



HAL
open science

De la vérité autoproclamée à la vraisemblance reconnue

Jean-Claude Regnier

► **To cite this version:**

Jean-Claude Regnier. De la vérité autoproclamée à la vraisemblance reconnue. Jean-Claude Gros; Daniel Gros; Paul Planchette; Jean-Claude Regnier; René Thomas. Enseigner la Statistique du CM à la Seconde Pourquoi? Comment?, IREM de Lyon Université Lyon1, pp.107-118, 1998. halshs-00406007

HAL Id: halshs-00406007

<https://shs.hal.science/halshs-00406007>

Submitted on 29 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

De la vérité autoproclamée à la vraisemblance reconnue

Comment peut-on reconnaître et contrôler les risques encourus lors d'une prise de décision en situation incertaine ? Quelle place et quel rôle tiennent les notions de significativité d'un résultat et de représentativité d'un échantillon dans la prise de décision ?

Jean-Claude Régnier

Les propos qui vont suivre visent à expliciter une procédure d'aide à la prise de décision utilisant un raisonnement statistique. Dans ce raisonnement, l'idée de vérité est remplacée par celle de vraisemblance. Entre deux hypothèses alternatives, il s'agit de décider celle que nous souhaitons tenir pour vraie parce qu'au vu des informations que nous détenons, elle nous apparaît comme la plus vraisemblable des deux. A notre avis, ce raisonnement est conduit plus ou moins spontanément et implicitement par chacun de nous. La statistique constitue alors un domaine scientifique dans lequel sont développés des outils conceptuels et des méthodes permettant de modéliser certaines catégories de situations de décision. Cette modélisation apporte des informations qui constituent une aide à la prise de décision. Cependant nous ne devons pas oublier qu'en fin de compte la décision appartient toujours au décideur qui traitera les informations recueillies selon son point de vue.

Un exemple dans l'évaluation scolaire d'une compétence.

A l'issue d'une séquence didactique, l'enseignant est soucieux de pouvoir porter un jugement relatif au niveau de compétence atteint par chaque apprenant dans le domaine. Un modèle simplifié de la situation d'évaluation réduit la décision à deux modalités : (H_0) *niveau de compétence insuffisant* contre (H_1) *niveau de compétence satisfaisant*.

Pour obtenir une aide à cette prise de décision, l'enseignant recourt à un questionnaire constitué d'une ou plusieurs questions indépendantes ou non les unes des autres, ouvertes, semi-ouvertes ou fermées. Chaque question est construite de manière que la production de la réponse par l'apprenant nécessite la mise en œuvre de la compétence dont le développement est visé.

L'apprenant peut, pour sa part, répondre ou ne pas répondre et quand il fournit une réponse, cette dernière peut être correcte ou incorrecte. Ceci amène à considérer qu'à chaque question est associée l'une des modalités :

1) réponse correcte (R), 2) réponse incorrecte (E) ou 3) non-réponse (N).

Nous présupposons qu'un questionnaire ne peut comporter qu'un nombre fini (au plus 500 selon notre avis) de questions . De ce fait, il ne peut apporter qu'une information partielle relativement à la compétence qui ne peut être cernée exhaustivement au travers d'un nombre fini de questions.

Un type de questionnaire dont la construction et l'usage s'avèrent assez aisés, est le questionnaire à choix multiples (QCM). Il consiste à poser une question et à proposer plusieurs réponses plausibles dont, dans notre situation, une seule est la réponse attendue, celle que l'on considère comme la réponse correcte.

question Q_i (énoncé) -->réponse(k modalités proposées): $m_{1i} \square m_{2i} \square \dots \square m_{ki} \square$

La variable de décision D communément en usage est le nombre de réponses correctes fournies au questionnaire. L'espace des résultats possibles est l'ensemble des nombres entiers de 0 à q pour un questionnaire comportant q questions. Si nous considérons que les questions sont mutuellement indépendantes et d'égales difficultés, cette variable n'est autre qu'une variable binomiale de paramètres q et π où π est la probabilité de fournir une réponse correcte.

Dès lors, nous percevons que la valeur de π est différente selon que nous nous plaçons sous l'hypothèse H_0 ou l'hypothèse H_1 . Il paraît vraisemblable de considérer que la valeur π_1 de π sous H_1 est supérieure à celle π_0 de π sous H_0 .

L'intérêt de ce modèle est qu'il nous permet de calculer directement la probabilité d'obtention d'un résultat k particulier de la variable de décision D "nombre de réussites" en fonction du nombre q de questions et de la probabilité de réussite. Les formules algébriques utiles sont :

$$\begin{aligned} \text{Prob}\{ D = k \} &= \binom{q}{k} \pi^k (1-\pi)^{q-k} = \frac{q!}{k! (q-k)!} \pi^k (1-\pi)^{q-k} \\ \text{Prob}\{ D \geq k \} &= \sum_{j=k}^q \text{Prob}\{ D = j \} \\ \text{Prob}\{ D < k \} &= 1 - \text{Prob}\{ D \geq k \} \end{aligned}$$

Examinons maintenant un exemple.

L'usage français consacre à l'ensemble des nombres entiers de 0 à 20 une place toute particulière dont nous ne connaissons nullement les raisons originelles. Cela se concrétise en considérant un questionnaire à 20 questions et "une note sur 20" qui, dans notre modèle, correspondra au nombre de questions ayant obtenu une réponse correcte du point de vue de l'enseignant-évaluateur.

Situation n ° 1

L'espace de décision est habituellement ainsi formé :

$\{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}$ est la zone de rejet de l'hypothèse de *compétence* ou celle d'acceptation de l'hypothèse d'*incompétence*

$\{ 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 \}$ est la zone d'acceptation de l'hypothèse de *compétence* ou celle de rejet de l'hypothèse d'*incompétence*

Essayons alors de mesurer les risques encourus. Pour cela, nous allons considérer les valeurs de π selon les deux hypothèses H_0 et H_1 . Nous posons que la probabilité π_0 de réussite à une question sous l'hypothèse (H_0) *niveau de compétence insuffisant* vaut 20% ($\pi_0 = 0,20$) et donc que la probabilité de réussite à une question sous l'hypothèse (H_1) *niveau de compétence satisfaisant* vaut 80% ($\pi_1 = 0,80$).

Le premier risque encouru par l'évaluateur est de commettre l'erreur de rejeter à tort l'hypothèse H_0 d'incompétence alors qu'elle est vraie. Cette erreur admet la probabilité α suivante d'être commise en fonction des conditions posées.

$$\alpha = \text{Prob}\{ D \geq 10 / H_0 \text{ vraie} \} = 0,003$$

Le second risque encouru par l'évaluateur est de commettre l'erreur de rejeter à tort l'hypothèse H_1 de compétence alors qu'elle est vraie. Cette erreur admet la probabilité β suivante d'être commise en fonction des conditions posées.

$$\beta = \text{Prob}\{ D < 10 / H_1 \text{ vraie} \} = 0,001$$

Ces deux calculs sont effectués à partir des formules rappelées plus haut à l'aide d'une table, d'une calculatrice ou d'un tableur (comme Excel 5).

Examinons maintenant la variation des valeurs de α et β selon les valeurs de π .

π_0	0,05	0,1	0,2	0,3	0,4	0,5	π_1	0,6	0,66	0,7	0,8	0,9
α	0,000	0,000	0,003	0,048	0,245	0,588	β	0,128	0,043	0,017	0,001	0,000

Il apparaît que si la probabilité de réussite sous l'hypothèse H_0 est inférieure ou égale à 0,3 alors le choix de 10 comme valeur critique frontière entre les deux zones assure un risque de 1^{ère} espèce inférieure à un niveau de 0,05. Ceci est réalisé dès que l'on construit un QCM avec quatre modalités dont une détermine la réponse correcte. Pour un QCM à 3 modalités la probabilité $\pi_0 = 1/3$ conduit une valeur de α entre 0,09 et 0,10.

Quant à l'erreur de 2^{nde} espèce, sa probabilité est inférieure à 0,05 dès que la probabilité de réussite sous H_1 est supérieure à 2/3.

Ainsi le choix de ce critère (à savoir : "10 réponses correctes sur 20 attendues") pour décider entre les deux hypothèses: *compétence* et *incompétence* demeure sous les conditions explicitées un moyen relativement sûr tout en mettant en évidence l'impossibilité de connaître la vérité puisqu'à chaque choix est attaché un risque d'erreur.

Situation n ° 2

Nous allons cette fois fixer la valeur maximale de α à 0,05. De là, nous cherchons la nouvelle partition de l'espace de décision $\{ 0, 1, 2, \dots, 18, 19, 20 \}$ en deux zones contiguës et complémentaires : une zone correspondant au rejet de l'hypothèse de *compétence* (ou zone d'acceptation de l'hypothèse d'*incompétence*) et une zone correspondant à l'acceptation de l'hypothèse de *compétence* (ou zone de rejet de l'hypothèse d'*incompétence*).

Nous supposons que le questionnaire contient des questions à 4 choix. Du point de vue de notre modèle, cela signifie que nous considérons que la conduite de réponse sous l'hypothèse H_0 est analogue à 20 tirages successifs avec remise dans une urne contenant 4 boules : trois boules noires et une boule blanche. La valeur de D correspond alors au nombre de boules blanches obtenues à l'issue des 20 tirages. La probabilité de réussite sous H_0 est ici de 0,25.

L'inéquation que nous avons à résoudre est :

$$\alpha = \text{Prob}\{ D \geq k / H_0 \text{ vraie} \} \leq 0,05$$

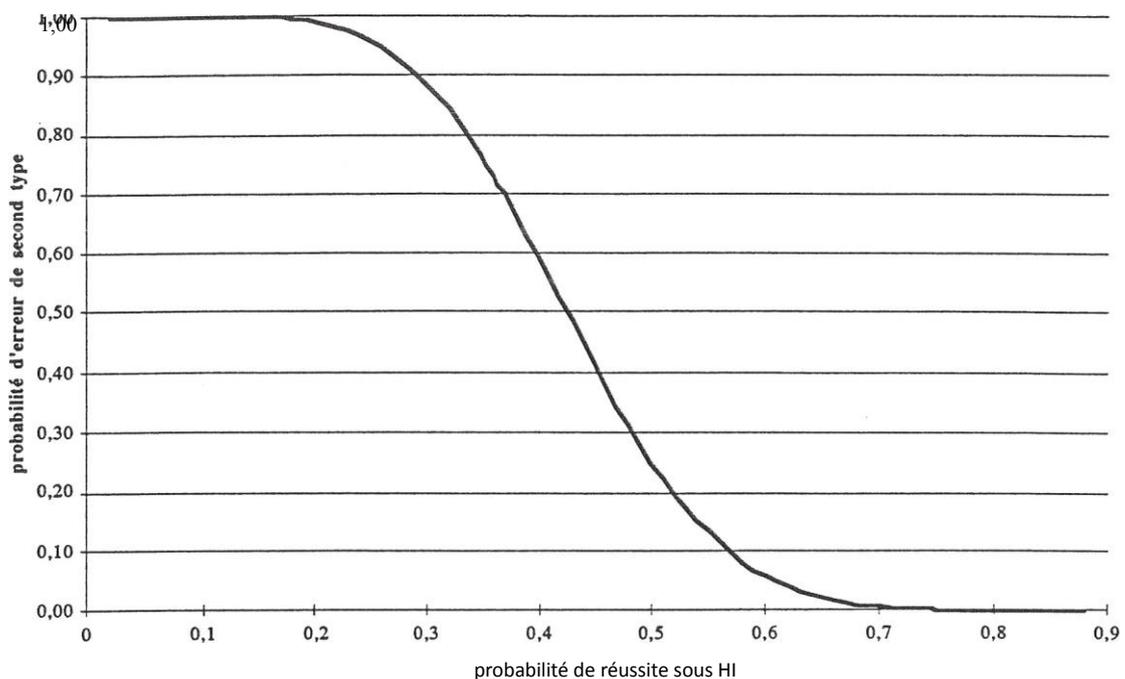
Par le calcul ou à l'aide de table, nous trouvons la valeur $k = 9$. Ainsi l'espace de rejet de H_0 au niveau maximum de risque de 0,05 est alors :

$$\{ 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 \} \text{ appelé zone critique } K.$$

Ce qui signifie qu'à l'issue d'une épreuve fondée sur un QCM à 4 modalités et 20 questions, tout apprenant ayant un score supérieur ou égal à 9 sera déclaré compétent par rejet de l'hypothèse d'incompétence avec un risque d'erreur de niveau inférieur à 0,05

Quand son score est inférieur strictement à 9, il est déclaré incompetent car ce score appartient à l'ensemble $\{ 0, 1, 2, 3, 4, 5, 6, 7, 8 \}$ appelé zone d'acceptation, qui constitue la zone de rejet de l'hypothèse de *compétence* ou zone d'acceptation de l'hypothèse d'*incompétence*

Mais nous ne pouvons oublier que ce score peut aussi être obtenu sous l'hypothèse H_1 pour diverses raisons. La compétence ne conduit pas inévitablement à une réponse correcte avec certitude. Il y a les erreurs que Gaston Bachelard nomme celles de l'esprit fatigué qui ne sont pas significatives d'une représentation erronée, d'une connaissance inadéquate ou fausse. Divers facteurs introduisent des fluctuations qui nécessitent d'adapter le modèle en traduisant la probabilité de réussite à une question par une valeur strictement inférieure à 1. La valeur de la probabilité β va dépendre de la probabilité de donner une réponse correcte sachant que celui qui répond est compétent. La courbe ci-dessous montre la variation de la probabilité de commettre une erreur de 2^{nde} espèce en fonction de la probabilité π de fournir une réponse correcte en étant compétent.



Au-delà d'une probabilité de 50% de réussite à une question sous l'hypothèse de compétence la valeur de β descend au-dessous de 25%.

Raisonnement statistique et test statistique

Le raisonnement utilisé prend appui sur le schéma suivant en considérant les deux hypothèses alternatives H_0 et H_1

		« Situation de nature » inconnue	
		Ho vraie H1 fausse	H1 vraie Ho fausse
Décision du chercheur	Rejeter Ho Conserv. H1	Erreur de première espèce α	Correcte
	Conserv. Ho Rejeter H1	Correcte	Erreur de seconde espèce β

		« Situation de nature » inconnue	
		Ho vraie H1 fausse	H1 vraie Ho fausse
Décision du chercheur	Rejeter Ho Conserv. H1	Prob{Rejeter Ho sachant Ho vraie}= α	Prob{Conserv. H1 sachant H1 vraie}= $1-\beta$
	Conserv. Ho Rejeter H1	Prob{Conserv. Ho sachant Ho vraie}= $1-\alpha$	Prob{Rejeter H1 sachant H1 vraie}= β

En exploitant l'exemple précédent nous pouvons constater que les valeurs de α et β varient entre 0 et 1 en sens inverses. Diminuer la probabilité du risque d'erreur de type I revient à augmenter celle du risque d'erreur de type II.

Certes il convient de rappeler que toutes les hypothèses formulables ne sont pas susceptibles d'être validées ou rejetées à l'aide d'un test statistique dont l'objet est de produire une procédure permettant de construire un espace de résultats pour une variable de décision à partir desquels le choix de l'une ou l'autre des hypothèses est réalisé avec un risque calculé. Mais lorsque nous pouvons construire un modèle dans lequel les hypothèses H_0 et H_1 sont caractérisées par des notions et des propriétés du domaine de la statistique mathématique, nous avons grand intérêt à rechercher dans la boîte à outils statistiques un test d'hypothèse adapté à la situation pour aider la prise de décision. Cependant les conditions d'application tant théoriques que pratiques doivent être abordées et discutées pour juger de l'adéquation du modèle théorique à la situation réelle étudiée. Sans cette précaution relative à l'applicabilité du test statistique, nous nous exposons à un risque d'erreur du type III : erreur de choix du modèle ou erreur de choix du test statistique. Lorsque deux tests peuvent être utilisés avec

des conditions théoriques et pratiques comparables, nous choisissons le plus puissant des deux, c'est à dire celui qui pour un niveau α donné conduit à un niveau de risque β le plus bas possible ou encore à une probabilité $1 - p$ de rejeter l'hypothèse H_0 alors que H_1 est vraie, la plus forte.

Un **test statistique** est donc une procédure logico-mathématique sur laquelle on s'appuie pour prendre une décision de choix entre deux **hypothèses alternatives** sur la base d'informations partielles issues d'un **échantillon**. Un test statistique d'hypothèse consiste à **définir explicitement une règle communicable de décision**.

Une hypothèse est dite *hypothèse simple* si elle peut se ramener à une confrontation simple, par exemple, une hypothèse centrée sur une valeur isolée d'un paramètre (moyenne, variance, écart-type, médiane, mode, coefficient d'asymétrie, coefficient d'aplatissement, etc...) du type $\Theta = \Theta_0$.

Une hypothèse est dite **hypothèse composite** dans le cas contraire, c'est à dire par exemple, si elle peut se ramener à une confrontation complexe avec plusieurs valeurs du type : $\Theta > \Theta_0$, $\Theta \geq \Theta_0$, $\Theta < \Theta_0$, $\Theta \leq \Theta_0$, $\Theta \neq \Theta_0$, ou $\Theta \in E$.

Comme nous l'avons dit plus haut si H_0 (dite *hypothèse nulle*) et H_1 (dite *hypothèse alternative*) sont ces deux hypothèses alternatives dont une seule est vraie, nous avons quatre cas décisionnels. Deux cas correspondent à une décision correcte et les deux autres, à une décision erronée. Dans l'usage courant l'hypothèse nulle H_0 joue un rôle prééminent qui conduit à contrôler le risque de première espèce α en posant généralement sa valeur égale à 0,01 ou le plus souvent à 0,05 mais parfois à 0,001. Cette hypothèse nulle H_0 est établie à partir de divers facteurs : la prudence, la subjectivité à laquelle on tient particulièrement, la facilité à formuler, ou encore le fait qu'il s'agit d'une hypothèse solidement établie et non contredite jusqu'alors par l'expérience que l'on met à l'épreuve dans le but de ne pas la rejeter.

La probabilité β du risque de seconde espèce est obtenue par le calcul. Cependant cela n'est possible que dans les cas où les distributions de probabilité sous l'hypothèse H_1 sont connues.

La probabilité $1 - \beta$ d'opter avec raison pour H_1 s'appelle la **puissance du test**.

Lorsqu'on a fixé α , il faut alors choisir une variable de décision :

- qui apporte le plus possible d'informations relatives au problème posé,

- dont la distribution de probabilité est différente selon que H_0 ou H_1 est vraie
- dont la distribution de probabilité est bien connue au moins sous la condition H_0 est vraie.

On définit alors la région critique K , c'est à dire l'ensemble des valeurs de la variable de décision qui conduisent à rejeter l'hypothèse nulle H_0 au profit de l'hypothèse alternative H_1 .

La région critique est déterminée par : $\text{Prob}\{\text{rejeter } H_0 / H_0 \text{ vraie}\} = \alpha$

On définit aussi la région d'acceptation A , c'est à dire l'ensemble des valeurs de la variable de décision qui conduisent à ne pas rejeter l'hypothèse nulle H_0 au détriment de l'hypothèse alternative H_1 . La région critique K et la région d'acceptation A sont deux ensembles complémentaires.

Dans cette perspective construire un test revient à construire une variable de décision explicite et à déterminer une région critique.

Démarche à suivre pour tester une hypothèse

- [1] identifier et formuler les deux hypothèses H_0 et H_1 , de telle sorte qu'elles correspondent à des notions et des propriétés du domaine de la statistique mathématique,
- [2] déterminer une variable de décision,
- [3] caractériser l'allure de la région critique K en fonction de l'hypothèse H_1 ,
- [4] déterminer en compréhension ou en extension par le calcul, la région critique K en fonction de la valeur α du risque de première espèce,
- [5] calculer, quand cela est possible, la puissance du test $1 - \beta$,
- [6] calculer la valeur expérimentale de la variable de décision à partir des valeurs obtenues sur un échantillon.
- [7] formuler la conclusion en termes de rejet ou non rejet de l'hypothèse nulle H_0 .

Les **tests statistiques** sont classés selon leur objet. Citons pour exemples ceux que l'on est amené à utiliser dans des recherches en Sciences de l'éducation et de la Formation:

Tests paramétriques, c'est à dire visant à tester une hypothèse relative à un ou plusieurs paramètres (essentiellement : moyenne, variance, proportion, etc...) d'une variable aléatoire spécifiée ou non :

- tests de comparaison d'une moyenne à une valeur fixée,
- tests de comparaison d'une variance à une valeur fixée,
- tests de comparaison d'une proportion à une valeur fixée,
- tests de comparaison d'une différence de deux moyennes à une valeur fixée,
- tests de comparaison de deux ou plusieurs moyennes,
- tests de comparaison de deux ou plusieurs variances,
- tests de comparaison de deux ou plusieurs proportions,
- tests d'indépendance de deux variables quantitatives

Tests non-paramétriques dont la construction est établie sur la base d'une *fonction des observations* issues d'un échantillon aléatoire, de distribution de probabilité indépendante de la connaissance de la distribution de probabilité de la variable étudiée sur la population.

- tests d'ajustement ou d'adéquation d'une distribution observée avec une distribution théorique
- tests d'indépendance de deux variables qualitatives,
- tests d'homogénéité des profils selon les modalités d'une variable qualitative - tests fondés sur les statistiques de rang.

À ce jour, la boîte à outils statistiques contient une panoplie extraordinaire de tests et continue à s'enrichir. De plus, rien n'empêche le chercheur aidé du statisticien de construire un test particulier adapté à la spécificité de la recherche conduite. Le chercheur en Sciences de l'éducation et de la formation soucieux de la mise à l'épreuve des hypothèses qu'il avance en réponse à une question centrale de sa problématique, peut dans la plupart des cas facilement trouver un outil adapté pour l'aider à prendre une décision sur la base de l'épreuve des faits. Cette perspective n'est en rien réduite à un point de vue quantitativiste étroit comme il est de coutume de le considérer. Elle se donne simplement comme but une réduction de la subjectivité en produisant des règles les plus explicites possibles de décision dont le mérite est de pouvoir être objet d'un débat méthodologique. C'est ainsi que le terme significatif prend un sens un peu moins flou que dans l'acception commune. Il ne s'agit pas d'une conception qui ferait inéluctablement basculer dans un impérialisme mécanique et froid de la prise automatique de décision fondée exclusivement sur la connaissance d'une quantité. En revanche cette conception éloigne du risque de l'enfermement dans un discours énonçant des propos qui paraissent auto-proclamer leur vérité sans en autoriser la falsification.

La perspective offerte par l'approche statistique des tests d'hypothèse ne conduit à aucun moment à l'incontestable puisque l'idée de risque y est explicitement présente et que la proposition choisie est acceptée non parce qu'elle est interprétée et proclamée vraie mais parce qu'elle est jugée la plus vraisemblable au regard des autres propositions au vu des informations recueillies dans le cadre d'un modèle théorique construit par le chercheur. Le doute demeure même après la décision prise. Ainsi nous ne pourrions pas, sur la base d'un test statistique d'hypothèse, obtenir une telle conclusion extraite d'un texte¹ du champ des sciences de l'éducation et de la formation:

¹ Dans ces textes, c'est qui soulignons

« Les conséquences de la mise en place de telles pratiques² ont été **incontestablement positives** dans le registre de la motivation des élèves; elles ont, **incontestablement**, contribué à réconcilier de nombreux élèves avec la pratique de la langue écrite, **ce qui n'est pas peu de chose**³...»⁴ Cet extrait et le suivant nous montrent le recours à *l'idée de signification des informations sur lesquelles se fonde l'opinion*.

« De plus, le développement d'institutions scolaires sur le modèle occidental dans les pays du Tiers Monde a produit, **de toute évidence**, des **effets souvent négatifs**, discréditant des savoirs traditionnels parfois fort utiles sans parvenir à transmettre des savoirs nouveaux de **manière significative**.»⁵

Cette idée de caractère **significatif des informations** fait aussi l'objet d'une tentative d'explicitation dans cette approche. En quel sens est-il utilisé ? dans celui de considérer ce que peut signifier un résultat empiriquement obtenu par comparaison à celui qu'on obtiendrait théoriquement sous certaines conditions (en particulier sous H_0 ou sous H_1). L'idée est alors de faire en sorte que le risque encouru lors de la prise de décision ne dépende que de la fluctuation de l'échantillonnage et non des erreurs d'observation, de mesure ou de report de résultats. Ici il y aurait lieu de réfléchir aux effets des approximations numériques par arrondi. Dans l'approche que nous avons développée dans cet article, l'idée de **significativité** est rattachée à celle de probabilité d'obtention d'un résultat par le seul fait du hasard (presque dans le sens étymologique où le résultat pourrait obtenu avec un jeu de dés adapté). Plus la valeur de la probabilité de cet événement est grande, moins le résultat est considéré comme **significatif**, c'est à dire apportant une signification au phénomène étudié et à l'hypothèse que l'on a émise à son propos. Ainsi dans notre exemple l'hypothèse H_0 d'incompétence est soumise à l'épreuve des faits. Obtenir un score supérieur ou égal à 9 est plus probable sous l'hypothèse H_1 de compétence que sous l'hypothèse H_0 d'incompétence. Un score de 12 sera dit **significatif** dans la mesure où il traduit l'idée que ce que nous observons (le score) est une réalisation de l'événement $\{D \geq 9\}$ qui admet une plus forte probabilité quand il est lié au phénomène (la compétence) que nous avons formalisé (recours au questionnaire, variable de décision "nombre de réussites") qu'au hasard en modélisant la conduite de réponse sous l'hypothèse H_0 (d'incompétence) par une suite de tirages dans une

² Il s'agit des pratiques pédagogiques liées aux « méthodes actives »

³ Expression que nous traduisons par significatif

⁴ Meirieu, Ph. (1993) Objectif, obstacle et situations d'apprentissage, in J. Houssaye (Coord) *La pédagogie : une encyclopédie pour aujourd'hui* ESF p 293

⁵ Ibidem p 290

urne contenant une boule "réponse correcte" et trois boules "réponse incorrecte". Pour donner un ordre de grandeur $\text{Prob}\{D \geq 9 / \pi_0 = 0,25\} \approx 0,041$ alors que $\text{Prob}\{D \geq 9 / \pi_1 = 0,80\} \approx 0,999$. Cette approche conduit à des énoncés du type suivant : «Au vu des résultats empiriques échantillonnaires, ces résultats sont significatifs alors il est plus raisonnable de prendre la décision de rejeter l'hypothèse H_0 si nous acceptons de prendre le risque de type I à un niveau α inférieur ou égal à 5% (1 % ou 0,1%) ou *ces résultats ne sont pas significatifs* alors il est plus raisonnable de conserver cette hypothèse H_0 mais avec un risque de type II à un niveau β souvent inconnu».

Pour achever cet article, nous évoquerons alors la notion de **représentativité** de l'échantillon utilisé pour obtenir une réalisation de la (ou des) variable(s) étudiée(s) et calculer la valeur prise par la variable de décision. Nous avons plusieurs fois trouvé dans des mémoires de licence, de maîtrise ou de DEA en sciences de l'éducation des propos émettant des réserves à l'égard de la **représentativité de l'échantillon** utilisé sans pour autant suspendre l'usage de l'information ainsi recueillie dans l'argumentation y compris en s'appuyant sur la **significativité**. Il nous a semblé qu'il s'agissait plus d'une précaution formelle que d'une remarque issue d'une analyse de la méthode d'obtention de l'échantillon. Ce que confirme la manière dont les informations échantillonnaires sont reprises dans l'argumentation. Ainsi abordée ni la **représentativité** ni la **non-représentativité** et le **bias** qu'elle induit, ne sont en fait établis autrement que par une opinion si nous nous en tenons à ce qui est écrit. En revanche, les erreurs possibles liées aux observations, aux mesures ne sont, elles, jamais évoquées. Au cours des accompagnements méthodologiques que nous avons avec les étudiants, la question de la **représentativité** est posée systématiquement mais d'abord et parfois même exclusivement par la taille de l'échantillon.

En quel sens pouvons-nous utiliser le qualificatif **représentatif** pour un échantillon d'une population (parente) ? ... dans le sens d'un fragment de cette population nous permettant d'obtenir une information la plus proche possible de celle que nous obtiendrions par recensement. De là, nous pouvons y voir l'idée d'un modèle réduit . En quelque sorte nous considérons que l'obtention de l'information sur la population résulte d'une transformation homothétique de l'information obtenue sur le fragment.

Une autre façon d'envisager cette notion de **représentativité** est de considérer la probabilité d'appartenance d'une unité statistique à l'échantillon. Ainsi en nous rattachant aux idées d'équité et de compensation, une "certaine" représentativité de l'échantillon paraît être

recevable si nous respectons la condition suivante : chaque unité statistique a la même chance d'appartenir à l'échantillon. Ainsi posée cette caractérisation de la **représentativité** d'un échantillon, nous sommes confronté au problème de la constitution d'un échantillon de taille n respectant cette contrainte. La méthode consiste à extraire successivement n unités au hasard avec ou sans remise en utilisant des tables de nombres aléatoires ou des fonctions aléatoires dans un tableur (Excel 5). De ce point de vue, la taille est plus un facteur d'accroissement de la quantité d'information que d'accroissement de la représentativité. Un échantillon de taille 1 est représentatif dans ce sens-là et il peut même comporter beaucoup d'informations : imaginons une boîte contenant 10000 vis identiques, il suffit de mesurer un élément pour en connaître la dimension et avoir l'information complète sur la population.

La connaissance du sens particulier des termes en usage est importante pour comprendre des textes énonçant des résultats fondés sur des approches statistiques. Pour illustrer nous livrons au lecteur le texte⁶ suivant issu d'une recherche sur le thème de l'autoévaluation et de la métacognition afin qu'il puisse exercer sa compréhension : «L'échantillon consiste en un groupe de 30 étudiants québécois de 11^{ème} année composé de 13 filles et 17 garçons (âge moyen : 16 ans). Les sujets du *groupe fort* (groupe du 3^{ème} quartile) partagent davantage la même opinion en ce qui concerne les questions qui seront choisies pour l'examen que les sujets du *groupe faible* (groupe du 1^{er} quartile). En effet le W de concordance de Kendall⁷ pour les huit étudiants-juges du *groupe fort* est de 0,2362 (significatif à 0,01) alors qu'il n'est que de 0,1208 ($p = 0,348$) pour le groupe faible. Ceci indique que les sujets ayant le moins bien réussi à l'examen sont aussi ceux qui sont le plus en désaccord sur le type de questions qui seront sélectionnées par l'enseignant

Les sujets du *groupe fort* manifestent également une plus grande concordance en ce qui concerne la préparation. En effet ce groupe a obtenu un $W=0,1946$ (significatif à 0,05) alors que le *groupe faible* a obtenu un $W=0,1381$ ($p = 0,190$). On peut interpréter ce résultat comme **signifiant** que le *groupe fort* a une plus grande tendance à faire porter les efforts d'étude sur les mêmes items que le *groupe faible* »

⁶ Laveault, D.& Fournier, Ch. (1990) Évaluation par objectifs : une approche métacognitive Revue *Mesure et Évaluation* vol 13.1 pp 54-57

⁷ Régnier, J.- C. (1996) Test de concordance des rangements de n objets selon p critères in *Méthodes quantitatives & statistique* , Tome I - p 91, Documents photocopiés - Université Lyon2 - ISPEF

Donner un sens plus spécifique et mieux circonscrit à des notions fréquemment usitées dans la langue commune ou correspondant à des pratiques concrètes, comme celles de risque, de significativité ou de représentativité qui interviennent dans nos prises de décision, en les replaçant dans les domaines de la statistique mathématique et de la statistique appliquée.