élément central de cette conférence de consensus



Tableau 3.5 Distribution de l'échantillon d'étalonnage selon le sexe, par groupe d'âge

Groupe d'âge	N Garçons	N Filles	Total
6 ans 0 mois 0 jour			
à 6 ans 5 mois 30 jours	18	21	39
6 ans 6 mois 0 jour			
à 6 ans 11 mois 30 jours	33	30	63
7 ans 0 mois 0 jour			
à 7 ans 5 mois 30 jours	28	25	53
7 ans 6 mois 0 jour			
à 7 ans 11 mois 30 jours	23	25	48
8 ans 0 mois 0 jour			
à 8 ans 5 mois 30 jours	23	23	46
8 ans 6 mois 0 jour			
à 8 ans 11 mois 30 jours	27	28	55
9 ans 0 mois 0 jour			
à 9 ans 5 mois 30 jours	28	24	52
9 ans 6 mois 0 jour			
à 9 ans 11 mois 30 jours	21	26	47
10 ans 0 mois 0 jour			
à 10 ans 5 mois 30 jours	17	21	38
10 ans 6 mois 0 jour			
à 10 ans 11 mois 30 jours	34	30	64
11 ans 0 mois 0 jour			
à 11 ans 5 mois 30 jours	19	19	38
11 ans 6 mois 0 jour			
à 11 ans 11 mois 30 jours	31	32	63
12 ans 0 mois 0 jour			
à 12 ans 5 mois 30 jours	27	29	56
12 ans 6 mois 0 jour			
à 12 ans 11 mois 30 jours	24	20	44
13 ans 0 mois 0 jour			
à 13 ans 5 mois 30 jours	23	22	45
13 ans 6 mois 0 jour			
à 13 ans 11 mois 30 jours	26	26	52
14 ans 0 mois 0 jour			
à 14 ans 5 mois 30 jours	25	29	54
14 ans 6 mois 0 jour			
à 14 ans 11 mois 30 jours	25	21	46
15 ans 0 mois 0 jour			
à 15 ans 5 mois 30 jours	28	23	51
15 ans 6 mois 0 jour			
à 15 ans 11 mois 30 jours	25	27	52
16 ans 0 mois 0 jour			
à 16 ans 5 mois 30 jours	26	24	50
16 ans 6 mois 0 jour			
à 16 ans 11 mois 30 jours	23	24	47
Total	554	549	1 103

Petite taille des échantillons

Critère Âge

Onze groupes d'âge, de 6 ans 0 mois à 16 ans 11 mois ont été constitués. Chaque groupe d'âge comprend entre 80 et 104 enfants.

WISC-V

Table A.1 Conversion des notes brutes totales aux subtests en notes standard, par groupe d'âge (suite)

Note standard	Âges 6:8-6:11															Note standard
	CUB	SIM	MAT	MCH	COD	VOC	BAL	PUZ	MIM	SYM	INF	SLC	BAR	COM	ARI	
1	0	—	0	0-2	0-5	0-2	0-1	0	0	0-1	0-1	—	0-16	0	0	1
2	1	—	1	3-4	6-8	3	2	—	1-2	2-3	2	—	17-19	—	—	2
3	2	0	2	5	9-10	4	3-4	1-2	3-4	4-5	3	0	20-22	1	1-2	3
4	3-4	1	3	6-7	11-12	5	5-6	3	5-6	6-7	4	1	23-25	2-3	3	4
5	5-6	2-3	4-5	8	13-15	6-7	7	4	7-8	8-10	5	2-3	26-28	4	4	5

4 509 218 enfants de 6 à 16 ans en France: 1103 = 0,024%; 1049=0,023%

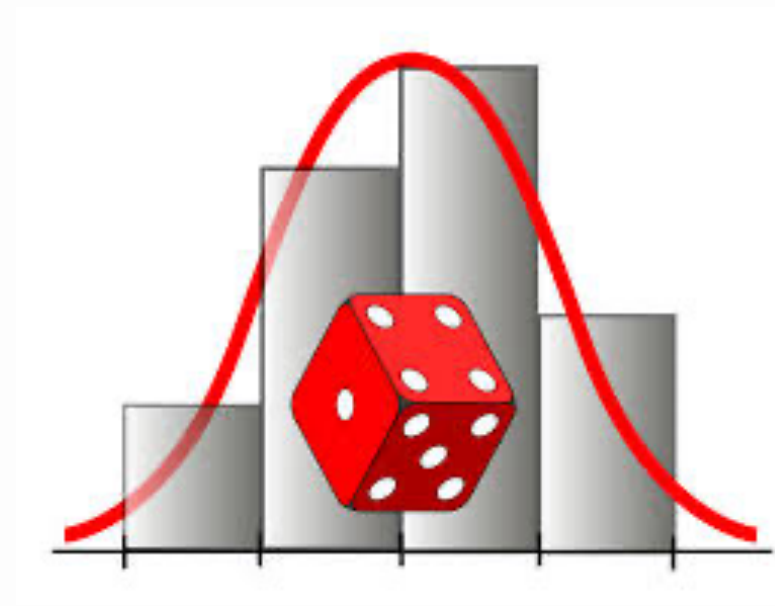
Scores normatifs tous les 4 mois (en réalité les normes sont basées sur des analyses statistiques inférentielles)

WISC-IV

Statistiques : estimation/prédiction sur base de la probabilité



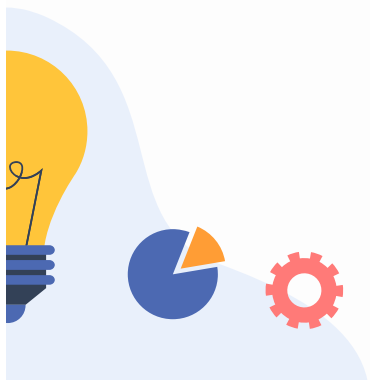
Prendre une décision : test d'hypothèse



A trois jours de la présidentielle américaine, un nouveau sondage sortit samedi soir annonce une victoire de Kamala Harris dans l'état de l'Iowa. Selon les chiffres recueillis auprès de 808 électeurs potentiels du 28 au 30 octobre, Kamala Harris devancerait Donald Trump en obtenant 47 % des intentions de vote contre 44 % pour le républicain. La marge d'erreur du sondage est de 3,4 points de pourcentage.

Si ces nouveaux chiffres surprennent autant c'est parce que l'Iowa est un état à forte tendance républicaine. En effet, il a été remporté par Trump avec plus de 9 points de pourcentage d'avance en 2016 et 8 points en 2020. "Il est difficile pour quiconque de dire qu'il a vu cela venir", a déclaré la sondeuse J. Ann Selzer.

Selon le sondage, ce sont les électrices qui se tourneraient plus vers Kamala Harris dans l'Iowa : "Le sondage montre que ce sont les femmes – en particulier celles qui sont plus âgées ou qui sont politiquement indépendantes – qui sont à l'origine du dernier virage en faveur de Kamala Harris".



Risque d'erreur



Donald Trump
Republican Party

56,3 %

899 659 votes

↑ A remporté 6 votes du Collège électoral



Kamala Harris
Democratic Party

42,3 %

676 339 votes



Robert Kennedy
Indépendant(e)

0,8 %

12 703 votes



Chase Oliver
Libertarian Party

0,4 %

6 883 votes

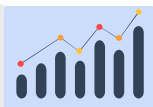


Libération

<https://www.liberation.fr> › International › Amérique

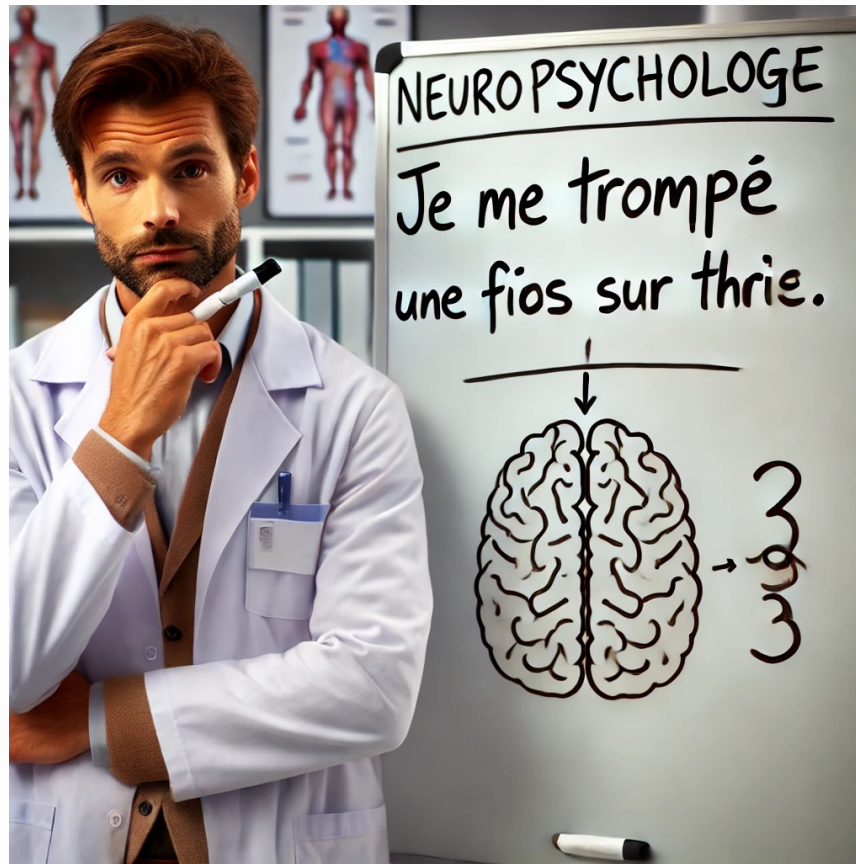
Résultats des élections américaines : le sondage mirage ...

il y a 4 heures — Les **résultats** de l'enquête d'un institut extrêmement réputé, qui donnait le week-end dernier Kamala Harris en tête dans cet Etat solidement ...





But = limiter le risque d'erreur



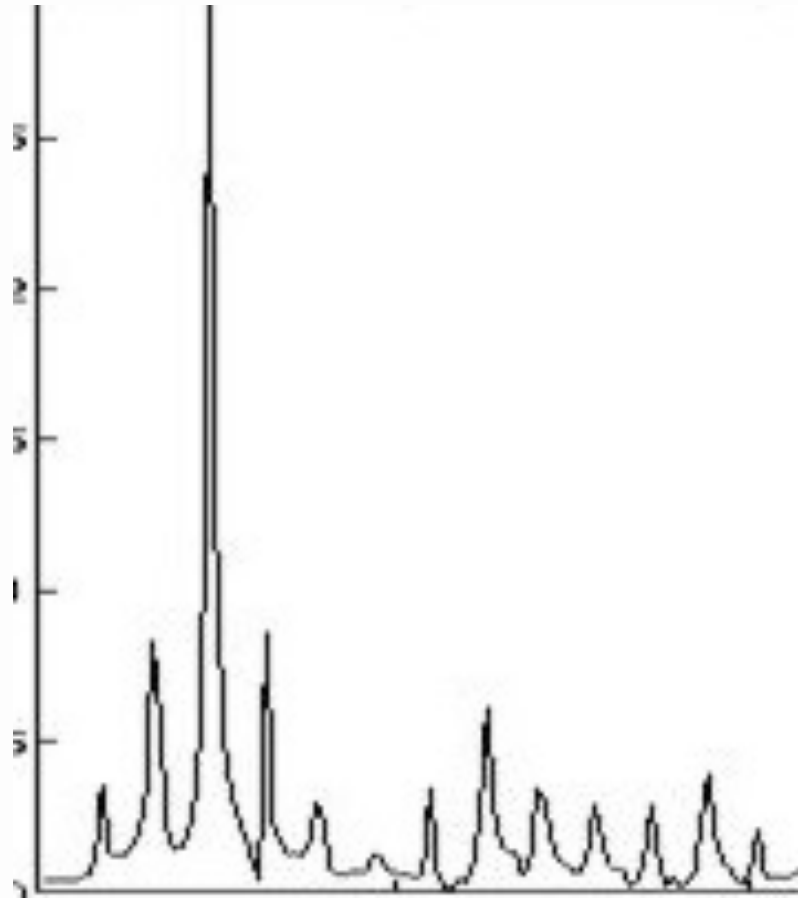
Rem: ChatGPT est dysorthographique



En sciences humaines les valeurs 0,05 et 0,01 sont souvent appelées seuil de rejet ou seuil de signification du test

Howell (2013)

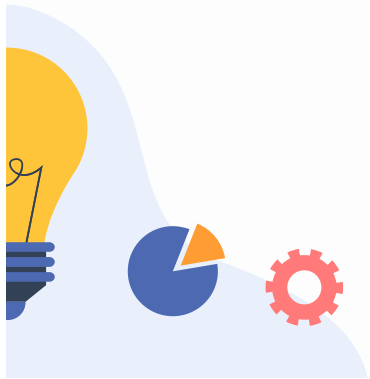
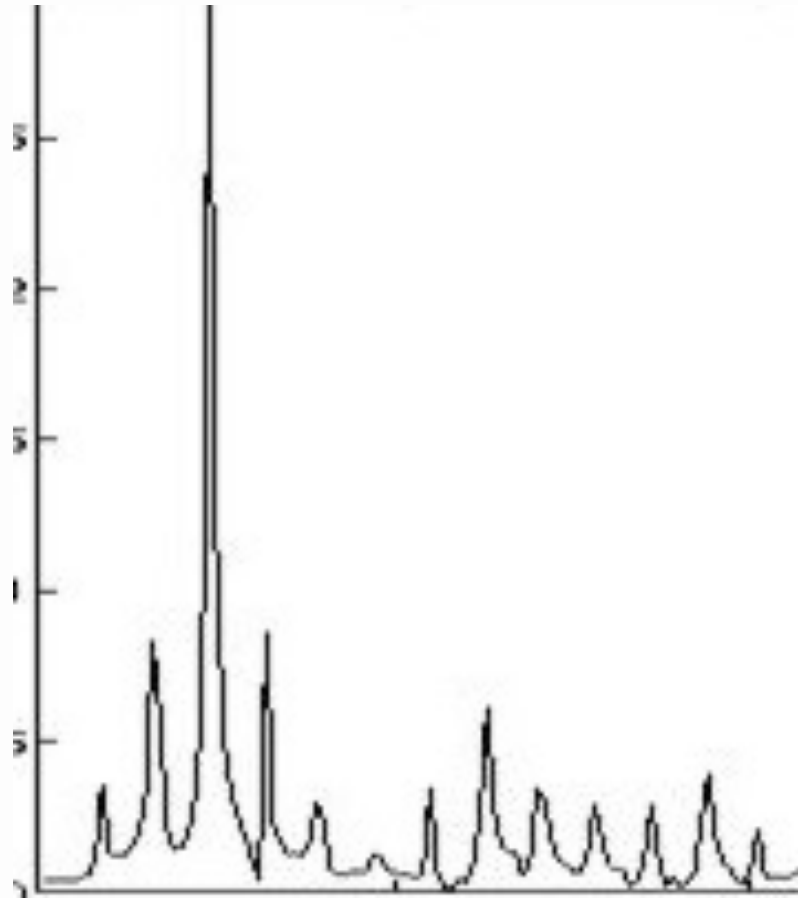
limiter le bruit



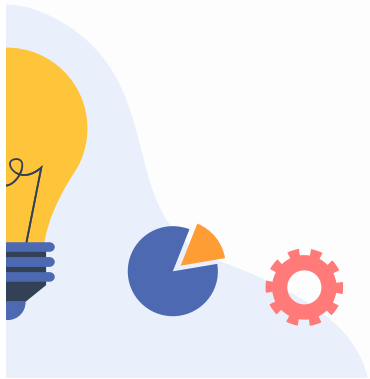
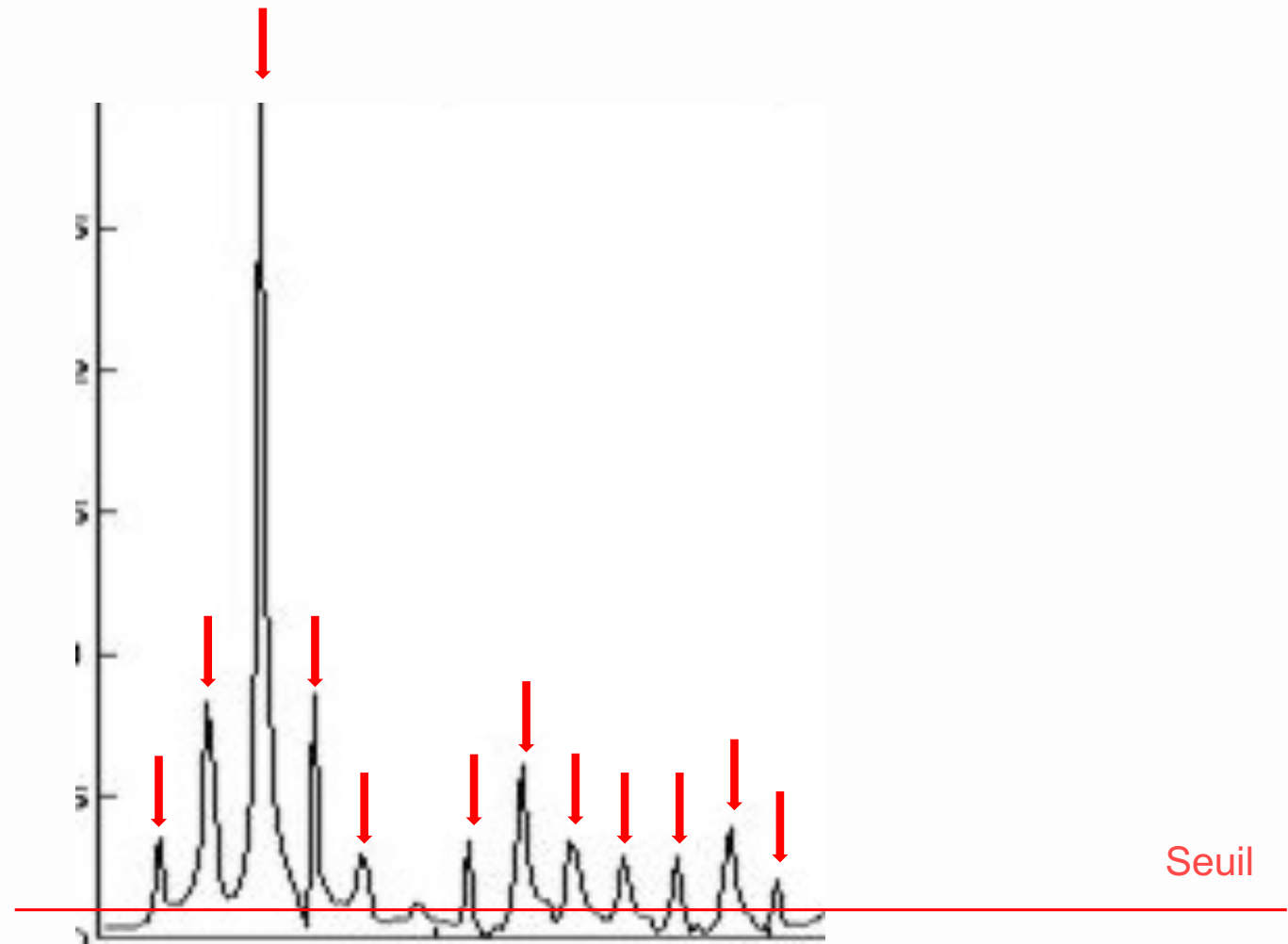
Limiter le bruit



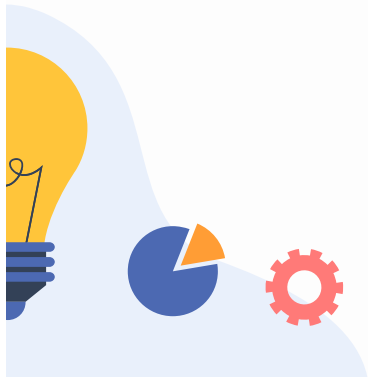
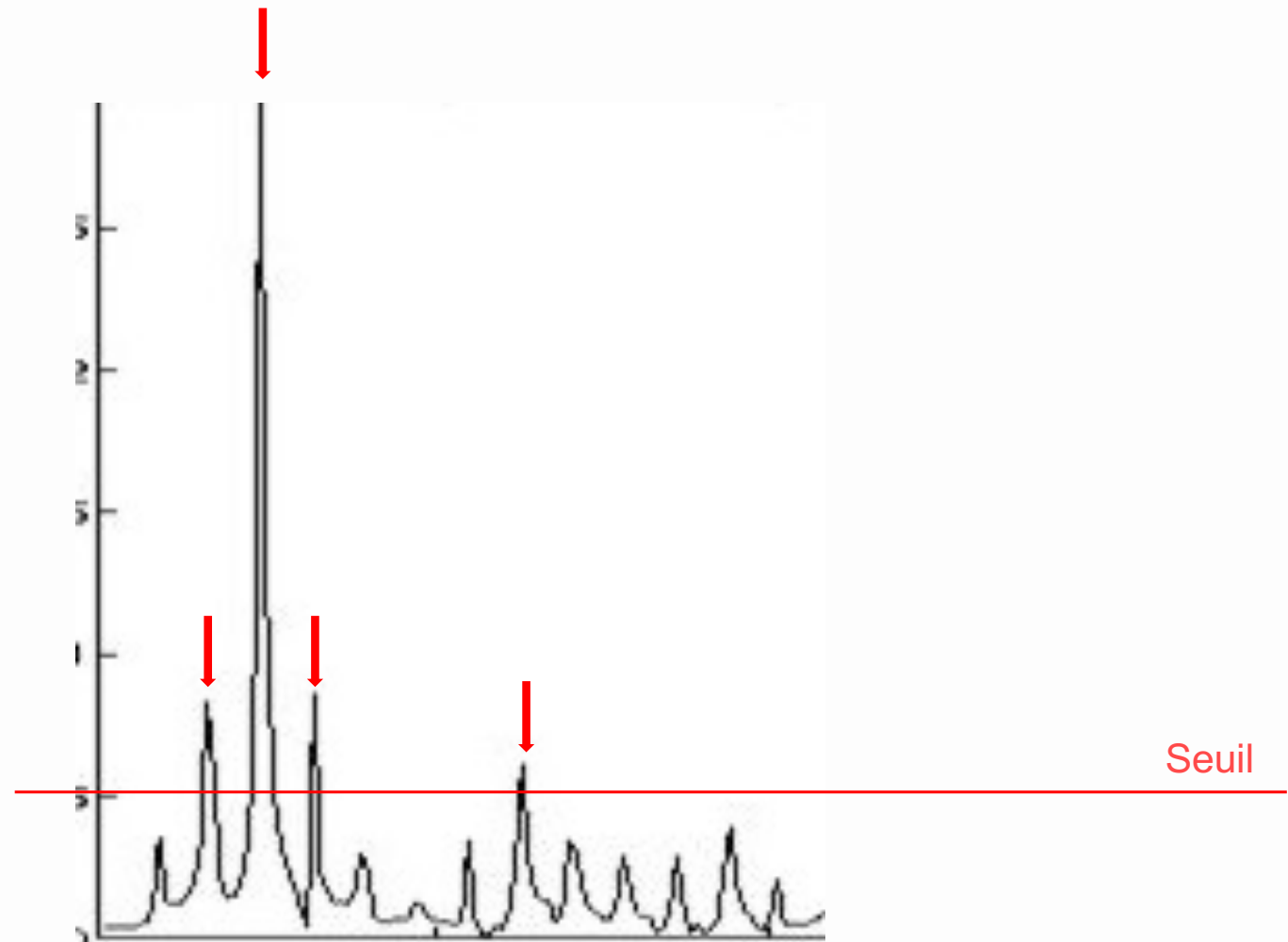
Indicateur d'un problème



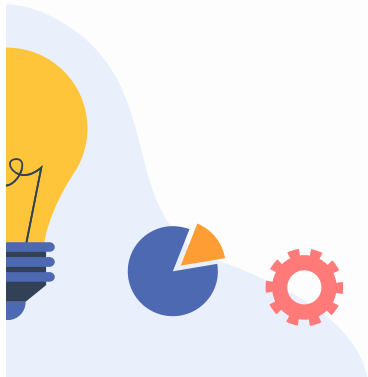
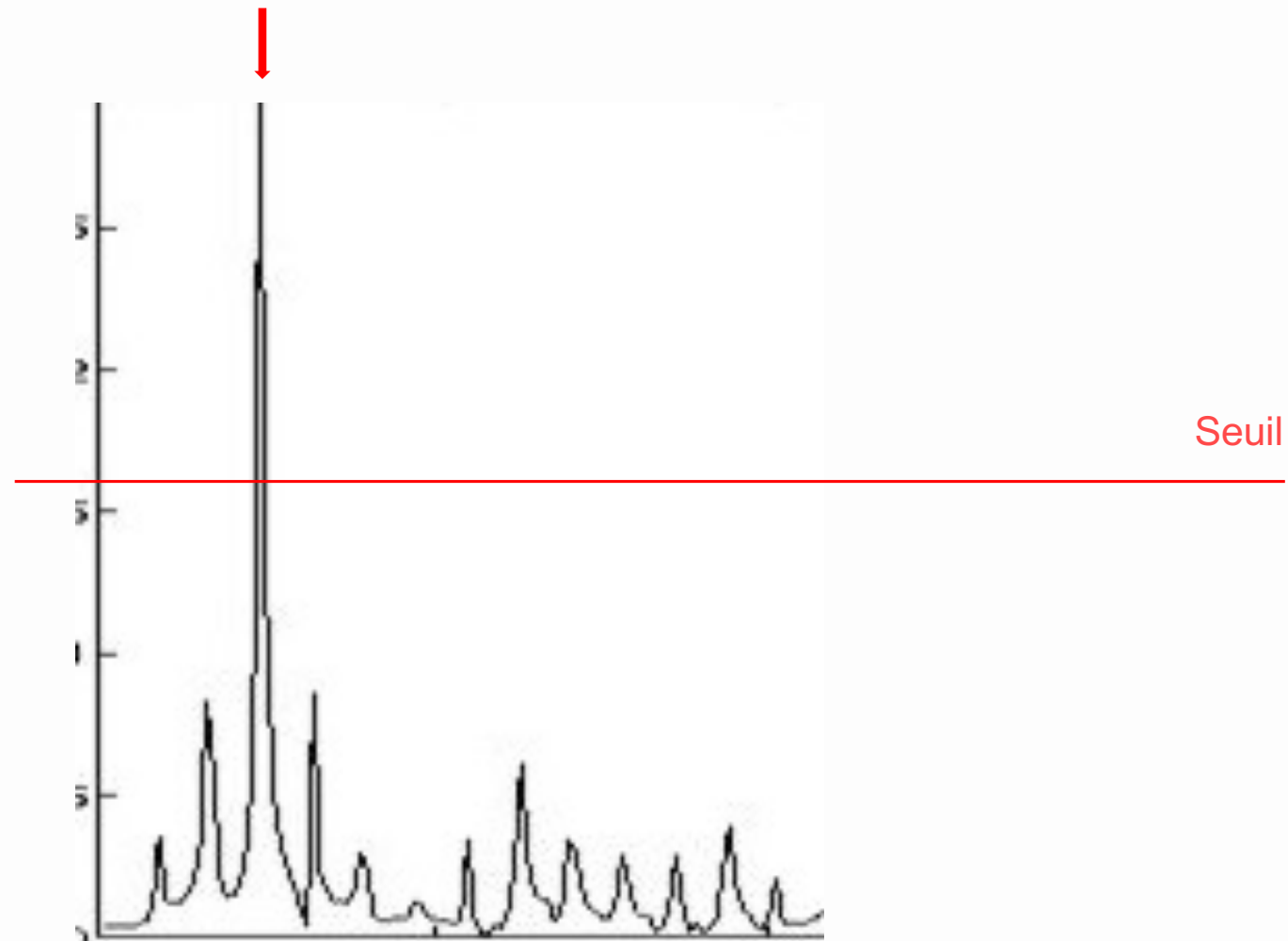
limiter le bruit



limiter le bruit

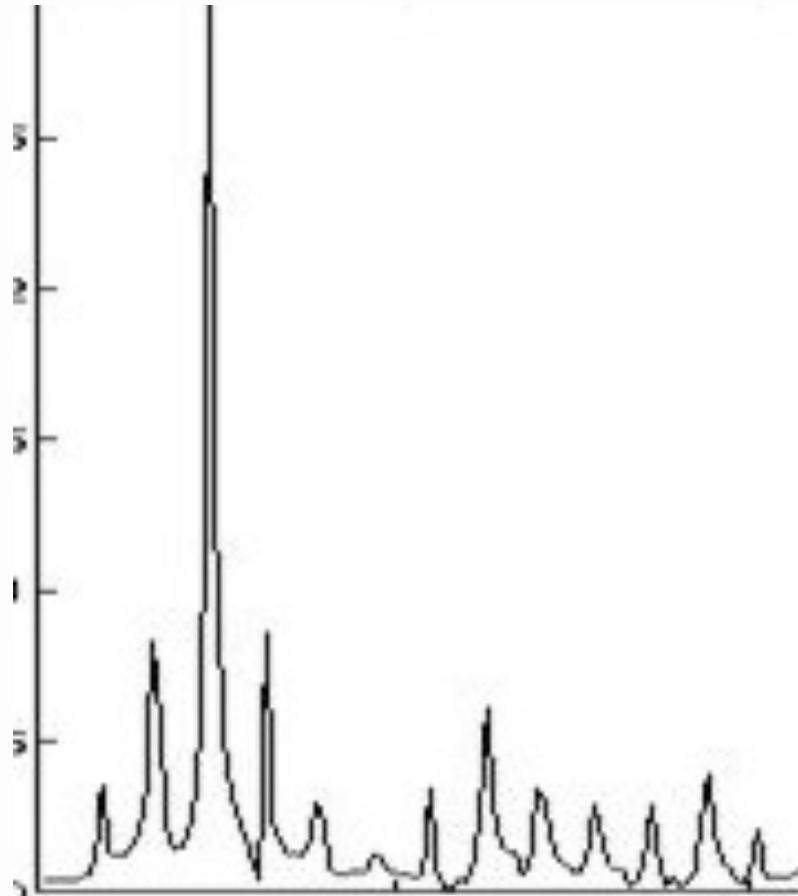


limiter le bruit



Limiter le bruit

Seuil



Hypothèse nulle (H_0)

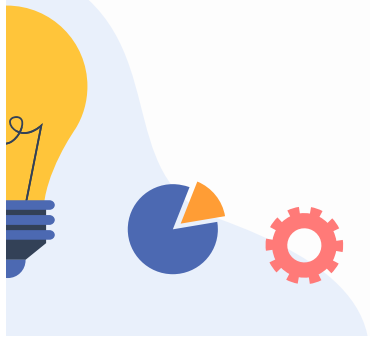
H_0 est l'hypothèse que les données sont extraites de populations avec des moyennes identiques à celle de la population d'échantillonnage

H_0 est donc l'hypothèse d'une égalité entre ce qui est observé et la population (distribution d'échantillonnage)

Ce que l'on teste dans les statistiques classiques, c'est la probabilité de cette égalité et le rejet de cette égalité si la probabilité est trop faible (rejet de H_0 si $p \leq 5\%$)

En fonction de la question posée, les statistiques réalisées peuvent différer.


En neuropsychologie, la demande la plus fréquente est de réaliser un état des lieux des fonctions et processus cognitifs altérés en vue de répondre à une question.





Question habituellement posée aux cliniciens:

Est-ce que la performance observée chez la personne testée est suffisamment rare pour que le/la neuropsychologue puisse prendre le risque d'erreur de rejeter l'hypothèse qu'elle fasse partie des performances de la population de référence ?





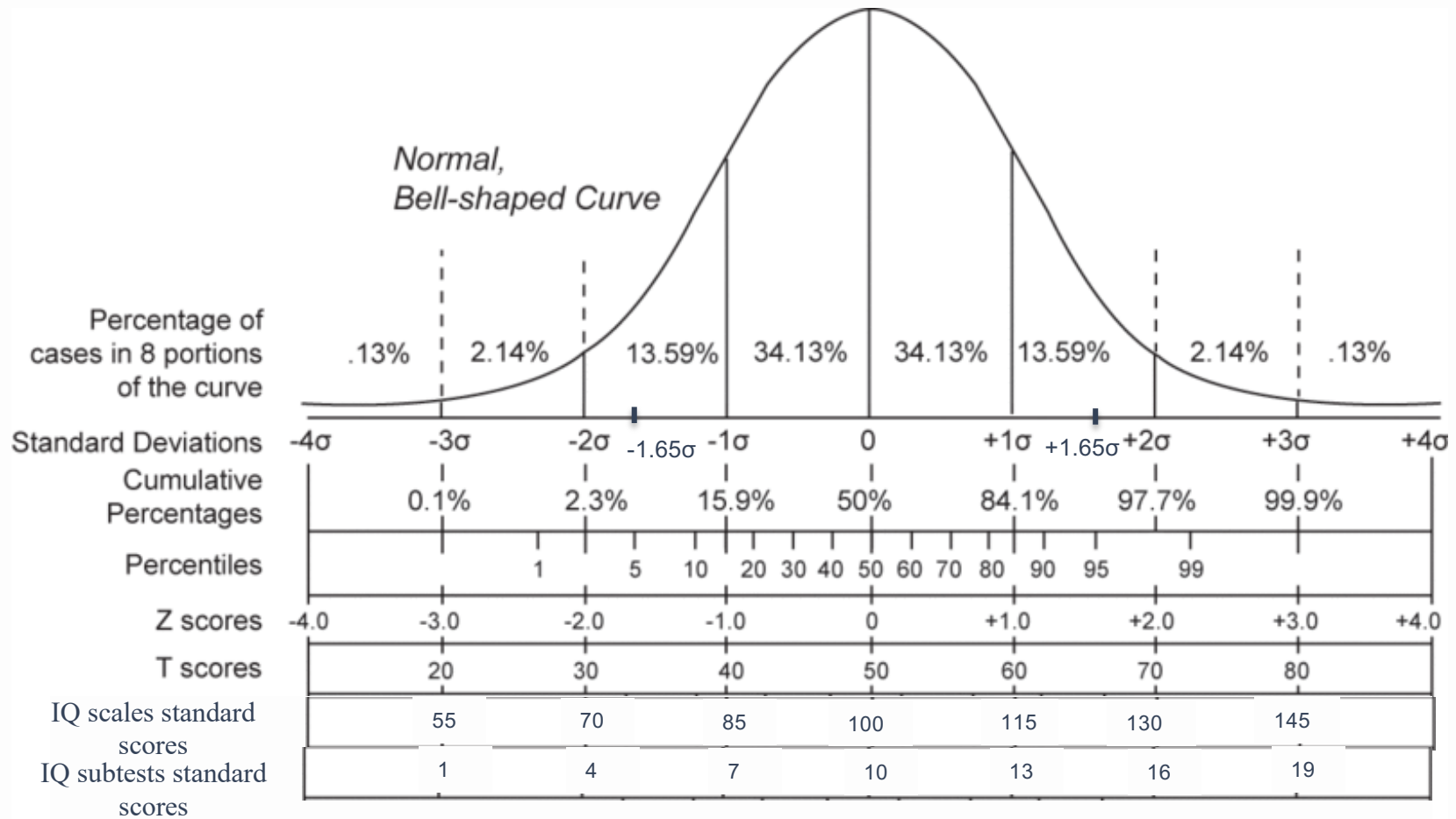
La logique des tests statistiques:

- Nous partons d'une hypothèse (la question posée à la/au neuropsychologue)
- Nous formulons l'hypothèse nulle (H_0 = hypothèse d'égalité)
- Nous traçons la distribution d'échantillonnage de la statistique en supposant que H_0 est vraie (échantillon normatif)
- Nous collectons les données (résultat du sujet/patient)
- Nous comparons les données à la distribution d'échantillonnage
- Nous rejetons ou conservons H_0 en fonction de la probabilité (risque d'erreur) d'arriver à des données comme celles obtenues



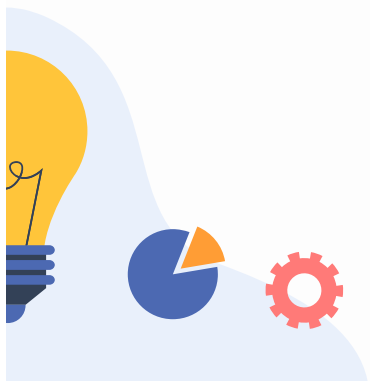
Adapté de Howell (2013)

Seuils de décision, statistiques et conventions



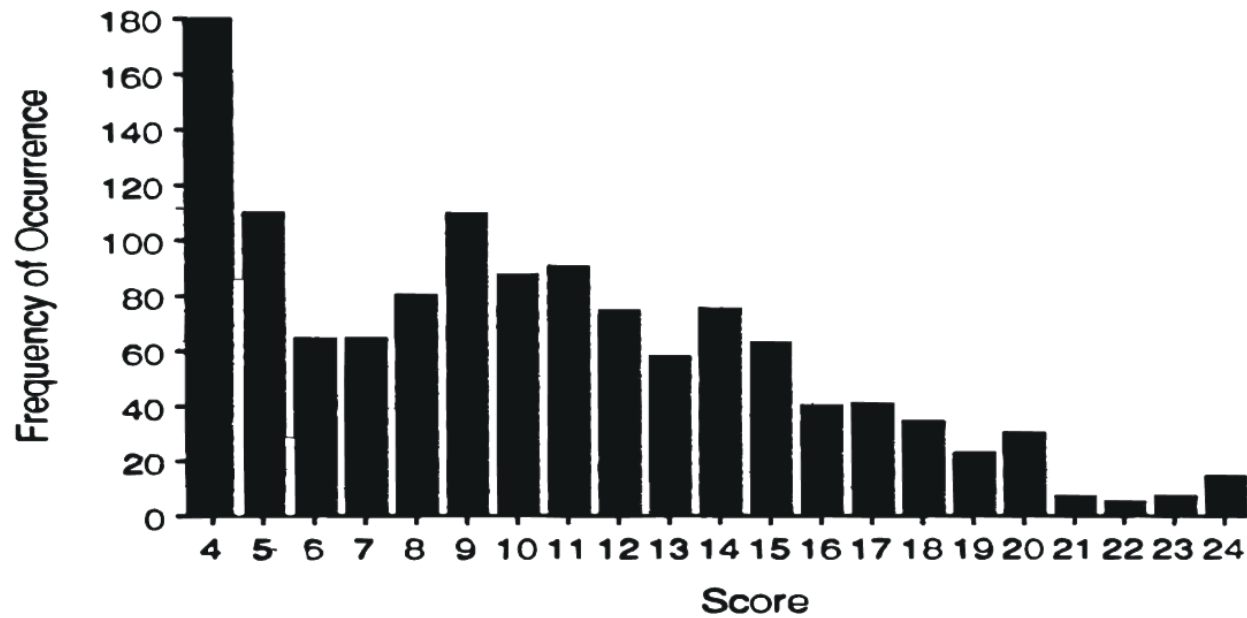
$$\text{Score } z = (\text{score observé} - \text{moyenne de l'échantillon}) / \text{écart-type de l'échantillon}$$

Un score à un test situé à deux écart-type au dessus de la moyenne ne veut pas dire que la personne obtient une performance supérieure à 97,7 % des autres personnes. Ce chiffre indique que seulement 2,3 % de l'échantillon issu de la population de référence obtient ce score ou davantage. J'ai donc une probabilité de 2,3 % de risque d'erreur si je dis que cette personne obtient un score significativement plus élevé que cette population



The Unicorn, The Normal Curve, and Other Improbable Creatures

Theodore Micceri¹
Department of Educational Leadership
University of South Florida



Les statistiques classiques ne testent pas la probabilité de l'hypothèse nulle mais la probabilité d'observer ces résultats si l'hypothèse nulle est vraie

Selon l'argument philosophique utilisé par Fisher nous ne pouvons jamais démontrer l'exactitude d'une hypothèse mais nous pouvons parfois en démontrer l'inexactitude. Observer 3000 personnes ayant deux bras ne prouve pas l'affirmation «Toute personne a deux bras».

Par contre, trouver une personne ne possédant qu'un bras ou n'en possédant aucun suffit pour prouver l'inexactitude de cette énonciation.

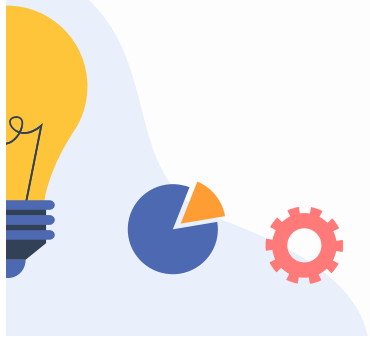
Tous les statisticiens s'accordent sur le fait qu'on ne peut jamais prétendre avoir prouvé l'hypothèse nulle avec les statistiques classiques.


Howell (2013)






Ne pas dire « Je vous aime » ne revient pas à dire « Je ne vous aime pas »





Pour tester la véracité de l'hypothèse nulle, il faut utiliser d'autres principes statistiques basés notamment sur la comparaison entre l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1). C'est le cas, notamment, de l'approche bayésienne.

Cette approche se base sur l'accumulation de données pour faire pencher la balance entre H_0 et H_1 et est donc peu adaptée à la situation de cas unique rencontrée en clinique neuropsychologique



Voir, par exemple, Heck et al. (2023)

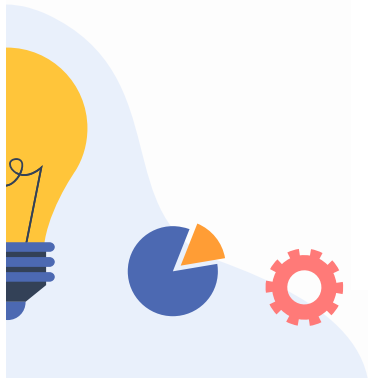
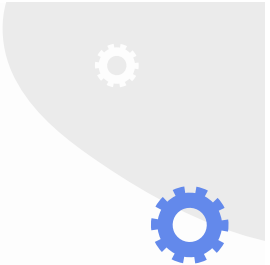
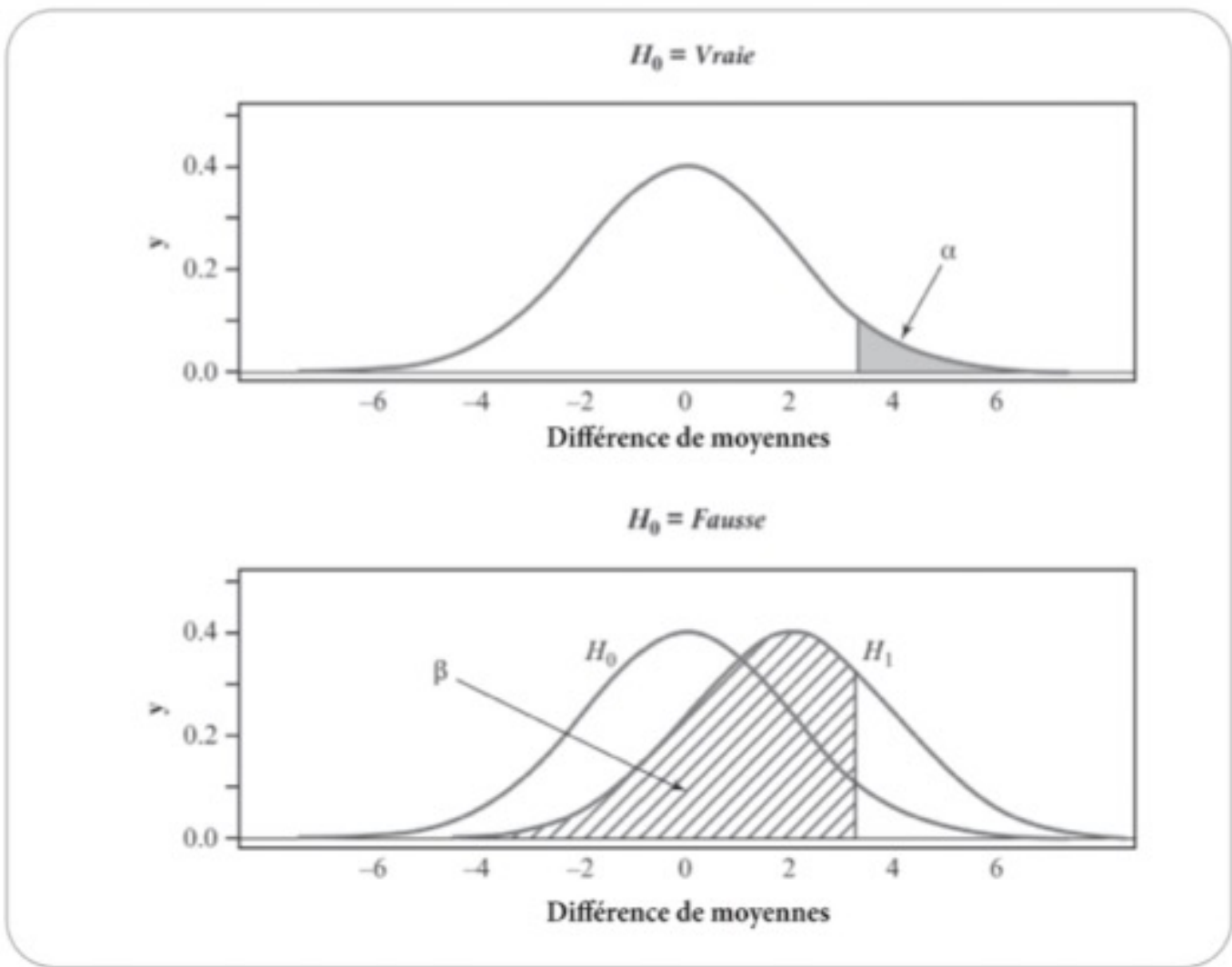


Types de risque d'erreur

Lorsque l'on prend une décision, nous pouvons commettre deux types d'erreur:

- Erreur de type I: risque de rejeter H_0 alors qu'elle est vraie = probabilité α (faux positif)
- Erreur de type II: risque de rejeter H_1 alors qu'elle est vraie = probabilité β (faux négatif) qui est basée sur H_1





Howell (2013; 2024)



Types de risque d'erreur

En général, en neuropsychologie nous ne connaissons pas la distribution de H_1 qui correspond aux résultats de la/des population(s) clinique(s) dont pourrait faire partie le sujet/patient. Nous basons donc notre décision sur H_0 avec comme risque d'erreur fixé le risque α de faux positif

Cependant, comme le risque d'erreur de type I (faux positif) influence le risque d'erreur de type II (faux négatif), mieux vaut éviter de fixer un seuil α trop bas qui augmenterait trop le risque de type II (β)





Test unilatéral ou bilatéral

Lorsque nous décidons de nous intéresser à un seul côté de la courbe de Gauss (par exemple, uniquement les résultats allant dans le sens d'une performance de bas niveau), nous réalisons un test unilatéral. Le risque d'erreur α , pour être de maximum 5%, sera donc placé à $p \leq 0,05$.

Par contre, lorsque nous décidons de nous intéresser aux résultats de haut et de bas niveaux, nous réalisons un test bilatéral. Le risque d'erreur α , pour être de 5%, sera donc placé à $p \leq 0,05/2$ c'est-à-dire à $p \leq 0,025$.

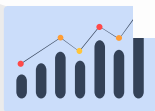
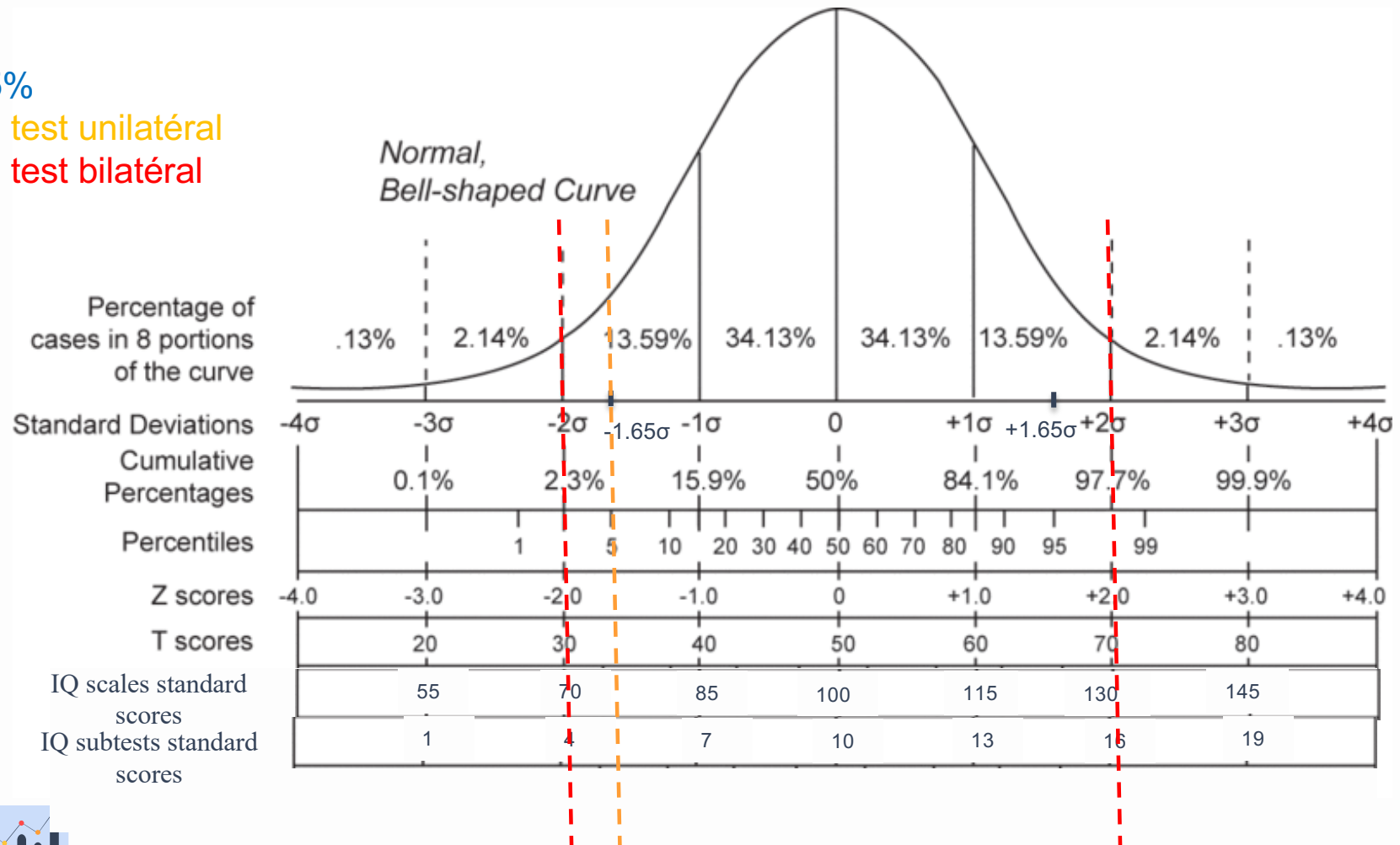


Test unilatéral vs. bilatéral

$\alpha = 5\%$

Seuil test unilatéral

Seuil test bilatéral





Test unilatéral ou bilatéral

Une autre manière de présenter les choses est que, lorsque l'on recherche uniquement une performance basse dans un test, on peut prendre le risque de ne regarder que de ce côté de la courbe de Gauss et donc garder un $p \leq 0,05$.

Par exemple, dans le cas d'une pathologie dégénérative, les neuropsychologues ne s'intéressent, en général, pas aux scores significativement de haut niveau de la personne mais seulement à ceux de bas niveau qui sont les conséquences probables de la pathologie.





Questions ou commentaires ?



05

Quels outils statistiques choisir ? Quel seuil de décision fixer ?

Patrick Fery (Hôpital Universitaire de Bruxelles – H.U.B, Université libre de Bruxelles – ULB; Fonds Erasme pour la recherche médicale)

01

02

03

04

05

06

07

08





Plan

01 Outils et seuils multiples

02 Le score z

03 Le t modifié

04 Le (per)centile

05 Le score seuil (cut-off)

06 Arbre décisionnel



01

Multiplicité d'outils et de seuils



Outils et seuils

Score Z

$\pm 1,645$

$\pm 1,96$

Percentile

$P \leq 5$

$P \leq 2,5$

t modifié

$p \leq 0,05$

$p \leq 0,025$

$p \leq 0,02$

Score seuil

$<$ ou $>$ ou = score fixe

Note standard

Moyenne

\pm

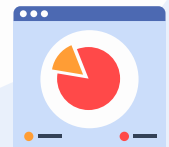
1,645 (1,96) écart-type

Equation de régression

score z:

$\pm 1,645$

$\pm 1,96$





02

Le score z à la loupe

$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

La question de l'effectif



UNILATERAL	BILATERAL
± 1,645	± 1,96

DISTRIBUTION NORMALE

**LE TAUX DE FAUX POSITIFS
EST INVERSEMENT PROPORTIONNEL
A LA TAILLE DU SOUS-GROUPE DE
REFERENCE.**

Table 1

Results From a Monte Carlo Simulation Study of the Percentage of Control Cases Classified as Exhibiting a Deficit (i.e., Percentage of Type I Errors) Using z and a Modified t Test When the Specified Error Rate Is 5%

Control sample N	Percentage of Type I errors
	z
5	10.37
10	7.57
20	6.25
50	5.53
100	5.28



$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

La question de la distribution

ASYMETRIE

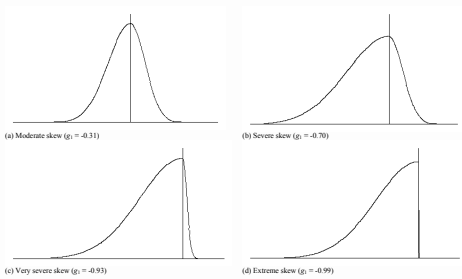


Table 2
Simulation Results: Percentage of Type I Errors (i.e., Percentage of Control Cases Classified as Exhibiting a Deficit) Using z and a Modified t Test for a Specified Error Rate of 5% When Sampling From (Negatively) Skewed Distributions

LE TAUX DE FAUX POSITIFS AUGMENTE AVEC LE DEGRE D'ASYMETRIE

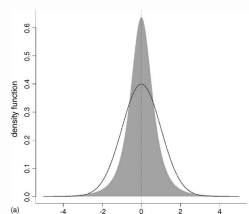
		Skew			
		Moderate ($\gamma_1 = -0.31$)	Severe ($\gamma_1 = -0.70$)	Very severe ($\gamma_1 = -0.93$)	Extreme ($\gamma_1 = -0.99$)
	<i>N</i>	<i>z</i>	<i>z</i>	<i>z</i>	<i>z</i>
Normale	5	11.48	12.50	13.23	13.39
	10	8.59	9.64	10.23	10.23
	20	7.23	8.20	8.72	8.72
	50	6.50	7.37	7.85	7.85
	100	6.20	7.11	7.56	7.56





$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

La question de la distribution



APLATISSEMENT

Table 1

Simulation results: percentage of Type I errors (i.e., percentage of control cases classified as exhibiting a deficit) using z and Crawford and Howell's method for a specified error rate of 5% when sampling from leptokurtic distributions

Control N	Normal distribution	Moderate leptokurtosis	Severe leptokurtosis
	z	z	z
5	10.36	10.42	10.34
10	7.53	7.47	7.42
20	6.26	6.10	5.88
50	5.48	5.20	4.91
100	5.23	4.93	4.53

LE TAUX DE FAUX POSITIFS N'AUGMENTE PAS AVEC LE DEGRE D'APLATISSEMENT



$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

La question de la distribution



ASYMETRIE ET APLATISSEMENT

Table 2
Simulation results: percentage of Type I errors using z and Crawford and Howell's method for a specified error rate of 5%; effects of combinations of skewness and kurtosis

Control N	Moderate skew	Severe skew	Very severe skew	Extreme skew
	z	z	z	z
No leptokurtosis				
5	11.28	12.51	13.24	13.37
10	8.50	9.59	10.22	10.34
20	7.10	8.12	8.74	8.85
50	6.31	7.29	7.79	7.97
100	6.08	7.01	7.56	7.64
Moderate leptokurtosis				
5	12.06	13.11	13.78	13.97
10	8.99	9.98	10.47	10.65
20	7.56	8.34	8.70	8.85
50	6.54	7.24	7.54	7.64
100	6.19	6.86	7.12	7.16
Severe leptokurtosis				
5	12.56	13.59	14.18	14.31
10	9.37	10.24	10.60	10.74
20	7.62	8.32	8.62	8.69
50	6.43	6.95	7.22	7.20
100	5.92	6.38	6.61	6.68

**LE TAUX DE FAUX POSITIFS
AUGMENTE AVEC LE
DEGRE D'ASYMETRIE ET LE
DEGRE D'APLATISSEMENT**



$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

UNILATERAL	BILATERAL
± 1,645	± 1,96

La question de l'effectif
et de la distribution: nuance!

**CELA NE CONCERNE PAS LES
SCORES Z ISSUS DE
L'UTILISATION D'UNE
EQUATION DE REGRESSION ET
DE L'ECART-TYPE RESIDUEL.**

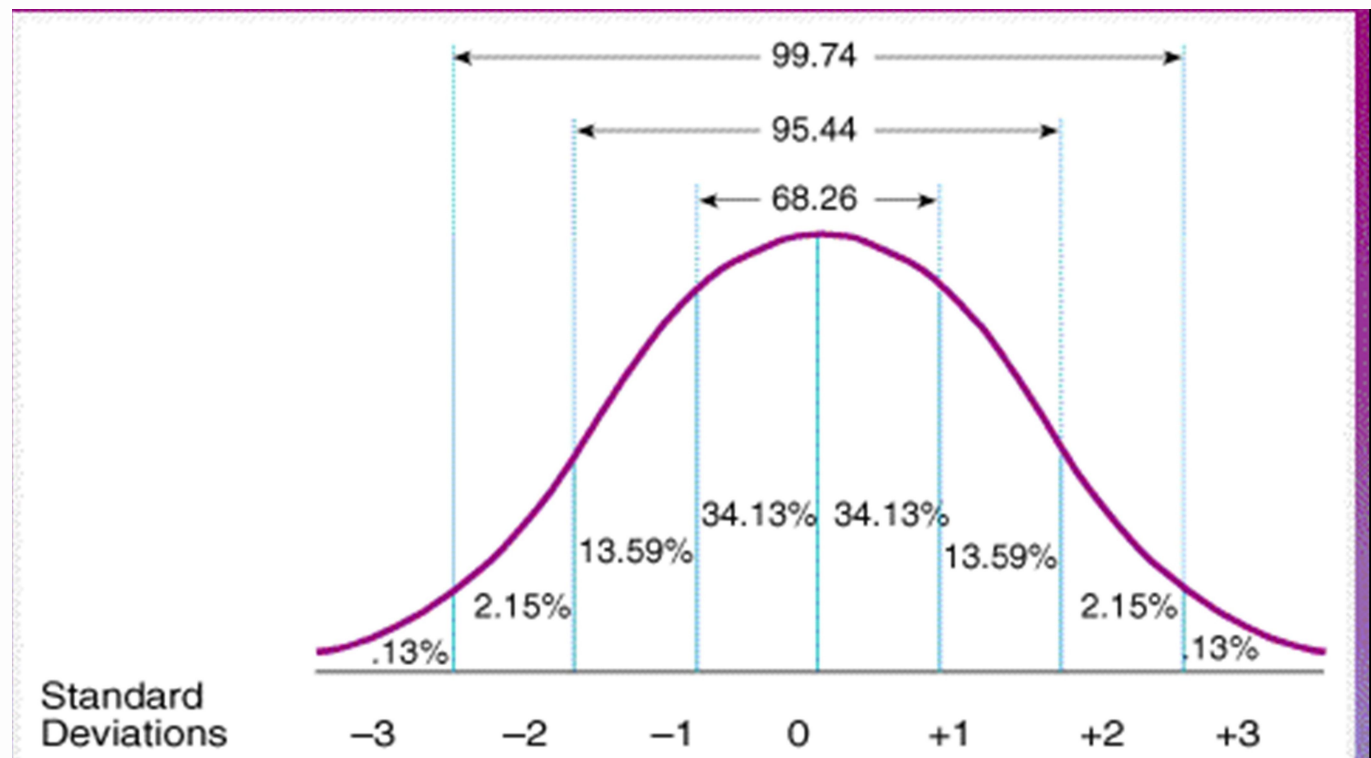


$$z = \frac{\text{score} - \text{moyenne}}{\text{déviation standard}}$$

Aire sous la courbe

exemple:
score $z = -1,87$
3,07%

L'estimation de la fréquence





03

**Une alternative:
le t modifié**

$$t = \frac{\text{score} - \text{moyenne}}{\text{dév. stand.} \sqrt{\frac{N2+1}{N2}}}$$

UNILATERAL	BILATERAL
$p \leq 0,05$	$p \leq 0,05$

DISTRIBUTION NORMALE

LE TAUX DE FAUX POSITIFS EST CONSTANT (5%) QUELLE QUE SOIT LA TAILLE DU SOUS-GROUPE DE REFERENCE.

La question de l'effectif

Table 1
Results From a Monte Carlo Simulation Study of the Percentage of Control Cases Classified as Exhibiting a Deficit (i.e., Percentage of Type I Errors) Using z and a Modified t Test When the Specified Error Rate Is 5%

Control sample N	Percentage of Type I errors
	t
5	5.01
10	5.00
20	5.00
50	5.03
100	4.98

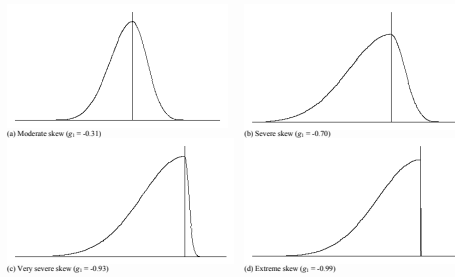
Sokal, R. R., & Rohlf, J. F. (1995). *Biometry*. San Francisco, CA: W.H. Freeman.

$$t = \frac{\text{score} - \text{moyenne}}{\text{dév. stand.} \sqrt{\frac{N2+1}{N2}}}$$

La question de la distribution



ASYMETRIE



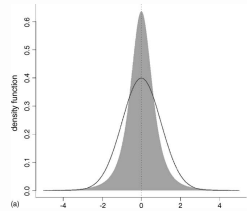
LE TAUX DE FAUX POSITIFS AUGMENTE AVEC LE DEGRE D'ASYMETRIE

$p \leq 0,025$

Table 2
Simulation Results: Percentage of Type I Errors (i.e., Percentage of Control Cases Classified as Exhibiting a Deficit) Using z and a Modified t Test for a Specified Error Rate of 5% When Sampling From (Negatively) Skewed Distributions

N	Skew			
	Moderate ($\gamma_1 = -0.31$)	Severe ($\gamma_1 = -0.70$)	Very severe ($\gamma_1 = -0.93$)	Extreme ($\gamma_1 = -0.99$)
	t	t	t	t
5	6.06	7.23	8.04	8.27
10	6.04	7.14	7.80	7.94
20	5.97	6.97	7.53	7.66
50	6.00	6.90	7.37	7.47
100	5.97	6.87	7.32	7.32

$$t = \frac{\text{score} - \text{moyenne}}{\text{dév. stand.} \sqrt{\frac{N2+1}{N2}}}$$



La question de la distribution



APLATISSEMENT

Table 1

Simulation results: percentage of Type I errors (i.e., percentage of control cases classified as exhibiting a deficit) using z and Crawford and Howell's method for a specified error rate of 5% when sampling from leptokurtic distributions

Control N	Normal distribution		Moderate leptokurtosis		Severe leptokurtosis	
	z	Crawford and Howell	z	Crawford and Howell	z	Crawford and Howell
5	10.36	5.02	10.42	5.42	10.34	5.64
10	7.53	4.98	7.47	5.23	7.42	5.39
20	6.26	5.01	6.10	5.06	5.88	4.97
50	5.48	4.99	5.20	4.80	4.91	4.59
100	5.23	4.98	4.93	4.73	4.53	4.37

**LE TAUX DE FAUX POSITIFS
N'AUGMENTE PAS AVEC LE
DEGRE D'APLATISSEMENT**

$p \leq 0,05$



$$t = \frac{\text{score} - \text{moyenne}}{\text{dév. stand.} \sqrt{\frac{N2+1}{N2}}}$$

La question de la distribution



ASYMETRIE ET APLATISSEMENT

LE TAUX DE FAUX POSITIFS AUGMENTE AVEC LE DEGRE D'ASYMETRIE ET LE DEGRE D'APLATISSEMENT

$p \leq 0,02$

Table 2

Simulation results: percentage of Type I errors using z and Crawford and Howell's method for a specified error rate of 5%; effects of combinations of skewness and kurtosis

Control N	Moderate skew		Severe skew		Very severe skew		Extreme skew	
	z	Crawford and Howell	z	Crawford and Howell	z	Crawford and Howell	z	Crawford and Howell
No leptokurtosis								
5	11.28	5.93	12.51	7.27	13.24	8.07	13.37	8.27
10	8.50	5.99	9.59	7.17	10.22	7.80	10.34	7.93
20	7.10	5.88	8.12	6.95	8.74	7.56	8.85	7.70
50	6.31	5.84	7.29	6.84	7.79	7.33	7.97	7.51
100	6.08	5.85	7.01	6.79	7.56	7.33	7.64	7.42
Moderate leptokurtosis								
5	12.06	7.12	13.11	8.31	13.78	9.09	13.97	9.31
10	8.99	7.42	9.98	7.85	10.47	8.37	10.65	8.56
20	7.56	6.98	8.34	7.35	8.70	7.71	8.85	7.86
50	6.54	6.16	7.24	6.87	7.54	7.17	7.64	7.26
100	6.19	6.01	6.86	6.67	7.12	6.95	7.16	6.98
Severe leptokurtosis								
5	12.56	7.93	13.59	9.10	14.18	9.81	14.31	9.96
10	9.37	7.39	10.24	8.35	10.60	8.71	10.74	8.85
20	7.62	6.74	8.32	7.45	8.62	7.78	8.69	7.84
50	6.43	6.11	6.95	6.64	7.22	6.91	7.20	6.89
100	5.92	5.78	6.38	6.24	6.61	6.46	6.68	6.48



Neuropsychologia 44 (2006) 666–677

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Testing for a deficit in single-case studies:
Effects of departures from normality

John R. Crawford^{a,*}, Paul H. Garthwaite^b, Adelchi Azzalini^c,
David C. Howell^d, Keith R. Laws^e

ADAPTER LES SEUILS DU SCORE Z



ASYMETRIE ET APLATISSEMENT

2%

UNILATERAL	BILATERAL
$\pm 1,645$	$\pm 1,96$



UNILATERAL	BILATERAL
$\pm 2,05$	$\pm 2,33$



$$t = \frac{\text{score} - \text{moyenne}}{\text{dév. stand.} \sqrt{\frac{N2+1}{N2}}}$$

L'estimation de la fréquence



p x 100

exemple:
p = 0,03
3%





04

Le (per)centile

5	42
10	44
25	52
50	56
75	58
90	59
95	60

$$\frac{(n C's: score < score\ cible) + ((n C's: score = score\ cible) \times 0,5)}{N} \times 100$$

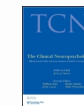


La question de l'effectif

au moins 10 personnes

pas plus de 10% ayant le score de bas niveau le plus bas

UNILATERAL	BILATERAL
P ≤ 5	P ≤ 2,5



The Clinical Neuropsychologist



ISSN: 1365-4648 (Print) 1744-4144 (Online) Journal homepage: <https://www.tandfonline.com/doi/full/10.1080/13654648.2020.1812202>

**On percentile norms in neuropsychology:
Proposed reporting standards and methods for
quantifying the uncertainty over the percentile
ranks of test scores**

Professor John R. Crawford, Paul H. Garthwaite & Daniel J. Slick

5	42
10	44
25	52
50	56
75	58
90	59
95	60

UNILATERAL	BILATERAL
$P \leq 5$	$P \leq 2,5$

**La question de la distribution
ne se pose pas**

**La question de la fréquence
donnée par le percentile**

exemple:

$P = 3$

3%





05

Le score seuil (cut-off)

SCORE \leftrightarrow SCORE SEUIL:

=

<

>

**La question de la distribution
ne se pose pas**

**La question de la fréquence
non estimable**

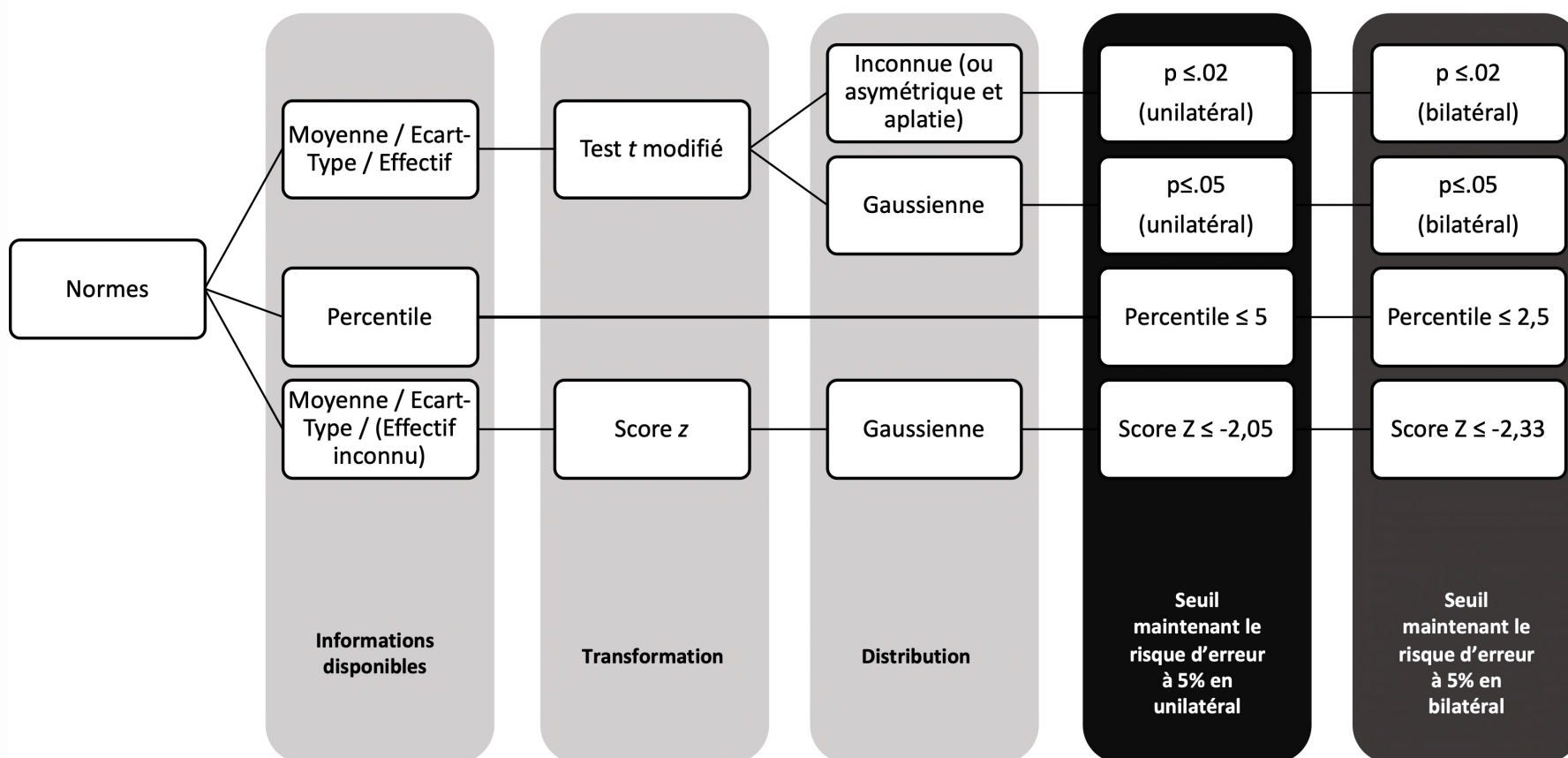
**Sensibilité et spécificité
souvent non disponibles**





06

Arbre décisionnel





Questions ou commentaires ?



06

Au-delà du score seuil : considérations lors de l'interprétation

Béatrice Degraeve, Université Catholique de Lille

01

02

03

04

05

06

07

08



Objectifs de la présentation



Multiplicité des scores

Nombre de tests,
nombre de scores :
quelles implications ?



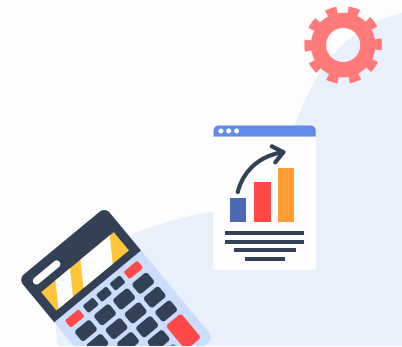
Les hypothèses cliniques

Hypothèse préalable :
Quelles répercussions ?



Conclusion et recommandations

Et alors qu'est-ce
qu'on fait ?





Objectifs de la présentation



Multiplicité des scores

Nombre de tests,
nombre de scores :
quelles implications ?



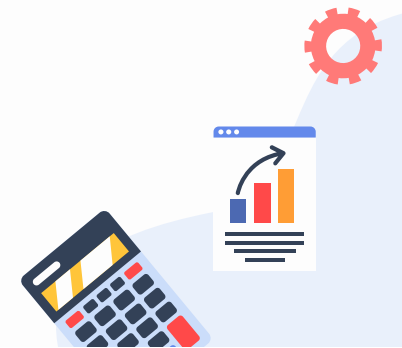
Les hypothèses cliniques

Hypothèse préalable :
Quelles répercussions ?



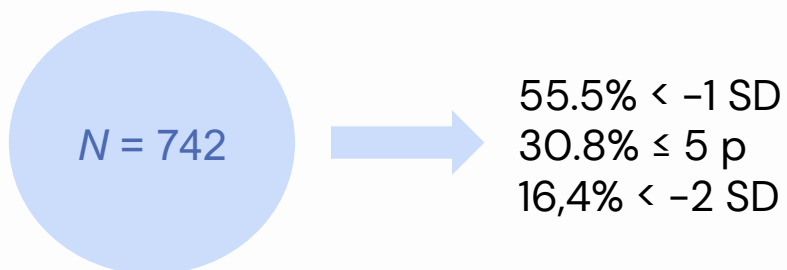
Conclusion et recommandations

Et alors qu'est-ce
qu'on fait ?



- Brooks et al. 2007, 2008

Fréquence des scores peu fréquents de bas niveau chez des adultes âgés sains (obtenus à la *Neuro-psychological Assessment Battery*, NAB ; **10 scores**)



Grand nombre de scores, ↗ risque de trouver des résultats « hors norme » sans qu'il y ait nécessairement de déficits cognitifs réels chez le patient

Journal of the International Neuropsychological Society (2007), **13**, 490–500.
Copyright © 2007 INS. Published by Cambridge University Press. Printed in the USA.
DOI: 10.1017/S1355617707070531

Substantial risk of “Accidental MCI” in healthy older adults: Base rates of low memory scores in neuropsychological assessment

BRIAN L. BROOKS,¹ GRANT L. IVERSON,^{2,1} AND TRAVIS WHITE³

¹British Columbia Mental Health & Addiction Services, Riverview Hospital, Coquitlam, British Columbia

²Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia

³Psychological Assessment Resources Inc., Lutz, Florida



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Archives of Clinical Neuropsychology 25 (2010) 14–21

Archives
of
CLINICAL
NEUROPSYCHOLOGY

Comparing Actual to Estimated Base Rates of “Abnormal” Scores on Neuropsychological Test Batteries: Implications for Interpretation

Brian L. Brooks^{a,b,*}, Grant L. Iverson^{c,d}

Synthèse Non Exhaustive des Études sur la Fréquence des Scores Peu Fréquents de Bas Niveau dans les Batteries Neuropsychologiques



	Batterie	Nombre de scores	N	>1.5 SD	> 2 SD
Schretlen et al. (2008)	ABC study	10	220–327	15% > 1	3.4% > 1
Testa and Schretlen (2006)	ABC study	25	269	16% > 2	3% > 2
Schretlen et al. (2008)	ABC study	25	220–327	40.1% > 1	14.4% > 1
Kiselica et al. (2020)	UDS3NB	21	2,608	42.1% > 0 20.2% > 1	23.9% > 0 9.4% > 1
Schretlen et al. (2008)	ABC study	43	220–327	56.8% > 1	23.8% > 1
Iverson et al. (2008)	WAIS-III/WMS-III	20	1250	43.3% > 0 28% > 1 17.7% > 2	26.6% > 0 14% > 1 5.7% > 2
Heaton et al. (2004)	E-HRNB	25	1189	59% > 0 24% > 1 15% > 2	28% > 0 10% > 1 5% > 2
Heaton et al. (1991)	E-HRNB	40	455	64% > 0	32% > 0
Iverson et al. (2008)	NAB	36	1269	70.2% > 0 48.5% > 1 15.6% > 4	44.3% > 0 21.8% > 1 12.2% > 2

Note. **ABC study** = Aging, Brain Imaging, and Cognition ; **E-HRNB** = Expanded Halstead–Reitan Neuropsychological Battery ; **UDS3NB** = Uniform Data Set Neuropsychological Battery ; **WAIS-III/WMS-III** = Wechsler Adult Intelligence Scale/Wechsler Memory Scale.



Synthèse Non Exhaustive des Études sur la Fréquence des Scores Peu Fréquents de Bas Niveau dans les Batteries Neuropsychologiques



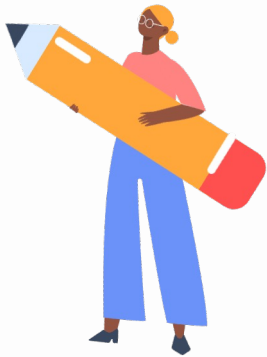
	Batterie	Nombre de scores	N	>1.5 SD	> 2 SD
Schretlen et al. (2008)	ABC study	10	220–327	15% > 1	3.4% > 1
Testa and Schretlen (2006)	ABC study	25	269	16% > 2	3% > 2
Schretlen et al. (2008)	ABC study	25	220–327	40.1% > 1	14.4% > 1
Kiselica et al. (2020)	UDS3NB	21	2,608	42.1% > 0 20.2% > 1	23.9% > 0 9.4% > 1
Schretlen et al. (2008)	ABC study	43	220–327	56.8% > 1	23.8% > 1
Iverson et al. (2008)	WAIS-III/WMS-III	20	1250	43.3% > 0 28% > 1 17.7% > 2	26.6% > 0 14% > 1 5.7% > 2
Heaton et al. (2004)	E-HRNB	25	1189	59% > 0 24% > 1 15% > 2	28% > 0 10% > 1 5% > 2
Heaton et al. (1991)	E-HRNB	40	455	64% > 0	32% > 0
Iverson et al. (2008)	NAB	36	1269	70.2% > 0 48.5% > 1 15.6% > 4	44.3% > 0 21.8% > 1 12.2% > 2

Note. **ABC study** = Aging, Brain Imaging, and Cognition ; **E-HRNB** = Expanded Halstead–Reitan Neuropsychological Battery ; **UDS3NB** = Uniform Data Set Neuropsychological Battery ; **WAIS-III/WMS-III** = Wechsler Adult Intelligence Scale/Wechsler Memory Scale.



Godefroy et al. (2014) : 724 témoins sains / 461 patients diagnostiqués avec troubles cognitifs.

Compromis entre sensibilité et spécificité en fonction du nombre de scores utilisés



Sensibilité : la capacité d'un test à identifier correctement les individus ayant la condition recherchée (vrais positifs). Une haute sensibilité signifie peu de faux négatifs.

Spécificité : la capacité d'un test à identifier correctement les individus sans la condition (vrais négatifs). Une haute spécificité signifie peu de faux positifs.

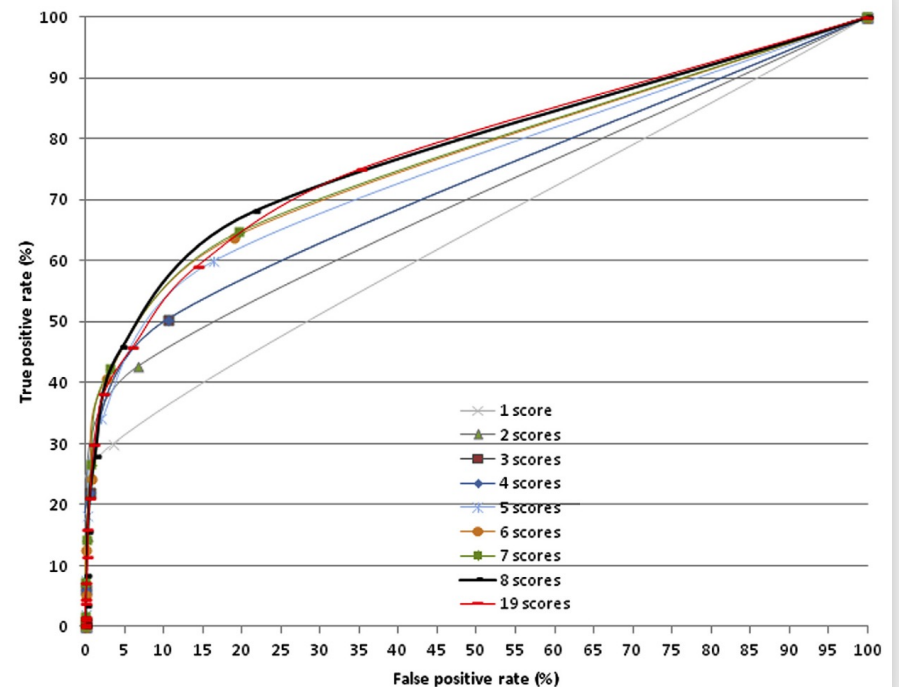
Chaque score supplémentaire ajouté à l'analyse augmente le taux de faux positifs de **2,61 %**

Research report

Validation of an integrated method for determining cognitive ability: Implications for routine assessments and clinical trials



Olivier Godefroy^{a,*}, Laura Gibbons^b, Momar Diouf^c, David Nyenhuis^d, Martine Roussel^a, Sandra Black^e, Jean Marc Bugnicourt^a and GREFEX study group¹



Courbes ROC (Receiver Operating Characteristic) en fonction du nombre de scores utilisés pour distinguer les patients des témoins.

Oltra-Cucarella et al. (2021)

- Lien entre le nombre de mesures administrées et la probabilité d'obtenir des scores peu fréquents de bas niveau.
- Plus les mesures neuropsychologiques sont corrélées, plus la probabilité d'obtenir un ou plusieurs scores faibles augmente



The Association Between the Number of Neuropsychological Measures and the Base Rate of Low Scores

Javier Oltra-Cucarella¹, Miriam Sánchez-SanSegundo¹, María Rubio-Aparicio¹, Juan Carlos Arango-Lasprilla^{2 3 4}, Rosario Ferrer-Cascales¹

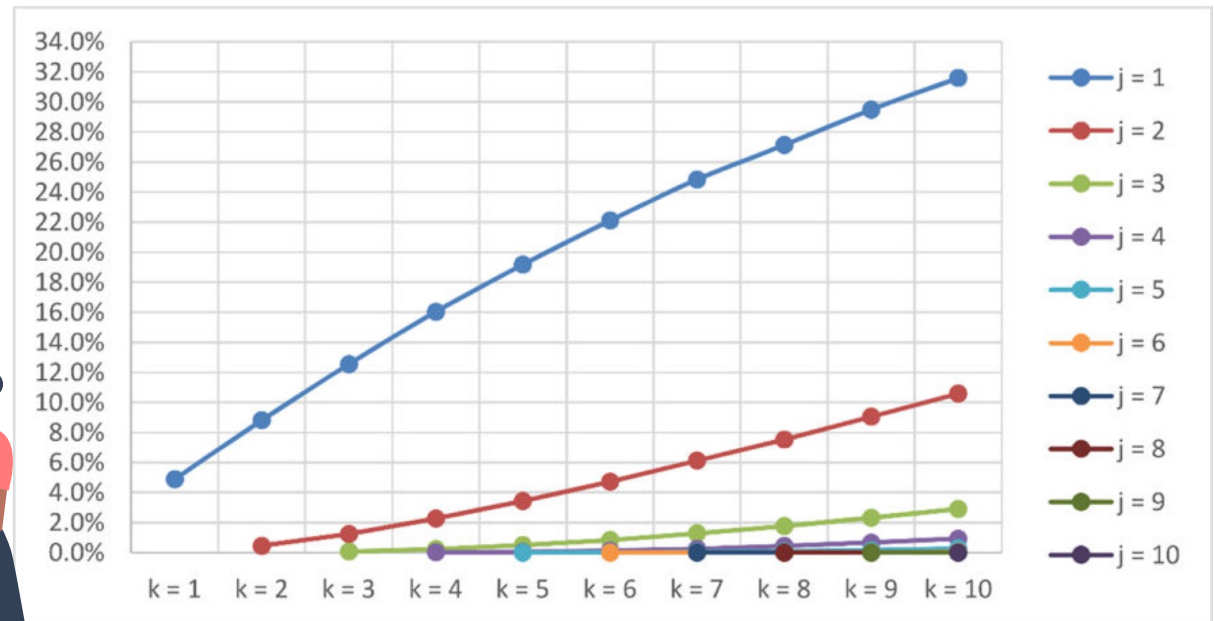


Fig. 1. Number of low scores according to the number of measures. k = number of measures. j = number of low scores or more

Constat : Il est courant d'obtenir quelques scores peu fréquents de bas niveau, même chez des personnes saines.

Et alors qu'est-ce qu'on fait ?

Pratique courantes des cliniciens :

- Agrégation des scores pour une vision d'ensemble.
- Taux de base pour considérer la variabilité cognitive normale.

Notre focus : le seuil de décision

Moduler le seuil de score en fonction du contexte

Utilisation d'un seuil plus conservateur avec de multiples tests.



Binder et al., 2009 ; Brooks et al. 2007, 2008 ; Godefroy et al., 2014 ; Iverson et al., 2008 ; Jak et al., 2009 ; Larrabee, 2012 ; Oltra-Cucarella, et al., 2018



Objectifs de la présentation



Multiplicité des scores

Nombre de tests,
nombre de scores :
quelles implications ?



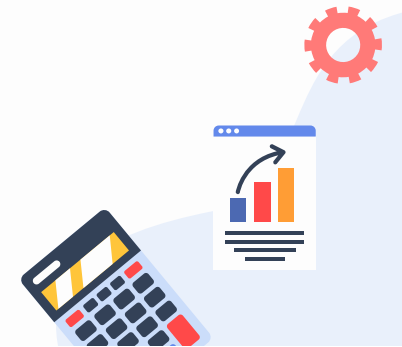
Les hypothèses cliniques

Hypothèse préalable :
Quelles répercussions ?



Conclusion et recommandations

Et alors qu'est-ce
qu'on fait ?



Avez-vous des hypothèses cliniques a priori (basées sur des connaissances diagnostiques spécifiques ou des observations préalables) ?



Oui → *Seuil libéral envisageable (risque 5%)*

Pourquoi ?

Les scores peu fréquents de bas niveau peuvent être attendus et cohérents avec le tableau clinique (ex : déficits de mémoire pour la maladie d'Alzheimer).



Non → *Seuil conservateur (risque 2%)*

Pourquoi ?

Sans prédictions précises, chaque score peu fréquents de bas niveau est pris avec prudence pour éviter des faux positifs.





Objectifs de la présentation



Multiplicité des scores

Nombre de tests,
nombre de scores :
quelles implications ?



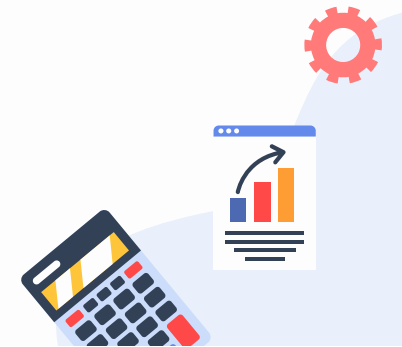
Les hypothèses cliniques

Hypothèse préalable :
Quelles répercussions ?

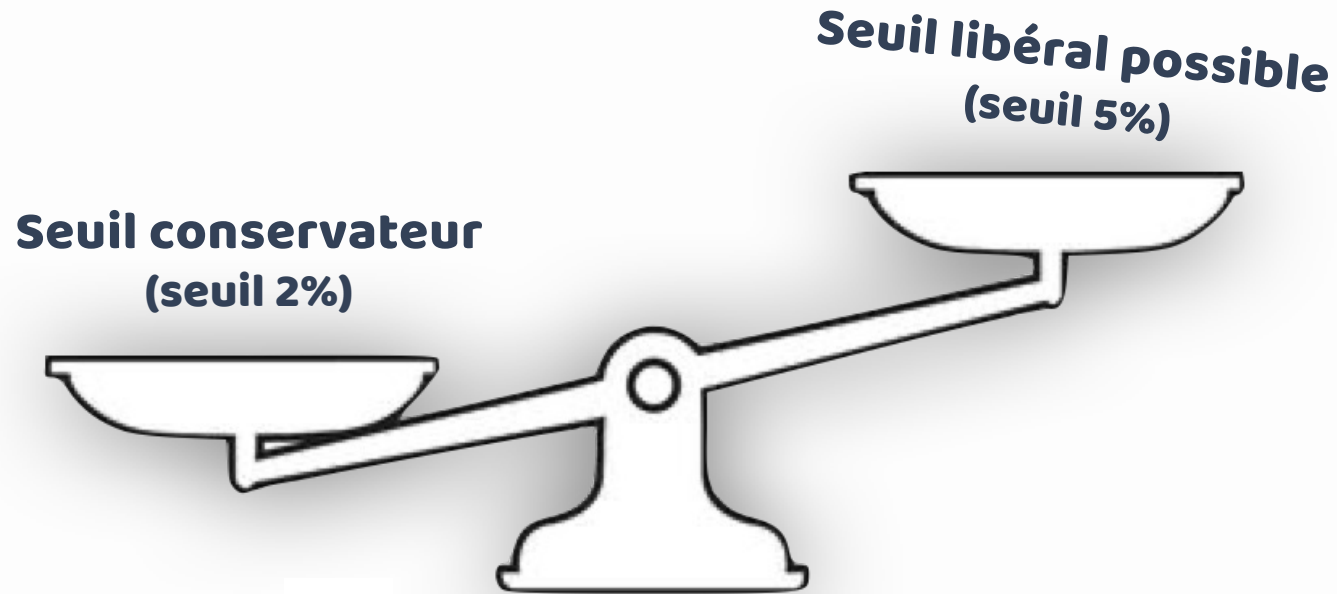


Conclusion et recommandations

Et alors qu'est-ce
qu'on fait ?



Au final qu'est ce qu'on fait ?



Nombre élevé de scores administrés



Pas d'hypothèses spécifiques



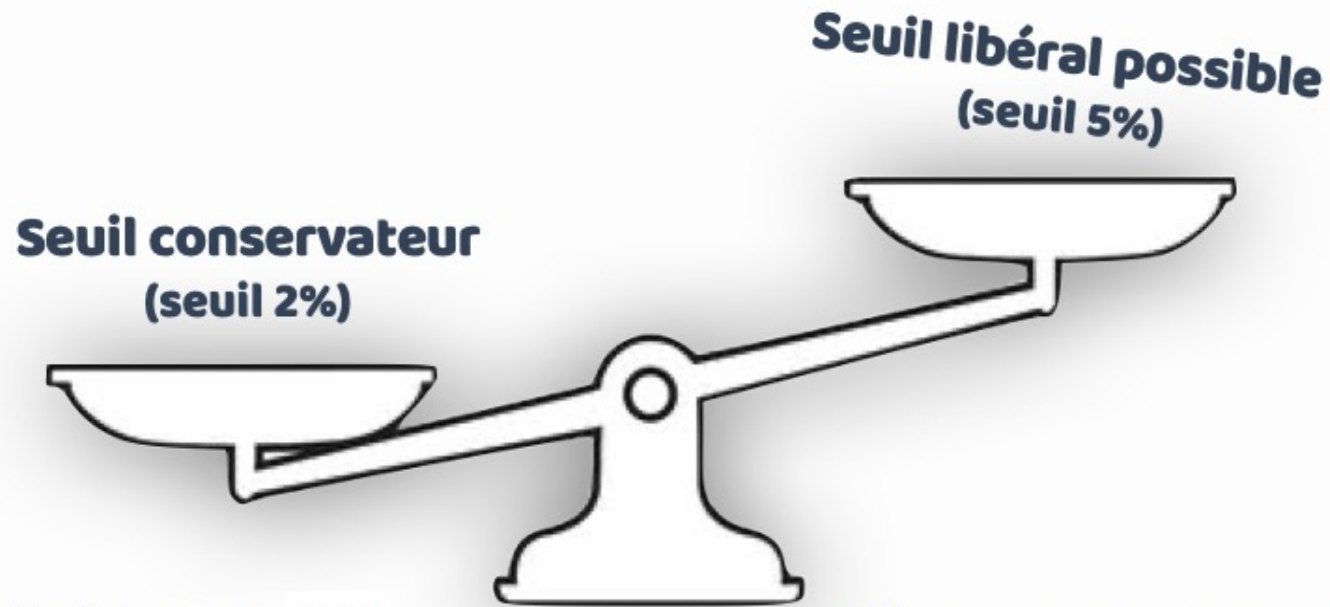
Nombre faible de scores administrés



Hypothèses spécifiques



Au final qu'est ce qu'on fait ?



Nombre élevé de scores administrés



Pas d'hypothèses spécifiques



Nombre faible de scores administrés



Hypothèses spécifiques

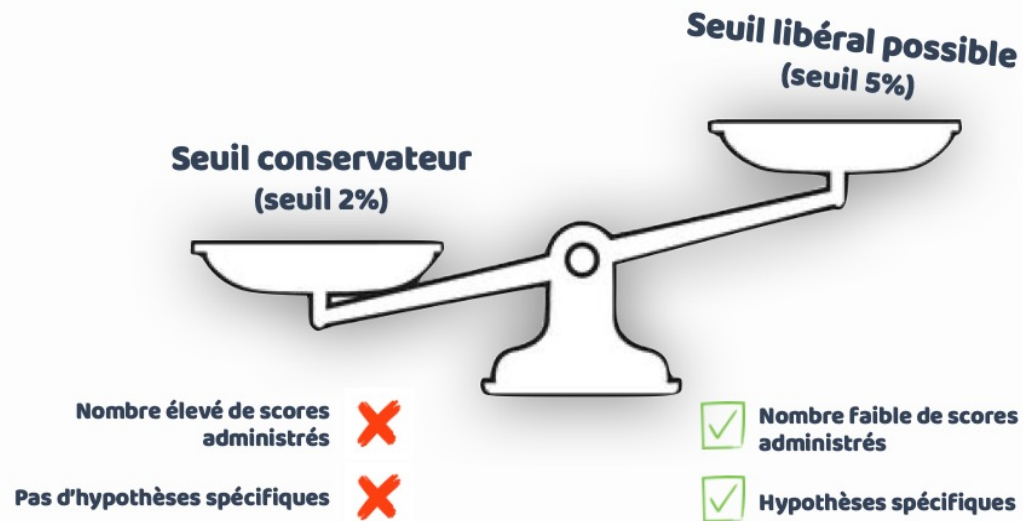


Au final qu'est ce qu'on fait ?



Décision Clinique

Décision Statistique

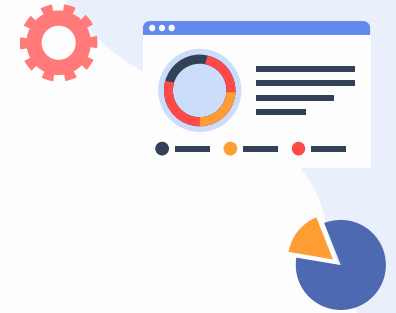


Interprétation et décision

En lien avec : anamnèse, antécédents, plaintes, observations...

Limites inhérentes aux tests





Questions ou commentaires ?



07

Des descripteurs sans ambiguïtés

Sylvie Willems et Patrick Fery



01

02

03

04

05

06

07

08



Des rapports plus pédagogiques

Idéalement, nos rapports devraient être lisibles

- par les patients
- par des professionnels variés

Mais, les équipes interdisciplinaires estiment nos rapports difficilement lisibles, bien que précieux pour la planification des soins (p. ex., Postal et al., 2018).





Des rapports plus pédagogiques

Augmenter la lisibilité des résultats

- Avec plus de transparence ;
- Avec plus de cohérence au sein du rapport ;
- Avec plus de cohérence entre les différents rapports concernant un même patient.



(Brooks et al., 2011 ; Guilmette et al., 2020)



Des descripteurs sans ambiguïtés

- Choisir des descripteurs qui concernent les scores et leur comparaison aux normes et éviter des descripteurs concernant l'état postulé de la fonction cognitive (AACN ; Guilmette et al., 2020)
- Choisir des descripteurs qui reflètent la logique de la comparaison normative (jugement de probabilité d'obtenir le score)
AACN : « score exceptionnellement bas »





Table 1. Recommended test score labels based on standard scores and percentiles for tests with normal distributions.

Standard Score	Percentile	Score Label
≥ 130	≥ 98	Exceptionally high score
120–129	91–97	Above average score
110–119	75–90	High average score
90–109	25–74	Average score
80–89	9–24	Low average score
70–79	2–8	Below average score
< 70	< 2	Exceptionally low score

Source de confusion



Proposition de qualificatifs



Valeur p associée au test t modifié, percentile et probabilité associée au score *standardisé*.



Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description	Seuil
$\geq 98\%$	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	Score fréquent	Score à interpréter ultérieurement en fonction du niveau de performance présumée
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau	Score jugé hors norme avec un seuil libéral
$\leq 2\%$	Score très peu fréquent de bas niveau	Score jugé hors norme avec un seuil conservateur



Proposition de qualificatifs

- associés à 5 catégories de scores

Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description	Seuil
$\geq 98\%$	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	Score fréquent	Score à interpréter ultérieurement en fonction du niveau de performance présumée
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau	Score jugé hors norme avec un seuil libéral
$\leq 2\%$	Score très peu fréquent de bas niveau	Score jugé hors norme avec un seuil conservateur



Proposition de qualificatifs

- associés à 5 catégories de scores

Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description	Seuil
$\geq 98\%$	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	Score fréquent	Score à interpréter ultérieurement en fonction du niveau de performance présumée
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau	Score jugé hors norme avec un seuil libéral
$\leq 2\%$	Score très peu fréquent de bas niveau	Score jugé hors norme avec un seuil conservateur

Proposition de qualificatifs



Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description	Seuil
$\geq 98\%$	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	Score fréquent	Score à interpréter ultérieurement en fonction du niveau de performance présumée
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau	Score jugé hors norme avec un seuil libéral
$\leq 2\%$	Score très peu fréquent de bas niveau	Score jugé hors norme avec un seuil conservateur





Scores de bas niveau ou de haut niveau

HAUT NIVEAU



6,26 m



2 m

10''



40''

BAS NIVEAU



Scores de bas niveau ou de haut niveau



HAUT NIVEAU



16 mots

RL/RI-16 RL1



310 msec

ALERTE TONIQUE



BAS NIVEAU

4 mots

621 msec



Proposition de qualificatifs



Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description
$\geq 98\%$	Score très peu fréquent de haut niveau
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau
Entre $> 5\%$ et $< 95\%$	Score fréquent
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau
$\leq 2\%$	Score très peu fréquent de bas niveau

Nombre de mot rappelés : $Z = -1,65$

Temps de réponse : ~~$Z = -1,65$~~



Temps de réponse : $Z = -1,65$

Nombre de mot rappelés : P5

Temps de réponse : ~~P95~~



Temps de réponse : P5



Proposition de qualificatifs



Fréquence estimée de la population ayant ce score ou moins (*= risque d'erreur associé à la décision de qualifier un score de hors-normes)	Description	Seuil
$\geq 98\%$	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	Score fréquent	Score à interpréter ultérieurement en fonction du niveau de performance présumée
Entre $> 2\%$ et $\leq 5\%$	Score peu fréquent de bas niveau	Score jugé hors norme avec un seuil libéral
$\leq 2\%$	Score très peu fréquent de bas niveau	Score jugé hors norme avec un seuil conservateur



Une explication claire

Fréquence estimée* de la population ayant ce score ou moins	Score z	QI	Note standard (n)	Note T	Description	Interprétation statistique
$\geq 98\%$	$\geq +2,05$	≥ 131	≥ 16	≥ 71	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	$\geq +1,65$ à $< +2,05$	≥ 125 à < 131	≥ 15 à < 16	≥ 67 à < 71	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	$> -1,65$ à $< +1,65$	> 75 à < 125	> 5 à < 15	> 34 à < 67	Score fréquent	
Entre $> 2\%$ et $\leq 5\%$	$> -2,05$ à $\leq -1,65$	> 69 à ≤ 75	> 4 à ≤ 5	> 30 à ≤ 34	Score peu fréquent de bas niveau	hors norme (seuil libéral)
$\leq 2\%$	$\leq -2,05$	≤ 69	≤ 4	≤ 30	Score très peu fréquent de bas niveau	hors norme (seuil conservateur)

Note. Valeur p associée au test t modifié, percentile et probabilité associée au score *standardisé*.

Une explication claire

Comparaison normative

En ce qui concerne l'interprétation des résultats aux épreuves, les scores obtenus sont comparés aux performances des personnes du groupe normatif (si possible de même âge, sexe et niveau d'éducation). Les résultats sont indiqués en référence à ces données normatives. Elles sont présentées sous forme de fréquence estimée de la population ayant ce score ou moins (la correspondance avec différents scores standardisés utilisés est également notée). Une fréquence de 5% (score z de -1,65, QI de 75, note standard de 5 ou score T de 34) correspond à une valeur en dessous de laquelle se trouve 5 % des performances des personnes du groupe normatif. Le tableau ci-dessous présente la grille d'interprétation des résultats quantitatifs et les scores seuils utilisés pour juger un score comme s'écartant des scores attendus.

Fréquence estimée* de la population ayant ce score ou moins	Score z	QI	Note standard (n)	Note T	Description	Interprétation statistique
$\geq 98\%$	$\geq +2,05$	≥ 131	≥ 16	≥ 71	Score très peu fréquent de haut niveau	
Entre $\geq 95\%$ et $< 98\%$	$\geq +1,65$ à $< +2,05$	≥ 125 à < 131	≥ 15 à < 16	≥ 67 à < 71	Score peu fréquent de haut niveau	
Entre $> 5\%$ et $< 95\%$	$> -1,65$ à $< +1,65$	> 75 à < 125	> 5 à < 15	> 34 à < 67	Score fréquent	
Entre $> 2\%$ et $\leq 5\%$	$> -2,05$ à $\leq -1,65$	> 69 à ≤ 75	> 4 à ≤ 5	> 30 à ≤ 34	Score peu fréquent de bas niveau	hors norme (seuil libéral)
$\leq 2\%$	$\leq -2,05$	≤ 69	≤ 4	≤ 30	Score très peu fréquent de bas niveau	hors norme (seuil conservateur)

Note. Valeur p associée au test t modifié, percentile et probabilité associée au score *standardisé*.

Exceptions à cette proposition

- Effet plafond => pas de « haut niveau »
- Score seuil associé à un score cut-off
- Tests de validité de performance



Questions ou commentaires ?



Merci !

Avez-vous des questions?

Beatrice.Degraeve@univ-catholille.fr

patrick.fery@hubruelles.be

hichem.slama@hubruelles.be

Sylvie.willems@uliege.be



Attestation de présence

<https://surveys.fplse.uliege.be/surveys/?w=xN&s=KWLCXSSIBC>





08

Conclusion

01

02

03

04

05

06

07

08

Merci !

