

Impact du facteur vocabulaire sur la compréhension des tests statistiques

Impact of the vocabulary factor on the understanding of statistical tests

Zahid El m'hamed^{1*}

¹ Centre Régional des Métiers de l'Éducation et de la Formation, Région de Rabat-Salé-Kénitra, Rabat, Maroc

Résumé. Une littérature abondante de travaux de recherches considère l'enseignement des tests statistiques comme un sujet difficile aussi bien pour l'enseignant que pour l'apprenant. Elle rapporte diverses difficultés, souvent en rapport avec des conceptions erronées, que l'on peut rencontrer à tous les âges et à tous les niveaux d'expertise. Les conclusions de ces recherches s'accordent sur le fait qu'en général, un étudiant de niveau moyen ne peut pas décrire l'idée de fondement de cette pratique statistique, seul un calcul mécanique est souvent restitué, basé beaucoup plus sur la mémorisation que sur la réflexion et l'interprétation. Certes, de multiples facteurs sont à l'origine de ce caractère difficile des tests statistiques. Ainsi, nous nous sommes focalisés dans cet article sur l'effet que pourrait avoir un facteur plus particulier sur l'interprétation, par nos étudiants, des notions et des expressions véhiculées lors de son enseignement. Il s'agit du *vocabulaire* utilisé par les procédures sous-jacentes à cet outil statistique. Afin de mener à bien ce projet, nous avons choisi un échantillon constitué de 195 étudiants, auquel nous avons administré un questionnaire composé de quatre questions suscitant chacune un aspect d'incompréhensions relatif au vocabulaire véhiculé. Les résultats auxquels nous nous sommes arrivés ont mis en évidence des anomalies à plusieurs niveaux chez ces étudiants, mettant ainsi en valeur que les difficultés rencontrées par nos apprenants relativement aux concepts mathématiques d'une manière générale pourraient parfois être dues à des facteurs qu'on néglige souvent dans notre enseignement et qui sont méconnus.

Abstract. The literature of didactic research declares that the teaching of statistical tests is difficult for the teacher as well as the learner. So, it has revealed that there are various difficulties, often with respect to some misconceptions, which are encountered at every age and level of expertise. The research findings show that a student with middle level cannot generally understand and describe the fundamental idea behind statistical tests. Instead, students generally rely on mechanical calculus, which is

*Corresponding author : mhamdizahid@yahoo.fr

based much more on memorization and computation, than on reflection and interpretation. Admittedly, many factors are at the origin of this difficult aspect of statistical tests. Thus, we focused in this article on the effect that could have a very particular factor on the interpretation, by our students, of the notions and expressions conveyed during his teaching. This is the *vocabulary* used by the procedures underlying this statistical tool. In order to carry out this project, we selected a sample composed of 195 students, to whom we administered a questionnaire made up of four questions each causing an aspect of misunderstanding relating to the vocabulary conveyed. The results we have reached have highlighted anomalies on several levels among these students, thus highlighting that the difficulties encountered by our learners in relation to mathematical concepts in general could sometimes be due to factors that are often overlooked and unrecognized in our teaching.

1 Introduction

Les tests statistiques sont une source de conceptions erronées très diverses que partagent la plupart des enseignants avec leurs apprenants [1 - 3]. De plus, une littérature abondante de travaux de recherches montre de manière évidente l'existence de difficultés de compréhension de cet outil statistique chez les utilisateurs à tous les âges et à tous les niveaux d'expertise [4 - 8]. Ainsi, le travail que nous exposons dans cet article s'inscrit dans le cadre de la continuité des réflexions et recherches menées sur la question d'exploration des conceptions erronées chez nos apprenants relativement aux tests statistiques[†]. La portée des conclusions dégagées à la lumière de l'analyse concernant les facteurs susceptibles de contribuer aux difficultés de compréhension des tests statistiques, bien qu'elle soit fondée sur la revue de plusieurs travaux didactiques antérieurs, ne pourra que gagner en pertinence si elle est accompagnée d'observations effectives auprès des apprenants. Très particulièrement, nous nous focalisons dans ce travail sur la question de recherche suivante :

« Le vocabulaire utilisé en tests statistiques a-t-il un effet négatif sur la compréhension de cet outil statistique par les étudiants en dernière année de licences scientifiques, relevant d'universités du Maroc ? ».

Mais auparavant, nous allons, dans un premier temps, poser notre problématique, dans un deuxième temps, présenter le cadre de référence théorique sur lequel ce travail est basé et, dans un troisième temps, mettre en œuvre quelques conceptions erronées relatives aux tests statistiques, susceptibles d'être causées par un tel vocabulaire.

2 Problématique

Les tests statistiques sont caractérisés par une spécificité étrange qui les différencie des autres outils de la boîte à outils de l'inférence statistique. En effet, ce qui est généralement enseigné aujourd'hui relativement à cet outil, dans des champs aussi variés que la psychologie, la sociologie, l'économie, l'éducation, la médecine, le droit et, plus récemment, l'écologie, n'est qu'un *rituel* (ou une logique hybride) souvent appelé : « *tests de signification de l'hypothèse nulle* » [9 - 12]. Ce rituel est, en quelque sorte, un amalgame d'idées incompatibles dérivant de la théorie des « tests de signification » de Fisher et de celle des « tests d'hypothèse » de Neyman-Pearson. Le caractère difficile de cet outil a permis de le qualifier comme étant le

[†] Tout au long de cet article, l'expression « tests statistiques » se rapporte à la fois au test de signification de Fisher et au test d'hypothèse de Neyman-Pearson.

foyer d'une opération croissante de critiques parfois fondées et dans la plupart des cas non fondées, et un sujet de discussions et de débats qui ont duré plusieurs années [13 - 17]. Aussi, il existe des détracteurs qui ont critiqué l'usage des tests statistiques en recherche expérimentale [14, 18], en proposant leur pure et simple exclusion de l'enseignement ou au moins les accompagner d'autres méthodes statistiques, telles que les intervalles de confiances, la taille de l'effet, la statistique bayésienne, la méta-analyse, la réplicabilité, etc. Cependant, la réalité qui s'impose est que cet outil statistique est utilisé par le passé, continue de l'être aujourd'hui et restera certainement pour plusieurs années encore.

Par ailleurs, plusieurs facteurs sont évidemment la cause de telles difficultés de compréhension des tests statistiques. Ainsi, nous sommes convaincus que ce caractère difficile ne sera surmonté qu'à travers une étude plus approfondie des causes sous-jacentes. De plus, nous pensons que la plupart de ces facteurs peuvent raisonnablement être repérés et bien compris. Mais quelques autres facteurs sont soit ignorés ou méconnus. Par ailleurs, nous allons nous focaliser dans cet article sur l'exploration chez nos étudiants de l'effet d'un facteur particulier, souvent déconsidéré et négligé par les études exploratoires déjà établies, relatives aux incompréhensions éventuelles en matière de statistique d'une manière générale et des tests statistiques d'une manière particulière. Ce facteur peut, d'une part, être une source de conceptions erronées relativement aux tests statistiques, et d'autre part, il n'a pas été considéré par les travaux déjà réalisés dans ce sens. Spécifiquement, il s'agit dans ce sens du « *vocabulaire* » utilisé par les tests statistiques, sur lequel se basent, *implicitement* ou *explicitement*, les procédures sous-jacentes à cet outil statistique.

Ainsi, nous faisons remarquer que le développement des concepts dépend intimement du vocabulaire utilisé pour les véhiculer dans les procédures sous-jacentes [19]. Nous ne pouvons pas savoir des concepts sans vocabulaire, et sans des mots représentant ces concepts. Nous choquons occasionnellement ici l'étudiant qui lie toujours le mot « *population* » à un ensemble *d'êtres humains* seulement. De plus, les étudiants ont notoirement des difficultés avec l'apprentissage des concepts statistiques. Ils ne comprennent pas bien des concepts de base. D'une manière équivalente, *les étudiants ne savent pas bien le vocabulaire de base, et alors ils auront un énorme trouble en développant des concepts plus complexes* [20]. Dans la plupart des cas, le vocabulaire statistique (de base) utilise des mots de tous les jours, souvent avec beaucoup de significations spécialisées. Les étudiants devront alors accomplir deux tâches : savoir la signification spécialisée, et en même temps imaginer sa signification dans la vie de tous les jours. De plus, beaucoup de ces concepts de base sont superficiellement simples, mais ils sont en réalité beaucoup plus complexes et/ou subtils. Par exemple, le terme « *population* » que nous avons déjà évoquée, peut signifier, dans le langage de tous les jours, à la fois l'ensemble des gens qui vivent dans un pays, ou l'effectif de tels gens. En statistique, la signification est restreinte à la série des gens seule. Effectivement, sa signification est beaucoup liée à l'idée du groupe d'*intérêt* sur lequel reposera l'étude en question [20].

Le caractère polysémique du vocabulaire utilisé, relativement aux tests statistiques, constitue l'un des principaux facteurs induisant des difficultés de compréhension de cet outil. En effet, les deux mots : « *Prouver* » et « *Vrai* », habituellement utilisés en procédures de tests statistiques, constituent d'énormes facteurs susceptibles de contribuer à une mauvaise appréhension de cet outil [20]. On utilise familièrement l'expression « *vous pouvez prouver[‡] quelque chose à l'aide de la statistique inférentielle !* ». Malheureusement, la réalité est toute autre : « *vous ne pouvez rien prouver à l'aide de cette discipline !* ». Ainsi, un problème sémantique se pose. En effet, pour les mathématiciens (incluant beaucoup de statisticiens), prouver une assertion, signifie la présentation d'une démonstration rigoureuse justifiant sa véracité. Pour les autres chercheurs, en l'occurrence ceux des sciences humaines qui utilisent

[‡] Le mot « *Prouver* » est utilisé ici dans son sens strict des mathématiques.

habituellement les tests statistiques dans leurs recherches, cela semble plutôt signifier : « *convaincs-moi tout simplement !* ».

D'une manière générale, dans plusieurs domaines tels que le droit, où les tests statistiques jouent un rôle très important, la notion de preuve n'est pas absolue. Une déclaration pourra être considérée comme « *prouvée* » si la preuve avancée en sa faveur semble être « *assez solide* », et que cela semble être une base satisfaisante à une action future. Ainsi, si on « *prouve* » qu'une personne est coupable de meurtre, cela signifie que la « *preuve* » avancée contre cette personne est assez solide, et qu'il est donc raisonnable de la condamner. « *Preuve* » dans ce cas signifie « *preuve au-delà du doute raisonnable* » et non pas « *au-delà du doute* ». Bref, dans ce type de contextes, « *prouver* » une déclaration signifie tout simplement : « *convaincs-moi dans la mesure que je le croie !* ». De plus, même la démonstration mathématique qu'on croyait être une preuve universelle et indiscutable, n'est pas ainsi parmi les mathématiciens. Ainsi, dans ce sens, Lamrabet [21] a pu établir une liste non exhaustive d'aspects mettant en évidence le caractère relatif des démonstrations mathématiques. On peut citer à titre d'exemples : « *la relativité par rapport à une époque donnée* », « *la relativité par rapport au courant de pensée mathématique adopté (formalisme, constructivisme, ...)* », « *la relativité par rapport aux moyens utilisés pour produire la démonstration (automatique, expérimental, ...)* », « *la relativité par rapport au degré d'exigence du lecteur* », ...

Par conséquent, la tendance à concevoir un test statistique d'une façon mécanique est accentuée dans l'esprit de plusieurs étudiants et professionnels. En tests statistiques, rien ne peut être plus loin de la vérité[§] du fait qu'ils se basent sur la notion de probabilité conditionnelle. Il est clair que des outils mathématiques sont utilisés en tests statistiques, et souvent ces outils sont très complexes, mais la situation reste la même que dans le domaine de l'architecture, par exemple, où les calculs mathématiques aident seulement à la réalisation du design d'un bâtiment. Cette conception est donc exacerbée par le vocabulaire généralement utilisé, à savoir l'utilisation des mots « *prouver* » et « *vrai* ». Le fait de parler de valeurs vraies d'un paramètre, de preuve, de règles de décision et de niveaux de signification 1%, 5%, est aussi trompeur. Un test statistique est non directement concerné par la valeur vraie d'un paramètre, mais par les modèles probabilistes qui traduisent au mieux la situation étudiée [22]. Accepter un modèle associé à une valeur spécifique d'un paramètre, signifie seulement qu'il est avantageux (*valable*) de baser nos prédictions sur un tel modèle ; le rejeter signifie qu'il est préférable d'utiliser un autre modèle, basé sur une autre valeur différente de ce paramètre.

De ce fait, les termes (polysémiques) du vocabulaire auxquels nous nous sommes intéressés dans ce travail apparaissent très simples, et ils sont en nombre de quatre. Très spécifiquement, il s'agit des termes « *Nulle* », « *Prouver* », « *Vraie* » et « *Significatif* » que comportent respectivement les expressions usuelles suivantes : « *Hypothèse nulle* », « *Prouver la véracité d'une hypothèse* », « *Hypothèse vraie* » et « *Résultat significatif* », qu'on ne cesse d'interpeller chaque fois qu'un test statistique est mis en pratique.

3 Cadre de référence théorique

Le cadre de référence théorique de notre recherche est principalement basé sur la théorie des champs conceptuels. C'est « une théorie cognitiviste qui vise à fournir un cadre cohérent et quelques principes de base pour l'étude du développement et de l'apprentissage des compétences complexes, notamment de celles qui relèvent des sciences et des techniques » [19]. En d'autres termes, dans cette théorie, Vergnaud [19] considère « un espace de problèmes ou de situations-problèmes dont le traitement implique des concepts et des

[§] Et rien ne peut être plus proche de la vérité.

procédures de plusieurs types en étroites connexions, ainsi que les représentations langagières et symboliques susceptibles d'être utilisées pour les représenter ». Plus précisément, ce qui nous intéresse dans cette théorie est les définitions données aux deux notions : « *concept* » et « *conception* ». En effet, ces dernières se manifesteront, comme nous allons le remarquer, dans toutes les étapes principales de la présente recherche, et seront fréquemment interpellées à plusieurs occasions.

3.1 Concept au sens de la théorie des champs conceptuels

Avant de mettre en place la définition d'un concept, il est indispensable de définir la notion de « *schème* », développée dans le cadre de la théorie des champs conceptuels. Ainsi, un schème est défini comme étant un procédé qui décrit l'organisation invariante de la conduite d'une personne dans une classe de situations. « C'est dans les schèmes qu'il faut rechercher les connaissances-en-acte du sujet, c'est-à-dire les éléments cognitifs qui permettent à l'action du sujet d'être opératoire » [19]. Il est composé de « *règles d'action* » (i.e., des règles auxquelles les stratégies ou les réponses dans les productions des élèves sont conformes), et « *d'invariants opératoires* » que la didactique des mathématiques a investi pour ses recherches. Vergnaud a pu distinguer trois types d'invariants opératoires :

- des invariants de type « *proposition* » dont les *théorèmes-en-acte* font partie. Un théorème-en-acte désigne « les propriétés des relations saisies ou utilisées par l'élève en situation de résolution de problème, étant entendu que cela ne signifie pas qu'il est capable pour autant de les expliciter ou de les justifier » ;
- des invariants de type « *fonction propositionnelle* » : ils sont indispensables à la construction de propositions ; les *concepts-en-acte* ou *catégories-en-acte* en font partie ;
- des invariants de type « *argument* » ; en mathématiques, les arguments peuvent être des objets, des nombres, etc.

Rappelons au passage qu'un concept se construit, pour l'individu, à travers les nombreuses situations qui le questionnent ou l'utilisent. Et le sens qui se construit est lié aux types de situations où l'individu l'a rencontré. Il reste seulement à préciser que *le langage, les représentations graphiques ou géométriques, les tableaux*, etc. participent aussi à la construction du sens du concept, et ils ont une double fonction de communication et de représentation.

En se basant sur ce qui précède, Vergnaud modélise un concept par « *un triplet de trois ensembles $C = \{S, I, S\}$ où :*

- *S est l'ensemble des situations qui donnent du sens au concept (la référence) ;*
- *I est l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié) ;*
- *S est l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les situations et les procédures de traitement (le signifiant)».*

3.2 Conception au sens de la didactique française des mathématiques

Après avoir mis en place le sens d'un concept à partir des situations qui le caractérisent (invariants opératoires et signifiants associés), il s'agit maintenant de construire la notion de « *conception* » afin de rendre compte des connaissances d'un individu sur un concept et du fonctionnement de ces connaissances à un moment donné.

Une première idée de conception à destination de la didactique française des mathématiques a été bâtie à partir des notions de représentation et de conception utilisées par

la psychologie cognitive, mais aussi à partir de la définition donnée à un concept. En effet, Brousseau [23] décrit une conception comme « un ensemble de règles, de pratiques, de savoirs qui permettent ensemble de résoudre une classe de situations et de problèmes de façon à peu près satisfaisante, alors qu'il existe une autre classe de situations où cette conception échoue, soit qu'elle suggère des réponses fausses, soit que les résultats soient obtenus plus difficilement et dans des conditions plus défavorables ».

De son côté, Artigue [24] considère que l'idée de conception a un objectif double. Elle permet de :

- mettre en évidence la pluralité des points de vue possibles sur un même concept, les modes de traitement associés, leur adaptation à la résolution de telle classe de problèmes ;
- différencier le savoir que l'enseignant veut transmettre des connaissances effectivement construites par l'élève.

Ensuite, le même auteur va s'appuyer sur la théorie des champs conceptuels de Vergnaud, en partant de l'idée que la notion de conception est l'analogie du côté du sujet du triplet $C = \{S, I, \mathcal{S}\}$ qui décrit un concept. Cette caractérisation est la référence classique dans les travaux de didactique des mathématiques, lorsqu'on parle de conception. Cependant, Artigue elle-même reconnaît que ce modèle n'est pas opératoire, et ce parce qu'il est difficile, voire impossible, d'une part, d'inférer à partir de l'observation de l'élève dans quelques situations, et d'autre part, à cause de la globalité de sa conception sur tel ou tel objet mathématique [24].

Pour terminer, il nous faut donner la dernière évolution de la notion de conception, construite par N. Balacheff dans un objectif qui nécessite de rendre calculables les situations didactiques. Ainsi, Balacheff [25] caractérise une conception C par « un quadruplet (P, R, L, Σ) dans lequel :

- P est un ensemble de problèmes sur lequel C est opératoire ;
- R est un ensemble d'opérateurs ;
- L est un système de représentation, il permet d'exprimer les éléments de P et de R ;
- Σ est une structure de contrôle, elle assure la non-contradiction de C . En particulier, un problème p de P est résolu s'il existe r de R et s de Σ tel que $s(r(p)) = \text{vrai}$ ».

4 Conceptions erronées susceptibles d'être causées par le vocabulaire utilisé en tests statistiques

Une littérature abondante de travaux de recherches considère que les étudiants se donnent à plusieurs conceptions erronées en liaison directe avec les aspects de compréhension des tests statistiques [26]. Nous faisons remarquer que la plupart de ces conceptions erronées sont probablement causées par le vocabulaire véhiculé par les procédures sous-jacentes, comme nous allons le voir dans ce qui suit.

4.1 Conception des tests statistiques comme générateur de résultats mathématiquement démontrés

La plupart des apprenants et des chercheurs expérimentés croient erronément que les résultats décisionnels (rejeter H_0 , accepter H_0 ou échouer à rejeter H_0) auxquels ils aboutissent à travers l'application d'un test statistique, soient mathématiquement démontrés et présentent un caractère déterministe et non probabiliste, absolu et non relatif [1, 3, 9, 18, 27]. Pire encore, ils peuvent même étendre cette idée de démonstration mathématique aux hypothèses et actions scientifiques modélisées par les hypothèses statistiques H_0 et H_1 .

Ceci est effectivement une erreur, du fait que, d'une part, la notion fondamentale sur laquelle se basent les tests statistiques, à savoir la probabilité conditionnelle, ne peut en aucun cas fonder des arguments relatifs à la certitude de la vérité ou de la fausseté des hypothèses H_0 et H_1 énoncées. D'autre part, n'importe quel résultat fourni à travers l'application d'un test statistique -ou de n'importe quelle autre procédure de la boîte à outils de l'inférence statistique - est basé sur un élément de jugement qu'on doit fixer explicitement avant ou après l'expérimentation conduite pour construire des données pertinentes, valides et fiables D_0 . Les exemples les plus usuels de cet élément de jugement sont la valeur prise pour le niveau de signification dans le cas des tests statistiques, et la probabilité *a priori* de l'hypothèse dans le cas de la statistique bayésienne.

De manière générale, nombre d'utilisateurs et de consommateurs de tests statistiques considèrent faussement que tout événement conditionnel des données \mathcal{D} sachant l'hypothèse H ($(\mathcal{D} | H)$), ou tout événement conditionnel des données \mathcal{D}' sachant d'autres données \mathcal{D} ($(\mathcal{D}' | \mathcal{D})$), auxquels on pourrait affecter respectivement la probabilité $P(H | \mathcal{D})$ et la probabilité $P(\mathcal{D}' | \mathcal{D})$ déterminées à partir de la probabilité $P(\mathcal{D} | H)$ autorisée par les tests statistiques, deviendront des événements certains ou incertains, et ce selon la conclusion (ou la décision) à laquelle on aboutira lors de l'étape de formulation du jugement, à savoir rejeter H_0 ou échouer à rejeter H_0 dans le cas des tests de signification, et rejeter H_0 ou accepter H_0 dans le cas des tests d'hypothèse. L'événement « *le résultat du test de signification est dû au seul hasard* » est un exemple significatif de cette erreur, dans la mesure où sa probabilité est considérée comme égale à la p-valeur, et deviendra sur la base d'un raisonnement erroné, un événement certain dans le cas où le résultat est jugé statistiquement significatif.

4.2 Conception de la signification statistique comme équivalente à la signification pratique**

Kirk [28] a défini la différence entre la signification statistique, fournie par application d'un test statistique, et la signification pratique, de la façon assez simple suivante :

«Statistical significance is concerned with whether a research result is due to chance or sampling variability; practical significance is concerned with whether the result is useful in the real world» [28, p. 746].

Par ailleurs, la controverse majeure sur l'interprétation des tests statistiques avait pour origine la supposition naïve selon laquelle un résultat statistiquement significatif est nécessairement un résultat remarquable [29]. Dès 1931, Tyler [30] avait commencé à reconnaître l'étendue de la mauvaise interprétation de la signification statistique, en déclarant que les interprétations généralement formulées à partir des études de l'époque indiquent clairement que nous sommes contraints de concevoir la signification statistique comme équivalente à la signification sociale. Il ajoutait que ces deux termes sont essentiellement différents et ne devraient pas être confondus : les différences qui sont statistiquement significatives ne sont pas toujours socialement importantes ; le contraire est également vrai : les différences qui ne sont pas avérées être statistiquement significatives, peuvent cependant être socialement significatives.

De manière similaire, Schafer [31] a noté que quelques chercheurs ne comprennent pas que statistiquement *significatif* et *important* sont deux choses différentes. Le terme *significatif* était sans doute mal choisi. De plus, Meehl [32] a récemment caractérisé l'utilisation du terme « *significatif* » comme étant « *cancéreux* » et « *trompeur* » et il a recommandé que les chercheurs interprètent les résultats en utilisant les intervalles de confiance au lieu des p-valeurs impliquées par les tests de signification. Certes, le résultat du

** La signification pratique est parfois appelée : « signification théorique », « signification substantielle », « signification sociale », « signification scientifique », etc.

jugement de la significativité statistique fourni par application d'un test statistique, à savoir: « Rejeter H_0 », « Accepter H_0 » ou « Echouer à rejeter H_0 », présente deux aspects favorisant le *non déterminisme* : d'une part, la région critique RC dans le cas d'un test d'hypothèse et la p-valeur dans le cas d'un test de signification, sont déterminées par l'intermédiaire d'un calcul probabiliste, et d'autre part, le statisticien utilise son jugement (la valeur prise pour le niveau de signification) pour formuler le résultat.

4.3 Conception de la signification statistique comme valeur scientifique d'une recherche ou d'une décision

Eysenck [33] a argumenté que le mot significatif est faussement utilisé comme un *schibboleth*^{††} qui différencie une recherche réussie d'une recherche non réussie, dans la mesure où le concept de p-valeur, sur lequel se basent les procédures de tests statistiques, est parfois considéré comme étant une mesure de la valeur scientifique d'une recherche : un résultat statistiquement significatif est meilleur que celui qui est statistiquement non significatif.

En fait, la signification statistique n'est pas une mesure de la valeur d'une recherche ou d'une décision. Un résultat (ou une décision) statistiquement non significatif n'est pas inférieur (en termes de valeur scientifique) à celui qui est statistiquement significatif, et ce pour plusieurs raisons. Premièrement, un test statistique permet tout simplement de choisir entre deux *modèles* [22]. Deuxièmement, il est vrai que l'hypothèse alternative H_1 est souvent celle qui est attendue pour être prouvée par le *chercheur ou le décideur*, mais il faut, pour ceci, se baser sur des bases scientifiques solides. Toutefois, la signification statistique devrait être considérée comme étant une mesure de satisfaction, et non un seuil ou une importance. Finalement, dans la plupart des cas, chaque test statistique, utilisé dans une recherche donnée, est très limité dans son rôle.

Il est par conséquent plus important d'identifier la valeur scientifique d'une recherche ou d'une décision d'abord en fonction d'un plan expérimental adéquat, susceptible de minimiser le bruit de fond causé par les variables parasites, et ensuite d'une sélection soignée de l'échantillon et d'un protocole plus précis du processus de mesure.

5 Etude exploratoire des conceptions auprès des étudiants

Nous détaillerons, dans ce qui suit, les étapes fondamentales sur lesquelles nous nous sommes basés pour mettre en place la partie expérimentale de ce travail.

5.1 Objectif de l'étude et construction des données par enquête par questionnaire

Nous avons conduit à la fin du deuxième semestre de l'année scolaire 2017/2018 une expérimentation auprès de 195 étudiants relevant de huit universités marocaines qui poursuivent leurs études en année de licences scientifiques (Mathématiques, Physique, Chimie, Génie électrique, Informatique, Enseignement) et économiques. Ces participants, âgés entre 22 ans et 25 ans, dont 95 sont de sexe féminin et 100 de sexe masculin, sont titulaires de baccalauréats de cinq types, à savoir : les Mathématiques, la Physique-Chimie, la Vie et Terre, les Sciences Techniques et l'Economie. De plus, 57% de ces participants avaient déjà suivi un cours sur les tests statistiques pendant leurs séjours à l'université. A ces participants de l'échantillon, nous avons administré pendant une durée d'une demi-heure un

^{††} Epreuve qui permet d'évaluer quelque chose de manière totale et décisive.

questionnaire composé de quatre questions (voir annexe). Chaque question traite les significations susceptibles d'être données à un terme usuellement utilisé relativement aux tests statistiques. Aussi, chaque question est composée d'un ensemble d'items à choix double. Les items de chaque question tournent autour des différentes significations des termes en relation avec la question correspondante. Les modalités de chaque item sont quelques significations possibles (*justes ou fausses*) qu'on pourrait imaginer.

Notons au passage que le choix de ce type d'échantillons d'étudiants dans notre travail est basé sur plusieurs raisons dont nous pouvons citer à titre d'exemples :

- Le fait que ces étudiants ont subi une formation scientifique durant leurs séjours à l'université et aux études au cycle du secondaire. C'est pour cela, il est probable que leurs conceptions relatives à ce qu'ils apprennent à l'école soient toujours à caractère absolu et non relatif.
- Il est probable que ces étudiants soient les plus qualifiés à se tromper par le vocabulaire véhiculé, surtout les problèmes susceptibles d'être impliqués par la traduction de la langue arabe à une autre langue étrangère, étant donné que l'enseignement des matières scientifiques au Maroc se fait en langue arabe jusqu'au baccalauréat et en langue française à l'université.

Nous avons signalé aux participants de lire attentivement chaque item avant de choisir la modalité qui leur semble la plus convenable, tout en leur demandant de justifier brièvement (et sans faire de calcul) à chaque fois leurs réponses. Enfin, nous leur avons aussi indiqué qu'ils peuvent répondre de *manière indépendante* aux différents items de chaque question. Il est important d'insister sur le fait que nous n'avons pas communiqué aux participants l'objectif du questionnaire. D'ailleurs, ce dernier est construit de telles sortes que les termes dont nous demandons leurs significations ne semblent pas relever du vocabulaire utilisé en tests statistiques.

5.2 Présentation et analyse *a priori* du questionnaire

Avant de procéder à l'analyse proprement dite des productions des étudiants, nous allons consacrer la prochaine section à l'analyse *a priori* du questionnaire, afin de mieux cerner le traitement statistique des données construites avec le questionnaire. Nous faisons remarquer que pour chaque item des différentes questions, la modalité considérée correcte (*selon la logique habituelle ou le contexte des tests statistiques*) est celle présentée cochée. Vu la nature de l'objectif fixé par notre présente recherche, qui vise, en quelque sorte, l'effet impliqué par le *vocabulaire* sur la signification qu'on pourrait éventuellement donner aux concepts utilisés en terme de procédures de tests statistiques, notre analyse se focalisera surtout sur les deux éléments essentiels suivants : *a*) la justification *scientifique* proprement dite de la fausseté ou de la véracité de chaque modalité en la liant éventuellement aux erreurs ou aux conceptions erronées usuelles en matière des tests statistiques, et ce étant donné que ce présent travail s'adresse aux utilisateurs expérimentés et potentiels de cet outil statistique ; *b*) la justification de la formulation par laquelle l'item d'une question donnée est construit et le terme (ou les termes) trompeur utilisé dans cette construction forçant un sujet qui n'a jamais suivi de cours sur les tests statistiques (ou même celui qui en a déjà suivi) à établir une correspondance erronée avec le terme ou l'expression objet de la question.

Notons aussi au passage que la mise en place d'un tel questionnaire sous sa forme définitive a été faite à travers l'accomplissement des deux tâches essentielles suivantes qui lui a donné sa légitimité d'être utilisé par notre recherche :

- Garantir la validité de cet instrument de mesure, en l'ayant présenté à 10 experts dans le domaine de l'enseignement de la statistique. Ainsi, nous avons demandé à ces experts

de donner leurs remarques et leurs suggestions concernant les éléments du questionnaire et ce dans l'objectif est de s'assurer de sa validité globale relativement à : la formulation linguistique des items, la concordance entre les items et les objectifs de la recherche, l'ajout ou la suppression de ce qu'ils voient convenable.

- Garantir la fiabilité des différents items de cet instrument de mesure en l'administrant pour une première fois à un échantillon exploratoire d'étudiants. A partir des résultats obtenus, nous avons calculé les différentes valeurs correspondantes au coefficient Alpha^{††} de Cronbach que nous avons trouvées variant entre 0,78 et 0,89 pour les différents items, et égale à 0,84 pour le questionnaire tout entier.

Question 1 : *Si l'on vous demande d'imaginer ce que vous pouvez comprendre par (ou ce que peut signifier pour vous) l'expression « Hypothèse Nulle », il surviendra à votre esprit :*

- | | | |
|---|--------------------------------------|--------------------------------------|
| A. Une hypothèse dont sa véracité sera sans valeur et n'aura aucune importance dans la vie sociétale. | <input type="radio"/> Oui | <input checked="" type="radio"/> Non |
| B. Une hypothèse dont sa véracité ne conduira à l'adoption d'aucune action nouvelle. | <input type="radio"/> Oui | <input checked="" type="radio"/> Non |
| C. Une hypothèse prévue pour être falsifiée ou rejetée, afin de gagner un support susceptible de permettre un progrès scientifique. | <input checked="" type="radio"/> Oui | <input type="radio"/> Non |
| D. Une hypothèse déjà connue d'être fautive avant même sa mise à l'épreuve. | <input type="radio"/> Oui | <input checked="" type="radio"/> Non |
| E. Une hypothèse exprimée sous forme d'une égalité ou d'une différence nulle. | <input type="radio"/> Oui | <input checked="" type="radio"/> Non |

Analyse *a priori* des items de la question 1

L'« hypothèse nulle » est spécialement réquisitionnée dans la théorie des tests de signification de Fisher. Dans cette théorie, une telle hypothèse est avancée pour être rejetée, afin d'acquiescer un support d'acceptabilité de l'hypothèse scientifique (ou de recherche) qui intéresse le chercheur. Une telle hypothèse est souvent considérée par erreur comme une hypothèse dont la véracité ne conduit à l'adoption d'aucune action nouvelle, en la confondant ainsi avec l'hypothèse H_0 de la théorie de Neyman-Pearson, incompatible avec la théorie de Fisher.

Notons au passage que l'hypothèse nulle est appelée ainsi, dans le sens de « *to be nullified* » ; c'est-à-dire à réfuter, et non nécessairement, comme on le trouve écrit dans certains manuels, d'une valeur de zéro pour le paramètre testé, même si c'est le cas le plus fréquent [34]. En général l'hypothèse nulle sera la négation de l'hypothèse à laquelle le chercheur s'intéresse réellement :

«... *the hypothesis that the phenomenon to be demonstrated is in fact absent*» [35].

De plus, avant sa mise à l'épreuve, l'hypothèse nulle devrait normalement être consistante avec l'état actuel de la connaissance scientifique. Malheureusement, plusieurs articles et chercheurs identifient dans leurs recherches l'hypothèse nulle comme étant une hypothèse « *prête-nom* »^{§§}, c'est-à-dire une hypothèse qui est connue pour être fautive avant même que des données soient construites – et elle est souvent utilisée de cette façon. Par ailleurs, rejeter une hypothèse nulle connue pour être fautive avant qu'on construise les données ne fournit pas de nouvelle information scientifique : on a simplement démontré qu'une déclaration déjà connue pour fautive l'est effectivement. Mais si l'hypothèse nulle est consistante avec l'état

^{††} C'est un indicateur qui sert à mesurer la cohérence interne (ou la fiabilité) des questions posées lors d'un test (les réponses aux questions portant sur le même sujet devant être corrélées). Sa valeur est inférieure ou égale à 1, étant généralement considérée comme « acceptable » à partir de 0,7.

^{§§} « *Straw person* » en langue anglaise.

actuel de la connaissance, alors la rejeter nous informera que nous avons trouvé quelque chose de nouveau, permettant à l'hypothèse de recherche (hypothèse nouvelle qui présente de l'intérêt) *d'avoir des chances* d'être acceptée [36].

Pareillement, bien que le résultat d'un test de signification soit « *rejeter l'hypothèse nulle* » ou « *échouer à rejeter l'hypothèse nulle* », il est possible dans ce dernier cas d'approfondir la recherche pour étudier la possibilité de résistance d'une telle hypothèse, qui restera encore valide jusqu'à nouvel ordre. Par ailleurs, il y a trois critères pour accepter l'hypothèse nulle, à savoir : (1) l'hypothèse nulle devrait être possible, (2) les résultats devraient être consistants avec l'hypothèse nulle (ou une p-valeur faible), et (3) un bon effort devrait être fait pour chercher un effet (c'est-à-dire pour rejeter l'hypothèse nulle). Selon Frick [37], dire qu'un bon effort est fourni dans une expérimentation donnée est fonction du degré de satisfaction des trois critères suivants : (a) l'adoption d'un plan expérimental adéquat, (b) l'obtention d'intervalles de confiances très larges, pour le paramètre testé, et (c) l'existence (ou non) d'un effet (lié) dans une expérimentation similaire, utilisant le même plan expérimental que à celle pour laquelle il reste à trouver un effet.

Il est vrai que l'hypothèse alternative H_1 est souvent celle que le *chercheur* (ou le *décideur*) souhaite prouver, mais il ne faut pas voir une absence d'information dans un résultat négatif. Nous savons qu'au rejet de l'hypothèse nulle H_0 , impliquant la signification statistique, est attaché un grand prix, parce que ce rejet est vu comme conduisant à la théorie en développement, tandis que l'échec du rejet de H_0 , ou même son acceptation, pourraient conduire à une impasse. La signification statistique devrait en revanche être considérée comme une mesure de satisfaction. Il est, par conséquent, plus utile d'attribuer à une recherche (ou une décision) une valeur qui est fonction de la qualité du plan expérimental adopté, dans la mesure où ce dernier réduit autant que possible les fluctuations d'échantillonnages et les erreurs de mesures expérimentales.

On peut remarquer que les différents items de cette question sont volontairement formulés à l'aide d'expressions traduisant une certaine nullité, qui peut être directement liée au terme « nulle », indiqué dans l'expression « hypothèse nulle ». On peut remarquer, par exemple, les expressions : « *sans valeur* », « *aucune importance* », « *ne conduira pas* », « *fausse* », « *égalité* », « *différence nulle* ». Cela permettra d'étudier si ce terme constitue vraiment un obstacle pour donner une signification exacte à l'expression « hypothèse nulle ». Il est vrai que l'expression indiquée dans l'item E est celle la plus habituellement enseignée aux étudiants, mais cela n'empêche pas d'étudier la possibilité d'existence des autres expressions signalées dans les autres items.

Question 2 : « *Prouver la Vérité d'une Hypothèse* » signifie pour vous :

- A. Présenter des arguments aussi solides que possible en faveur de la véracité de cette hypothèse, tout en fixant un seuil exigé et réfléchi d'évidence, et en exposant la force de preuve ou la puissance de votre raisonnement. Oui Non
- B. Présenter des arguments absolument convainquant comme dans le cas d'une démonstration purement mathématique envers la véracité de cette hypothèse. Oui Non

Question 3 : L'expression « *Hypothèse Vraie* » signifie pour vous :

- A. Une telle hypothèse est un modèle reflétant au mieux le monde réel, permettant à la plupart des gens rationnels de l'accepter comme modèle réalisable Oui Non
- B. La véracité d'une telle hypothèse est seulement une affaire de jugement, dans la mesure où un individu peut considérer cette hypothèse vraie alors que son voisin le plus proche peut décider le contraire. Oui Non

- C. On a présenté des arguments absolument convainquant envers la véracité de cette hypothèse, comme dans le cas d'une démonstration purement mathématique, et que cette véracité dure pour toujours. O Oui X Non
- D. Une telle hypothèse est un modèle dans lequel on tolère qu'il puisse exister quelques peu de cas pour lesquels cette hypothèse ne tient pas, mais sans pour autant l'étiqueter comme étant non valide dans le monde réel. X Oui O Non

Analyse *a priori* des items des questions 2 et 3

Remarquons que l'analyse des deux expressions (prouver et vrai) objets de ces deux questions avait déjà commencé dans l'introduction de ce travail. Néanmoins, nous allons approfondir une telle analyse en présentant sur ce sujet quelques autres éléments de valeur. De plus, contrairement à l'expression « *hypothèse vraie* » qui est explicitement utilisée dans le langage usuel des tests statistiques, l'expression « *Prouver la Véracité d'une Hypothèse* » est implicitement emmagasinée dans l'esprit de l'utilisateur de cet outil qui sollicite les tests statistiques afin de lui permettre de prouver quelque chose. Signalons aussi que c'est volontairement que nous avons situé ces deux questions dans le même panier d'analyse. La raison en est que les deux termes sous-jacents sont étroitement liés, dans la mesure où *la véracité d'une hypothèse* ne sera établie qu'en s'appuyant sur des éléments de *preuve* convaincants. De plus, les différents items de ces deux questions sont tous formulés avec des termes et des expressions induisant soit un caractère relatif considéré comme valide dans le contexte des tests statistiques et suivant la logique de formulation d'une conclusion relativement à un niveau de signification, soit un caractère absolu.

Il est à signaler qu'en termes de procédures de tests statistiques, il est préférable d'utiliser le terme « *modèle* » plutôt qu'hypothèse, si l'on veut souligner qu'en tests statistiques, on comparera des modèles et non des hypothèses. En effet, des hypothèses (statistiques) sont des modélisations d'*hypothèses scientifiques* plus abstraites, reflétant le monde réel. Ainsi, accepter, par exemple, l'hypothèse nulle signifie que *le modèle nul* pourrait donner de meilleures prédictions. Un autre point que nous souhaitons soulever à propos des tests statistiques est que dans un projet de recherche, un test statistique est utilisé seulement pour *aider à décider* en s'aidant de l'analyse statistique. A cela s'ajoute le fait qu'il s'agit seulement d'un outil de la boîte des statistiques, parmi plusieurs qui peuvent être utilisés. Cela permet de conclure que son usage est limité par définition.

Il est vrai que les données construites soutiennent toujours le modèle alternatif. Si ce n'était pas le cas, aucun test statistique choisissant automatiquement le modèle nul ne devrait être utile ou nécessaire. Un test de signification, par exemple, est utilisé dans le but de falsifier le modèle nul en faveur de l'acceptabilité du modèle alternatif. L'indicateur de mesure généralement utilisé, est la *p-valeur* qui est égale à « *la probabilité d'obtenir le résultat de l'échantillon (ou même plus extrême) en supposant que le modèle nul est vrai* ». Cette approche en termes de modèles des tests statistiques est plus longuement discutée par Mclean [22].

Ainsi, l'usage de la terminologie telle que « *la valeur vraie de la moyenne* » conduit à des déclarations comme « *l'hypothèse nulle est presque toujours fausse* ». Bien que ceci puisse être vrai, il est hors de propos. Un test statistique n'est pas directement concerné par la valeur vraie d'un paramètre, mais par les modèles escomptés. Accepter le modèle nul signifie seulement que l'usage de ce modèle, fondé sur la valeur spécifiée du paramètre, est avantageux (valable) ; le rejeter signifie qu'il est préférable d'utiliser un modèle fondé sur une valeur différente.

Cependant, juger de la qualité de la preuve pour conclure quant au rejet du modèle nul ou à l'échec de son rejet est une tâche subjective. Quelqu'un pourrait considérer une déclaration

comme étant prouvée, alors que son proche voisin penserait le contraire. Si la preuve statistique est suffisamment forte, le lecteur pourra accepter le résultat comme « prouvé » dans un sens faible. Cet argument s'applique en particulier aux tests statistiques.

Par ailleurs, dans une situation de recherche, la décision est provisoire dans la mesure où il faut toujours avoir présent à l'esprit que l'acceptation ou le rejet d'une hypothèse peuvent ne pas durer longtemps. Les recherches qui seront menées dans le futur pourront contredire les conclusions d'aujourd'hui.

Question 4 : Si l'on vous demande ce que peut signifier pour vous l'expression « Résultat Significatif », vous penserez à un résultat qui :

- A. Est forcément remarquable et important dans la vie sociétale Oui Non
- B. Peut ne pas être ni remarquable, ni important dans la vie sociétale. Oui Non
- C. N'est pas dû au hasard, dans la mesure où il existe une cause bien connue et bien définie qui l'a suscité. Oui Non
- D. Peut-être remarquable et important dans la vie sociétale. Oui Non

 Analyse a priori des items de la question 4

L'expression « résultat significatif » est très habituellement utilisée dans les procédures de tests statistiques. Elle est utilisée pour signifier qu'un test statistique a permis de rejeter l'hypothèse nulle. Evidemment, cette expression est utilisée à tort pour remplacer l'expression « résultat statistiquement significatif », dans la mesure où l'adverbe « statistiquement » y est introduit pour mettre en évidence la différence qui pourrait exister entre une « signification statistique » fournie par application d'un test statistique, et une « signification pratique » (cf. § 4.2).

Certes, le résultat (existence ou non de la signification statistique) fourni par application d'un test statistique, à savoir : « Rejeter H_0 », « Accepter H_0 » ou « Echouer à rejeter H_0 », possède deux aspects favorisant le non déterminisme : D'une part, la région critique RC, dans le cas d'un test d'hypothèse, et la p-valeur, dans le cas d'un test de signification, sont déterminées par un calcul probabiliste, et d'autre part, le statisticien utilise son jugement (la valeur α , prise pour le niveau de signification) pour déterminer un tel résultat. Par conséquent, l'omission, dans cette question, de l'adverbe statistiquement dans l'expression « résultat statistiquement significatif » et la formulation des items sous-jacents pourraient conduire non seulement un apprenant qui n'a jamais suivi de cours sur les tests statistiques, mais aussi un utilisateur expérimenté de l'outil statistique, à choisir des modalités erronées des différents items de cette question.

Le tableau 1 ci-dessous représente un récapitulatif des éléments essentiels, caractéristiques des différents items relatifs aux quatre questions du questionnaire.

Tableau 1. Récapitulatif des éléments caractéristiques des différents items.

| Question et « Concept implicite » | Terme trompeur du concept | Signification ^{†††} du concept | Items | Eléments d'ambiguïté du vocabulaire, impliqués par l'item |
|-----------------------------------|---------------------------|--|-------|---|
| Question n° 1 : « hypothèse » | « Null » | « Hypothèse prévue pour être falsifiée » | A | « sans valeur », « aucune importance » |

*** Il s'agit du concept implicite dans une procédure de tests statistiques.

††† Il s'agit de la signification (vraie) du concept dans le vocabulaire usuel adopté par les procédures de tests statistiques.

| | | | | |
|--|------------------|--|---|--|
| | | | B | « ne conduira pas », « aucune action nouvelle » |
| | | | C | - |
| | | | D | « connue d'être fausse » |
| | | | E | « égalité », « différence nulle » |
| Question 2 : « Prouver la véracité d'une hypothèse » | « Prouver » | « Prouver de manière relative et non absolue » | A | « arguments aussi solides que possible », « fixation d'un seuil exigé d'évidence » |
| | | | B | « arguments absolument convainquant », « démonstration purement mathématique » |
| Question 3 : « Hypothèse vraie » | « Traite » | « Vérité à caractère relatif et non absolu » | A | « modèle reflétant au mieux le monde réel », « la plupart des gens l'acceptent » |
| | | | B | « affaire de jugement de l'individu » |
| | | | C | « arguments absolument convainquant », « démonstration purement mathématique » |
| | | | D | « modèle », « on tolère » |
| Question 4 : « Résultat significatif ^{***} » | « Significatif » | « Le résultat fourni par un test statistique pourrait être pratiquement ^{§§§} significatif, comme il se peut qu'il ne pourrait pas l'être » | A | « forcément remarquable et important » |
| | | | B | « peut ne pas être ni remarquable, ni important » |
| | | | C | « n'est pas dû au hasard », « une cause bien définie l'a suscité » |
| | | | D | « peut-être remarquable et important » |

5.3 Codage et outil d'analyse retenus pour le traitement des réponses des étudiants

Pour le traitement des réponses des étudiants, l'analyse factorielle des correspondances multiples (AFCM), nous a semblé être un outil efficace pouvant tenir compte de ce fait, puisqu'il va permettre d'analyser et d'interpréter les réponses des étudiants en termes de liens qui existent entre les différentes modalités des différents items du questionnaire. Dans le tableau 2 suivant, nous résumons, les différents questions et items du questionnaire, ainsi que les significations et les codages des modalités associées. Ce tableau est, en quelque sorte, un complément de l'analyse *a priori* dont nous avons commencé à présenter l'essentiel dans la section 5.2 de cet article.

^{***} L'expression « Résultat significatif » est souvent utilisée à la place de l'expression « Résultat statistiquement significatif ».

^{§§§} L'adverbe « pratiquement » peut être remplacé par : « socialement », « théoriquement », « scientifiquement », « substantiellement », ...

Tableau 2. Codage des modalités des différents items des quatre questions du questionnaire.

| Question | Items | Modalité correcte | Codes**** |
|-------------------|-------|-------------------|-----------|
| <i>Question 1</i> | A | Non | 1Ar/1Ae |
| | B | Non | 1Br/1Be |
| | C | Oui | 1Cr/1Ce |
| | D | Non | 1Dr/1De |
| | E | Non | 1Er/1Ee |
| <i>Question 2</i> | A | Oui | 2Ar/2Ae |
| | B | Non | 2Br/2Be |

| Question | Items | Modalité correcte | Codes |
|-------------------|-------|-------------------|---------|
| <i>Question 3</i> | A | Oui | 3Ar/3Ae |
| | B | Oui | 3Br/3Be |
| | C | Non | 3Cr/3Ce |
| | D | Oui | 3Dr/3De |
| <i>Question 4</i> | A | Non | 4Ar/4Ae |
| | B | Oui | 4Br/4Be |
| | C | Non | 4Cr/4Ce |
| | D | Oui | 4Dr/4De |

5.4 Analyse a posteriori et interprétation des réponses des étudiants

Rappelons que nous avons choisi l’AFCM pour analyser les données de cette étude. Nous présentons, dans ce qui suit, les éléments nécessaires à l’application de cette méthode multidimensionnelle ainsi que les résultats d’analyse auxquels nous sommes arrivés à travers son application.

5.4.1 Valeurs propres et inertie totale

L’AFCM**** appliquée au tableau disjonctif complet issu du codage des réponses des étudiants, a conduit à des valeurs propres non nulles de moyenne 0,07 (l’inverse du nombre d’items qui est égal à 15). Par ailleurs, les valeurs propres supérieures à cette moyenne sont présentées dans le tableau 3 suivant :

Tableau 3. Tableau des valeurs propres.

| λ_1 | λ_2 | λ_3 | λ_4 | λ_5 | λ_6 | λ_7 |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0,135 | 0,114 | 0,108 | 0,093 | 0,082 | 0,076 | 0,070 |

L’inertie totale est égale à 1 (nombre de valeurs propres non nulles, divisé par le nombre d’items).

5.4.2 Nombre d’axes retenus dans l’analyse

Nous rappelons tout d’abord que l’inertie totale en AFCM appliquée à un tableau disjonctif complet n’a pas de signification statistique**** ; elle ne dépend pas des observations, elle dépend uniquement du nombre de variables et du nombre de modalités impliquées. Par ailleurs, le calcul des taux d’inertie liés aux sept valeurs propres ci-dessus conduit à des pourcentages décroissant de 13,5% à pratiquement 1,7%, ce qui donne une idée assez pessimiste des parts d’informations obtenues par l’AFCM. Nous avons donc eu recours à la formule proposée par Benzecri [38], qui dans une telle situation, va permettre une meilleure appréciation des taux d’inertie. L’application de cette formule aux inerties initiales, ont donné lieu aux pourcentages d’inerties corrigées et leurs cumuls, illustrés dans le tableau 4 suivant:

**** Le codage adopté pour les modalités est de la forme : « nXy », où n : numéro de la question de la modalité, X : lettre alphabétique désignant l’item de la modalité, et y : lettre alphabétique désignant la modalité de l’item X appartenant à la question numéro n. Les modalités de types « nXr » ou « nXe » (c’est-à-dire : y= r ou y= e) désignent respectivement la réussite ou l’échec à l’item X de la question numéro n.

**** Cette analyse a été conduite à l’aide du logiciel Statistica (Version 6).

**** Ce n’est pas le cas pour le calcul d’inertie en analyse de correspondances simples d’un tableau de Burt.

Tableau 4. Pourcentages d'inerties corrigées (benzécriennes) et leurs cumuls.

| Valeur propre | λ_1 | λ_2 | λ_3 | λ_4 | λ_5 | λ_6 | λ_7 |
|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| %Inertie corrigée | 48 | 23 | 18 | 7 | 2 | 1 | ≈ 0 |
| Cumul % Inertie corrigée | 48 | 71 | 89 | 96 | 98 | 99 | 100 |

Par conséquent, le premier axe représente 48% de l'information totale, le deuxième axe 23%, le troisième axe 18% et le quatrième axe 7% (les axes suivants chutent à presque une part d'information représentant moins de la moitié du quatrième axe). Ces valeurs indiquent que l'essentiel de l'information est lié aux quatre premiers axes factoriels qui représentent ensemble 96% de l'information totale.

Par ailleurs, ces quatre valeurs propres sont différentes ; cela signifie que nous sommes pratiquement en présence de quatre valeurs propres simples. Par conséquent, nous allons interpréter séparément les axes : 1, 2 et 3.

Avant d'entamer une telle étude, nous allons consacrer la prochaine section à la présentation de quelques éléments d'analyse globale que nous considérons essentiels, et que nous avons pu déduire en mettant en œuvre les différents plans et axes factoriels retenus.

5.4.3 Eléments d'analyse globale

De prime abord, en nous basant sur des éléments élémentaires de la statistique descriptive, nous voudrions tout simplement signaler que *le vocabulaire* adopté par les procédures de tests statistiques est nettement ouvert à plusieurs interprétations (fallacieuses) par nos participants. Ceci est déduit des pourcentages élevés des réponses erronées de ces derniers, aux différents items du questionnaire. Nous rappelons au passage que ces réponses erronées sont repérées à la fois chez les participants qui avaient déjà suivi un cours sur les tests statistiques que chez ceux qui n'en avaient jamais suivi.

De plus, il est évident que bien que les items du questionnaire fassent référence à une exploration conceptuelle autour des tests statistiques, et non pas aux performances des étudiants en situation de résolutions de problèmes sur cet outil, la plus grande part de l'information (1^{er} plan factoriel) contenue dans les réponses des étudiants se traduit presque en termes de performances des étudiants, autrement dit en termes de réussites-échecs. Notons aussi au passage, que dans les oppositions de groupes de modalités à fortes contributions relatives aux axes factoriels retenus pour l'analyse, nous avons une opposition nette de modalités contraires, ce qui confère au questionnaire une certaine homogénéité.

Par ailleurs, nous aborderons, dans les trois sections ci-après, l'interprétation de chacun de ces trois axes factoriels, en essayant de manière générale, de traduire les différentes proximités (*locales et globales*) entre les modalités à fortes contributions relatives, remarquées dans chacun des deux côtés opposés de tels axes, ainsi que les différentes oppositions entre telles modalités. Nous nous basons dans notre analyse sur les deux principes usuels, (a) et (b) ci-après, de l'AFCM : (a) la proximité entre modalités d'items différents est exprimée en terme d'association : ces modalités correspondent aux points moyens des individus qui les ont choisies et sont proches parce qu'elles concernent globalement les mêmes individus ou des individus semblables ; (b) la proximité entre deux modalités d'un même item est exprimée en terme de ressemblance : par construction, les modalités d'un même item s'excluent. Si elles sont proches, cette proximité s'interprétera en termes de ressemblance entre les groupes d'individus qui les ont choisies (vis-à-vis d'autres variables actives de l'analyse).

5.4.4 Interprétation du 1^{er} axe factoriel

L’AFCM conduit à une inertie de 48% pour l’axe 1. Les modalités qui contribuent le plus à la construction de cet axe sont celles entourées dans la figure 1 ci-dessous^{§§§§}. Elles ont une contribution comprise entre 1,7% et 16,1%. La qualité de représentation de ces modalités par rapport à un tel axe est comprise entre 0,092 et 0,550.

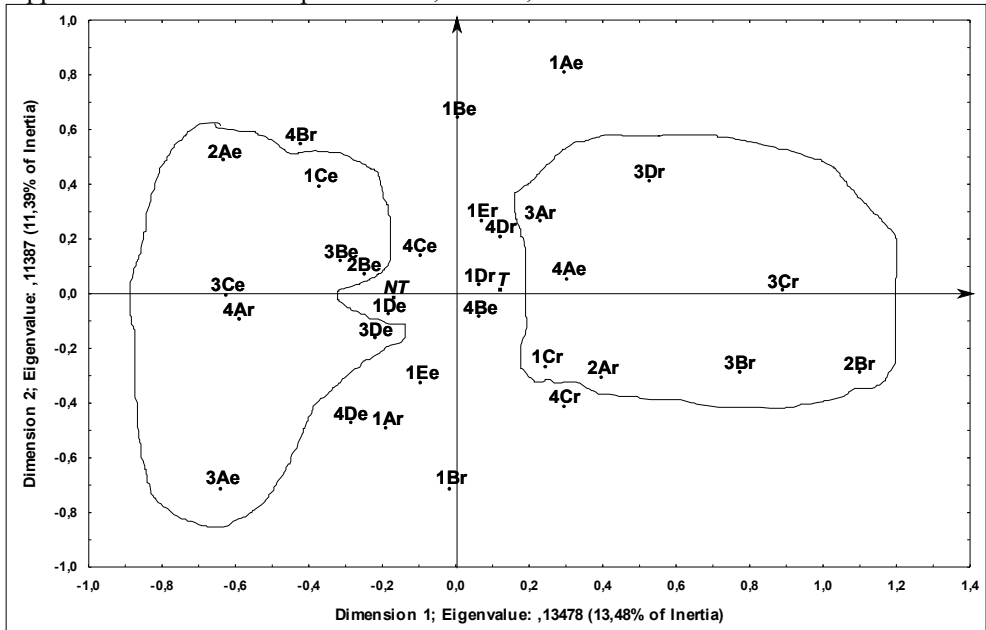


Fig. 1. Plan factoriel (1, 2).

Dans ce premier plan factoriel, le 1^{er} axe présente l’opposition fournie par le tableau 5 suivant :

Tableau 5. Positionnements des modalités sur l’axe 1.

| A gauche de l’axe 1 | A droite de l’axe 1 |
|--|--|
| 3Ae, 2Ae, 3Ce, 4Ar, 1Ce, 3Be, 2Be, 3De | 3Ar, 1Cr, 4Ae, 2Ar, 3Dr, 3Br, 3Cr, 2Br |

Si nous ne nous tenions pas compte des deux modalités 4Ar et 4Ae dans cette opposition, nous pourrions remarquer que, d’un côté, nous avons des modalités d’échecs et, de l’autre, leurs correspondantes de réussites, il suffira alors de donner une interprétation à l’un des groupes de modalités opposées, du fait que l’autre groupe aura tout simplement l’interprétation contraire de ce dernier. Nous essayons, dans le paragraphe suivant, d’accomplir soigneusement une telle interprétation.

Ainsi, les différentes modalités de réussites, se trouvant du côté droit du 1^{er} axe factoriel, et ayant de fortes contributions relatives dans la construction de cet axe, traduisent une maîtrise du vocabulaire utilisé tout au long de la procédure sous-jacente aux tests de signification (i.e., aux tests statistiques selon le paradigme de Fisher). En effet, la définition adéquate qu’il faut donner à l’hypothèse nulle, concept qui représente un point de départ pour une telle procédure, est assurée par la modalité 1Cr. Cette dernière avance qu’une telle

^{§§§§} Notons au passage que ces modalités occupent des positions correspondant aux modalités ayant les plus grandes coordonnées en valeur absolue par rapport au 1^{er} axe factoriel. Sinon, le reste des modalités occupent quasiment des positions intermédiaires, avec en outre de faibles contributions relatives.

hypothèse (nulle) est instaurée pour la falsifier ou la rejeter afin de gagner un support susceptible de permettre un progrès scientifique.

Pareillement, *le caractère relatif* de la notion de *preuve*, qui fait partie du vocabulaire implicitement utilisé pour aboutir à la phase de conclusion d'une telle procédure, est bien appréhendé par les étudiants. Une telle preuve signifie la présentation d'arguments aussi solides que possibles tout en fixant un seuil exigé et réfléchi d'évidence, comme le traduit la modalité 2Ar. Elle ne signifie pas la présentation d'arguments absolument convainquant comme dans le cas d'une démonstration purement mathématique, comme l'exprime la modalité 2Br. La maîtrise du vocabulaire utilisé en procédure de tests de signification est atteinte par la compréhension de *la nature relative* du résultat (ou conclusion) auquel on aboutira dans la phase finale d'une telle procédure.

Une telle maîtrise est justifiée par la réussite des étudiants à tous les items de la question 3. Rappelons qu'un tel résultat ne concerne que la véracité (ou la fausseté) d'une hypothèse parmi celles impliquées dans une procédure de tests de signification. Ainsi, l'expression « *hypothèse vraie* », par exemple, adoptée par le vocabulaire des tests de signification, a pris un ensemble de significations vraies qui sont suscitées par ces différents items. Elles sont toutes à caractère relatif : « *c'est un modèle reflétant au mieux le monde réel* » (3Ar), « *la véracité d'une hypothèse est une affaire de jugement de l'individu qui l'évalue* » (3Br), « *la véracité de l'hypothèse n'est pas purement mathématique et ne dure pas pour toujours* » (3Cr), et enfin, « *il peut exister quelques peu de cas pour lesquels cette hypothèse ne tient pas, sans l'étiqueter comme étant fausse* » (3Dr).

De plus, « le résultat est significatif » n'est qu'une expression souvent utilisée pour exprimer autrement un cas possible de résultat auquel on aboutirait dans l'étape finale de la procédure sous-jacente à un test de signification. Ce cas de résultat est, en quelque sorte, équivalent à : « *l'hypothèse nulle n'est pas vraie* » ou « *l'hypothèse nulle est rejetée* ». Cependant, bien que les étudiants réussissent à répondre aux différents items constituant une telle procédure comme nous avons indiqué dans le paragraphe ci-dessus, ils échouent, comme le montre la modalité 4Ae, à interpréter cette expression, en avançant qu'un tel cas de résultat est forcément remarquable et important dans la vie sociétale. Ce qui met en évidence l'effet négatif que pourrait impliquer le vocabulaire utilisé, sur la compréhension des tests statistiques.

Une solution préconisée, susceptible d'aider à surmonter cette difficulté, est de remplacer, lors d'un enseignement sur cet outil statistique, une telle expression usuelle par celle la plus juste suivante : « *le résultat est statistiquement significatif* », en insistant chaque fois que possible, en classe de cours, sur l'adverbe « *statistiquement* » et ce afin de faire habituer les étudiants de ce langage statistique, d'une part, et de les rappeler de manière assez insistante du caractère *probabiliste* (et *non déterministe*) du résultat qu'un test de signification pourrait fournir, de l'autre. D'une manière générale, surmonter les trois habitudes de langage suivantes pourrait aider à éviter les inconscientes interprétations erronées en tests statistiques:

- *Dire* : « *statistiquement significatif* » plutôt que « *significatif* ». Cette phrase permettra de mettre en question l'association que la plupart des gens conçoivent erronément entre le rejet d'une hypothèse nulle et l'obtention d'un résultat important.
- *Ne pas dire des choses telles que* : « *mes résultats sont proches de la signification statistique* ». Ce langage n'a pas de sens dans la logique des tests statistiques.
- *Ne pas dire des choses comme* : « *le test statistique a évalué si les résultats sont dus au hasard* ». Ce langage donne l'impression que la réplique puisse être évaluée par l'outil de tests statistiques.

En conclusion, ce 1^{er} axe factoriel peut être interprété comme *un axe d'échec/réussite au vocabulaire utilisé en tests de signification, accompagné d'une réussite/échec au fait d'exprimer d'une autre façon le résultat d'application de tels tests.*

5.4.5 Interprétation du 2^{ème} axe factoriel

L'AFCM conduit à une inertie de 23% pour l'axe 2. Les modalités qui contribuent le plus à la construction de cet axe sont celles entourées dans la figure 2 ci-dessous****. Elles ont une contribution comprise entre 2,5% et 14,5%. La qualité de représentation de ces modalités par rapport à un tel axe est comprise entre 0,06 et 0,45.

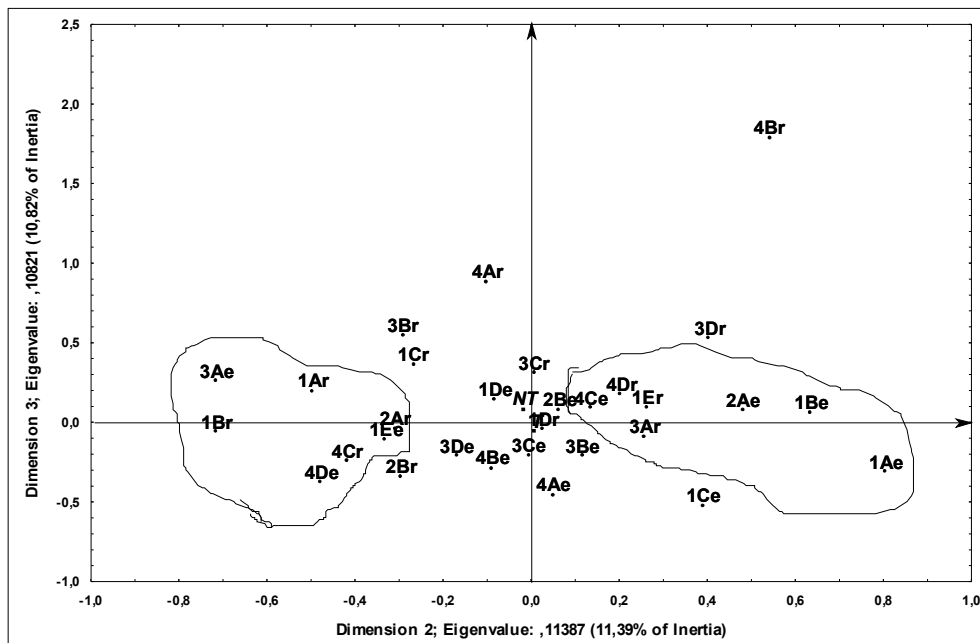


Fig. 2. Plan factoriel (2, 3).

Dans ce deuxième plan factoriel, le 2^{ème} axe présente l'opposition fournie par le tableau 6 suivant :

Tableau 6. Positionnements des modalités sur l'axe 2.

| A gauche de l'axe 2 | A droite de l'axe 2 |
|-----------------------------------|-----------------------------------|
| 3Ae, 1Br, 1Ar, 4De, 4Cr, 1Ee, 2Ar | 4Ce, 4Dr, 3Ar, 1Er, 2Ae, 1Be, 1Ae |

Dans cette opposition, nous avons, d'un côté, des modalités d'échecs et, de l'autre, leurs correspondantes de réussites, il suffira alors de donner, presque comme dans le cas du 1^{er} axe factoriel, une interprétation à l'un des groupes de modalités opposées, du fait que l'autre groupe aura tout simplement l'interprétation contraire de ce dernier. Ainsi, nous nous focalisons dans ce qui suit sur le groupe de modalités, situé dans le côté positif de cet axe.

Nous aborderons une telle interprétation, en portant une attention particulière à la modalité 1Be, et ce pour deux raisons. La première raison est que sa modalité opposée (1Br) concerne la définition qu'il ne faut pas donner au concept de « l'hypothèse nulle ». Notons au passage que ce concept, ainsi que ladite appellation qu'on lui a affectée, sont propres au paradigme des tests de signification (de Fisher) seul. Ce concept d'hypothèse nulle est plus ou moins équivalent à : « une hypothèse prévue pour être falsifiée, afin de gagner un support susceptible de permettre un progrès scientifique » selon ce paradigme. Rappelons aussi que la notion d'hypothèse, qu'elle soit appelée nulle ou « non nulle ! », constitue la base et le point de départ des procédures des différents outils de la statistique inférentielle, dans la

**** Ibid

mesure où elle représente l'objet sur lequel se fondent les différentes étapes de telles procédures, à savoir : « l'étape de prouver la véracité d'une hypothèse », « l'étape d'interpréter la véracité d'une hypothèse », « l'étape d'exprimer autrement la véracité (ou la fausseté) d'une hypothèse », etc. La deuxième raison est qu'une telle modalité (1Be) traduit l'appropriation des étudiants d'une définition erronée à ce concept en la confondant ainsi avec « l'hypothèse testée » propre au paradigme des tests d'hypothèse de Neyman-Pearson. Un tel paradigme est considéré comme la pièce maîtresse essentielle et un acteur pondérant de l'amalgame qui envahit l'enseignement, l'apprentissage et la mise en pratique des tests statistiques. Une telle hypothèse, selon ce paradigme, signifie : « aucune action à prendre » ou « ne rien faire, et laisser les choses comme vous les avez trouvées ».

Nous nous plaçons donc dans le cadre du paradigme des tests d'hypothèse, suscité par la modalité 1Be; puis nous allons essayer de mettre en place des interprétations éventuelles aux proximités remarquées entre les différentes modalités de réussites et d'échecs formant avec 1Be un même groupe au côté positif du 2^{ème} axe factoriel, tout en essayant de mettre en œuvre les différentes spécificités du vocabulaire impliqué dans un tel paradigme d'une manière générale, et la signification (erronée) donnée par les étudiants au concept d'hypothèse nulle d'une manière particulière.

Ainsi, la modalité 1Ae traduit le fait que le terme « nulle » suscité dans l'expression « hypothèse nulle » implique une certaine ambiguïté chez les étudiants qui le considèrent erronément comme équivalent aux termes : « sans valeur » et « aucune importance ». Ceci est dû à la confusion que font ces étudiants entre ces deux termes et le terme « aucune » que comporte l'expression « aucune action à prendre » utilisée par le paradigme des tests d'hypothèse pour désigner « l'hypothèse testée ». Ce qui renforce une autre fois l'effet négatif que pourrait avoir le vocabulaire adopté, sur la compréhension des tests statistiques. En revanche, la modalité 1Er traduit le fait que tels étudiants réussissent à comprendre que le terme « nulle » ne signifie pas « une égalité » ou « une différence nulle » comme il est presque toujours erronément imaginé par la plupart des apprenants et même des experts. Cette mauvaise perception constitue des principaux éléments sur lesquels se base le rituel utilisé en termes d'application des tests statistiques. De ce fait, la proximité remarquée entre 1Be et 1Er nous permet de déduire qu'une telle réussite est une exclusivité des étudiants appropriant le paradigme de tests d'hypothèse de Neyman-Pearson, que d'autres.

L'ambiguïté et la mauvaise stabilité de compréhension, débutées par la non maîtrise de la signification de « l'hypothèse nulle », que nous avons fait remarquer ci-dessus, continue à être mise en évidence dans les autres étapes des procédures sous-jacentes aux tests statistiques. En effet, la modalité 2Ae traduit le fait que les étudiants ne considèrent la véracité d'une hypothèse comme prouvée que si l'on présente pour ceci des arguments absolument solides comme dans le cas d'une démonstration purement mathématique. Cependant, la modalité 3Ar exprime que selon ces mêmes étudiants, si ce type de preuve avait permis de conclure la véracité d'une telle hypothèse, alors cette dernière (l'hypothèse) ne serait considérée qu'un modèle réalisable parmi d'autres qui sont possibles, reflétant au mieux ce qui se passe effectivement dans le monde réel.

Finalement, comme nous avons indiqué lors de l'interprétation du 1^{er} axe factoriel, dire que « le résultat est significatif » est tout simplement une autre façon d'exprimer la non véracité de « l'hypothèse nulle » dans le cas des tests de signification de Fisher. Mais, il peut aussi être utilisé pour exprimer la décision de « rejet de l'hypothèse testée » dans le cas du paradigme des tests d'hypothèse de Neyman-Pearson. Il est vrai que la modalité 4Dr est en concordance avec la modalité 3Ar, du fait qu'elles traduisent toutes les deux le caractère relatif d'un tel résultat. Ceci est déduit des termes suivants qu'elles suscitent : « modèle », « peut être », ...

Mais, la modalité 4Ce avançant qu'exprimer un type de résultat de cette façon signifie qu'il n'est pas dû au hasard et de plus il existe une cause bien définie qui l'a suscité, met en

évidence le caractère ambigu que nous avons indiqué ci-dessus. En tout cas, ceci accentue une autre fois l'effet négatif que pourrait impliquer le vocabulaire utilisé en tests statistiques, étant donné que les étudiants sont presque toujours familiarisés par la conclusion erronée usuelle suivante : « *le résultat n'est pas dû au hasard* », dans le cas où l'on rejette l'hypothèse nulle.

En conclusion, ce 2^{ème} axe factoriel peut donc être interprété comme *un axe de maîtrise partielle du vocabulaire habituel, utilisé en procédures de tests d'hypothèse de Neyman-Pearson*.

5.4.6 Interprétation du 3^{ème} axe factoriel

L'AFCM conduit à une inertie de 18% pour l'axe 3. Les modalités qui contribuent le plus à la construction de cet axe sont celles entourées dans la figure 3 ci-dessous⁺⁺⁺⁺. Elles ont une contribution comprise entre 2% et 27%. La qualité de représentation de ces modalités par rapport à un tel axe est comprise entre 0,11 et 0,51.

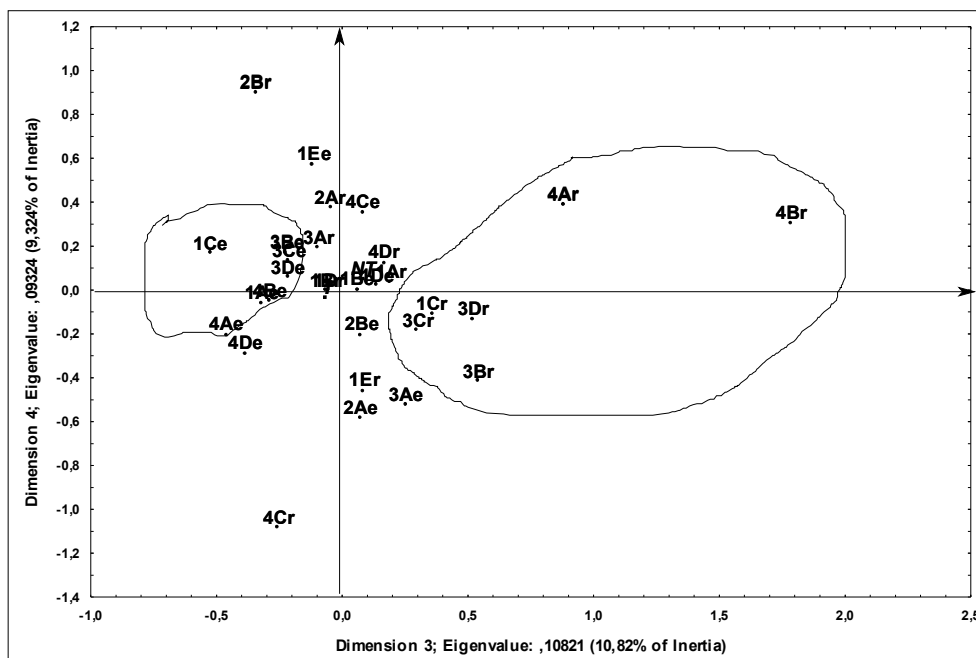


Fig. 3. Plan factoriel (3, 4).

Dans ce troisième plan factoriel, le 3^{ème} axe présente l'opposition fournie par le tableau 7 suivant :

Tableau 7. Positionnements des modalités sur l'axe 3.

| A gauche de l'axe 3 | A droite de l'axe 3 |
|-----------------------------------|------------------------------|
| 1Ce, 4Ae, 2Ae, 4Be, 3Be, 3Ce, 3De | 3Cr, 1Cr, 3Dr, 3Br, 4Ar, 4Br |

Dans cette opposition, nous avons, comme dans le cas du 2^{ème} axe factoriel, d'un côté, des modalités d'échecs et, de l'autre, leurs correspondantes de réussites, il suffira alors de donner une interprétation à l'un des groupes de modalités opposées, du fait que l'autre groupe

⁺⁺⁺⁺ *Ibid*

aura tout simplement l'interprétation contraire de ce dernier. Ainsi, comme dans le cas de l'interprétation du 2^{ème} axe factoriel, nous nous focalisons dans ce qui suit sur le groupe de modalités (de réussites), situé dans le côté positif de cet axe.

De prime abord, la modalité 1Cr nous permet de conclure que ce groupe de modalités est approprié par des étudiants qui comprennent bien la définition donnée dans les tests de signification (de Fisher), au concept de « *l'hypothèse nulle* ». Ce qui constitue un indice adéquat pour une bonne appréhension de ce paradigme. De plus, les modalités 3Br et 3Cr traduisent le fait que la signification donnée à l'expression « *hypothèse vraie* » est à caractère relatif et non absolu, puisque d'après tels étudiants une telle véracité est, d'un côté, une question de jugement de l'individu qui l'évalue (3Br) ; et il suffit de présenter des arguments relativement convaincants pour la prouver, plus le fait que cette véracité ne dure pas pour toujours (3Cr), de l'autre. Il est aussi à signaler que la façon par laquelle les items 4A et 4B sont construits justifie la proximité remarquée entre les deux modalités 4Ar et 4Br, dans la mesure où l'une est une conséquence logique de l'autre. Ainsi de telles réussites expriment que le caractère relatif donné par les étudiants de ce groupe à l'expression « *résultat significatif* », qui la considèrent comme une affaire qui n'est forcément ni remarquable ni important dans la vie sociale.

Il est vrai que dans ce groupe de modalités, on manque les deux modalités 2Ar et 2Br indiquant le caractère relatif du terme « *prouver* », utilisé dans le raisonnement associé aux tests de signification, afin de pouvoir conclure que tout le vocabulaire adopté n'a pas d'influence négative sur toutes les étapes des procédures sous-jacentes à ce paradigme. Mais, les positionnements de ces deux modalités dans ce 3^{ème} plan factoriel, montrant qu'elles sont non discriminantes, nous permettent de nous assurer que la signification donnée par les étudiants au terme « *prouver* » est au moins n'est pas absolu.

De ce fait, nous pouvons conclure que ce 3^{ème} axe factoriel est *un axe d'échec/réussite du vocabulaire utilisé en tests de signification de Fisher*.

6 Conclusions et perspectives didactiques

Nous nous sommes focalisés dans ce travail sur le *vocabulaire* comme un facteur qui envahit la compréhension des tests statistiques. Les raisons pour cela viennent du fait que c'est le dernier souci qui fort probablement préoccupe la plupart des enseignants de statistique (et de probabilités). Ces derniers risquent de le déconsidérer souvent durant leurs cours dispensés à leurs étudiants. Pareillement, ils peuvent négliger son ampleur et ses effets néfastes sur la compréhension des concepts statistiques qui sont habituellement connus pour être plus liés à la vie de tous les jours que les autres concepts de mathématiques d'une manière générale. Par ailleurs, l'importance excessive donnée dans ce travail à ce facteur ne devrait pas nous faire oublier des difficultés, fort probablement d'être causées par les autres facteurs impliquant des gravités à ne pas négliger, et qui méritent d'être objet d'études plus approfondies.

Ainsi, les résultats d'analyse factorielle multidimensionnelle des réponses des étudiants interrogés ont donné lieu à des axes factoriels avec des interprétations en termes de *maîtrise du vocabulaire utilisé en tests statistiques*, axes qui se déclinent selon l'importance de leur part d'inertie absolue (i.e., importance de leur part d'information contenue dans l'ensemble des réponses) suivant les interprétations respectives ci-après :

- *Un échec/réussite au vocabulaire utilisé en tests de signification de Fisher, accompagné d'une réussite/échec au fait d'exprimer autrement le résultat d'application de tels tests (Axe 1).*
- *Une maîtrise partielle du vocabulaire habituel, utilisé en procédures des tests d'hypothèse de Neyman-Pearson (Axe 2).*

- *Un échec/réussite du vocabulaire utilisé en procédures des tests de signification de Fisher (Axe 3).*

En plus de la focalisation, dans notre analyse, sur les modalités des différents items, l'étude des individus auxquels nous avons administré notre questionnaire, bien qu'elle soit succincte, a été aussi parmi nos préoccupations. Pour cela, nous avons créé une variable supplémentaire à deux modalités : *T* et *NT*, exprimant respectivement si ou non les individus étudiés avaient déjà suivi un cours sur les tests statistiques. Les positionnements de ces deux modalités sur les trois plans factoriels (cf. Fig. : 1, 2 et 3) montrent que seul le 1^{er} axe factoriel a permis une certaine discrimination modérée entre ces individus. Ce qui met en œuvre l'effet négatif que pourrait avoir le vocabulaire utilisé sur la compréhension des tests statistiques. Une telle discrimination est ainsi traduite par le fait que contrairement aux individus n'ayant jamais suivi de cours sur les tests statistiques, ceux ayant déjà suivi expriment seulement une certaine réussite quant au vocabulaire utilisé en tests de signification, mais accompagné d'un échec au fait d'exprimer autrement le résultat d'application de tels tests.

Il est vrai que parmi toutes les modalités des différents items de ce questionnaire, seules les modalités, 1Dr et 1De, de l'item 1D n'avaient pas fait partie de notre analyse, et ce du fait qu'elles ne discriminent pas entre les individus situés de part et d'autres des trois axes factoriels considérés. Par ailleurs, les pourcentages non trop faibles des individus ayant choisi ces deux modalités (27% pour 1De contre 73% pour 1Dr) ne nous ont pas permis d'exclure cet item de notre analyse factorielle. De toute façon, l'absence d'une même tendance générale entre la véracité ou la fausseté de cet item et celles des autres items du questionnaire, est possible. Il est donc probable que les étudiants se soient basés sur le hasard pour choisir la modalité qui leur semble adéquate, de cet item. Par ailleurs, si l'on admet que l'utilisation d'un tel hasard existe vraiment, ceci ne donnera pas raison à quiconque de critiquer la façon par laquelle cet item est construit, en avançant, par exemple, qu'il comporte une certaine ambiguïté qui a poussé les étudiants de procéder ainsi. Au contraire, l'échec à cet item n'explique pas seulement la difficulté que pourrait impliquer le vocabulaire, objet de notre travail, mais aussi une certaine incompréhension plus large dérivant de conceptions erronées dont souffrent même certains chercheurs chevronnés qui se basent sur le rituel pour mener leurs recherches. En effet, ils choisissent souvent comme concurrentes à leurs hypothèses alternatives (ou de recherche), des hypothèses nulles qu'ils considèrent comme des « *hypothèses nulles* », qui ne sont pas *théoriquement enracinées* et qui sont souvent fausses même avant de les mettre sous test [37]. En procédant ainsi, ils croient de manière erronée qu'ils ont démontré la véracité de leurs hypothèses de recherche. En fait, ils veulent atteindre une telle véracité à tout prix, en convoquant ainsi une conception erronée usuelle qui considère la véracité de l'hypothèse de recherche avoir plus de valeur que celle de l'hypothèse nulle.

Cette caractéristique d'interférence entre des difficultés relatives au vocabulaire utilisé et de conceptions erronées n'est pas une spécificité de l'item 1D seul, mais c'est un dénominateur commun des différents items de la question 1. Le contenu de cette dernière met en évidence une base théorique enrichissante relative à la notion « *d'hypothèse nulle* » qui sera l'objet d'une étude didactique détaillée, que nous allons mener dans le plus proche possible, à côté d'autres études relatives à d'autres facteurs susceptibles d'induire des difficultés éventuelles de compréhension des tests statistiques. Nous pouvons citer, à titre d'exemple, le facteur « *d'ambiguïté linguistique* », que pourraient impliquer les expressions, surtout à caractère probabiliste, sur l'interprétation des différents concepts véhiculés dans les procédures sous-jacentes à cet outil statistique...

Références

1. M. Oakes, *Statistical Inference: A commentary for the social and behavioural sciences* (Chichester, England: Wiley, 1986).
2. R. Falk, C. W. Greenbaum, Significance test die hard. *Theory & Psychology* **5(1)**, 75-98 (1995).
3. H. Haller, S. Krauss, Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research* **7(1)**, 1-20 (2002).
4. J. Poitevineau, B. Lecoutre, the interpretation of significance levels by psychological researchers: The .05-cliff effect may be overstated. *Psychonomic Bulletin and Review* **8**, 847-850 (2001).
5. Z. Elm'hamedi, *Contribution à une ingénierie didactique pour l'enseignement et l'apprentissage des tests statistiques à l'université*, thèse de doctorat (Université Sidi Mohammed Ben Abdellah, Fès, 2010).
6. Z. Elm'hamedi, Effets d'un apprentissage empirique sur la compréhension du concept de moyenne arithmétique. *Annales de Didactique et de Sciences Cognitives* **19**, 129 – 169 (2014).
7. Z. Elm'hamedi, *L'apprentissage empirique et son impact sur la compréhension du concept de moyenne arithmétique*, dans actes du 5^{ème} Colloque Francophone International sur l'Enseignement de la Statistique (CFIES 2017), 06-08 septembre 2017, Grenoble, France (2017).
8. Z. Elm'hamedi, La compréhension du concept de moyenne arithmétique : au-delà des connaissances calculatoires. *Revue internationale scientifique de l'éducation et de la formation (RSIEF)* **2**, 76-99 (2019a).
9. G. Gigerenzer, The superego, the ego, and the id in statistical reasoning. In G. Keren, C. Lewis (Eds.) *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (Hillsdale, NJ: Erlbaum, 1993).
10. M. Zaki, Z. Elm'hamedi, *Mesures de support facilitant une meilleure appréhension du concept de tests statistiques*, dans actes du Symposium International Forapeval_st2006, 23-24 novembre 2006, Fès, Maroc (2006).
11. M. Zaki, Z. Elm'hamedi, Eléments de mesures pour un enseignement des tests statistiques. *Annales de Didactique et de Sciences Cognitives* **14**, 153-194 (2009).
12. Z. Elm'hamedi, M. Zaki, *Les tests statistiques : Quelques aspects de critiques non fondées*, dans actes du 1^{er} Congrès International sur la Statistique Mathématique, 22-24 avril 2008, Alep, Syrie (2008).
13. Z. Elm'hamedi, M. Zaki, *Les tests statistiques entre Fisher et Neyman-Pearson*, dans actes de la 3^{ème} Biennale du Réseau Marocain de la Didactique des Sciences (REMADIS), 12-14 mai 2005, Errachidia, Maroc (2005).
14. M. Zaki, Z. Elm'hamedi, Aspects de quelques critiques non fondées de la théorie des tests statistiques. *Annales de Didactique et de Sciences Cognitives* **18**, 139-171 (2013).
15. Z. Elm'hamedi, *La logique hybride dans l'utilisation des tests statistiques*, dans actes du 1^{er} Colloque International sur la Formation et l'Enseignement des Mathématiques, 07-08 avril 2016, El Jadida, Maroc (2016).
16. Z. Elm'hamedi, Effets du vocabulaire et de l'ambiguïté linguistique sur la compréhension des tests statistiques. *Annales de Didactique et de Sciences Cognitives* **24**, 133-181 (2019b).

17. Z. Elm'hamedi, Difficultés de compréhension du concept de niveau de signification, (à paraître dans la revue *Education & Didactique*) (2021).
18. R. B. Kline, *Beyond significance testing: Reforming data analysis methods in behavioral research* (DC: APA Books, Washington, 2004).
19. G. Vergnaud, La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques* **10(2-3)**, 133-170 (1991).
20. A. L. Mclean, Statistacy: Vocabulary and hypothesis testing. *ICOTS* **6** (2002)
21. D. Lamrabet, *La démonstration mathématique : structure et situations*, dans actes du Congrès international : Universalité et Localité en Sciences, mars 2001, Rabat, Maroc (2001).
22. A. L. Mclean, On the nature and role of hypothesis tests. Department of Econometrics and Business Statistics Working Paper 4/2001 (2000).
23. G. Brousseau, Fondements et méthodes de la didactique des mathématiques. *Recherches en Didactique des Mathématiques* **7.2**, 33-115 (1986).
24. M. Artigue, Ingénierie didactique. *Recherches en Didactique des Mathématiques* **9.3**, 281-308 (1990).
25. N. Balacheff, *Modèle de connaissances pour le calcul de situations didactiques*, dans Conference : XII^e école d'été de didactique des mathématiques, août 2003, France (2003).
26. Z. Elm'hamedi, Quelques aspects de l'incompréhension des tests statistiques. *Recherches en Didactique des Mathématiques* **39/2**, 167 – 211 (2019c).
27. A. Vallecillos, Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios, (University students' understanding of the logic of hypothesis testing). *Recherches en Didactique des Mathématiques* **15(3)**, 53-81 (1995).
28. R. E. Kirk, Practical significance: A concept whose time has come. *Educational and Psychological Measurement* **5**, 746-759 (1996).
29. L. G. Daniel, Kerlinger's research myths: An overview with implications for educational researchers. *Journal of Experimental Education* **65**, 101-112 (1997).
30. R. W. Tyler, What is statistical significance? *Educational Research Bulletin* **10**, 115-118 (1931).
31. J. P. Schafer, Interpreting statistical significance and nonsignificance. *Journal of Experimental Education* **61(4)**, 383-387 (1993).
32. P. E. Meehl, The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In L. I., Harlow, S. A., Mulaik et J. H., Steiger (Eds.) *What if there were no significance tests?* 393-426 (Mahwah, NJ: Erlbaum, 1997).
33. H. J. Eysenck, The concept of statistical significance and the controversy about one-tailed tests. *Psychological Review* **67**, 269-271 (1960).
34. J. Poitevineau, *Méthodologie de l'analyse des données expérimentales : Etude de la pratique des tests statistiques chez les chercheurs en psychologie, approches normative, perspective et descriptive*, Ph. D (Université de Rouen, Rouen, 1998).
35. R. Fisher, *The Design of Experiments* (Edinburgh: Oliver & Boyd, 1935).
36. M. Granaas, Hypothesis testing in psychology: throwing the baby out with the bathwater? *ICOTS* **6** (2002).
37. R. W. Frick, Accepting the null hypothesis. *Memory & Cognition* **23(1)**, 132-138 (1995).
38. J. P. Benzecri, Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers de l'analyse des données* **4**, 377-378 (1979).

ANNEXE : Questionnaire

Question 1 : *Si l'on vous demande d'imaginer ce que vous pouvez comprendre par (ou ce que peut signifier pour vous) l'expression « Hypothèse Nulle », il surviendra à votre esprit :*

- | | | |
|---|---------------------------|---------------------------|
| A. Une hypothèse dont sa véracité sera sans valeur et n'aura aucune importance dans la vie sociétale. | <input type="radio"/> Oui | <input type="radio"/> Non |
| B. Une hypothèse dont sa véracité ne conduira à l'adoption d'aucune action nouvelle. | <input type="radio"/> Oui | <input type="radio"/> Non |
| C. Une hypothèse prévue pour être falsifiée ou rejetée, afin de gagner un support susceptible de permettre un progrès scientifique. | <input type="radio"/> Oui | <input type="radio"/> Non |
| D. Une hypothèse déjà connue d'être fautive avant même sa mise à l'épreuve. | <input type="radio"/> Oui | <input type="radio"/> Non |
| E. Une hypothèse exprimée sous forme d'une égalité ou d'une différence nulle. | <input type="radio"/> Oui | <input type="radio"/> Non |

Question 2 : *« Prouver la Véracité d'une Hypothèse » signifie pour vous :*

- | | | |
|---|---------------------------|---------------------------|
| A. Présenter des arguments aussi solides que possible en faveur de la véracité de cette hypothèse, tout en fixant un seuil exigé et réfléchi d'évidence, et en exposant la force de preuve ou la puissance de votre raisonnement. | <input type="radio"/> Oui | <input type="radio"/> Non |
| B. Présenter des arguments absolument convainquant comme dans le cas d'une démonstration purement mathématique envers la véracité de cette hypothèse. | <input type="radio"/> Oui | <input type="radio"/> Non |

Question 3 : *L'expression « Hypothèse Vraie » signifie pour vous :*

- | | | |
|--|---------------------------|---------------------------|
| A. Une telle hypothèse est un modèle reflétant au mieux le monde réel, permettant à la plupart des gens rationnels de l'accepter comme modèle réalisable. | <input type="radio"/> Oui | <input type="radio"/> Non |
| B. La véracité d'une telle hypothèse est seulement une affaire de jugement, dans la mesure où un individu peut considérer cette hypothèse vraie alors que son voisin le plus proche peut décider le contraire. | <input type="radio"/> Oui | <input type="radio"/> Non |
| C. On a présenté des arguments absolument convainquant envers la véracité de cette hypothèse, comme dans le cas d'une démonstration purement mathématique, et que cette véracité dure pour toujours. | <input type="radio"/> Oui | <input type="radio"/> Non |
| D. Une telle hypothèse est un modèle dans lequel on tolère qu'il puisse exister quelques peu de cas pour lesquels cette hypothèse ne tient pas, mais sans pour autant l'étiqueter comme étant non valide dans le monde réel. | <input type="radio"/> Oui | <input type="radio"/> Non |

Question 4 : *Si l'on vous demande ce que peut signifier pour vous l'expression « Résultat Significatif », vous penserez à un résultat qui :*

- | | | |
|---|---------------------------|---------------------------|
| A. Est forcément remarquable et important dans la vie sociétale. | <input type="radio"/> Oui | <input type="radio"/> Non |
| B. Peut ne pas être ni remarquable, ni important dans la vie sociétale. | <input type="radio"/> Oui | <input type="radio"/> Non |
| C. N'est pas dû au hasard, dans la mesure où il existe une cause bien connue et bien définie qui l'a suscité. | <input type="radio"/> Oui | <input type="radio"/> Non |
| D. Peut-être remarquable et important dans la vie sociétale. | <input type="radio"/> Oui | <input type="radio"/> Non |