American Journal of Innovative Research and Applied Sciences. ISSN 2429-5396 I www.american-jiras.com

ORIGINAL ARTICLE

LA LOGIQUE HYBRIDE DANS L'UTILISATION DES TESTS STATISTIQUES

HYBRID LOGIC IN THE USE OF STATISTICAL TESTING

| Zahid El M'hamedi |

Centre Régional des Métiers de l'Education et de la Formation | Casablanca-Settat | Département de Mathématiques | Settat | Maroc |

| Received 02 October 2019 | | Accepted 24 October 2019 | | Published 01 November 2019 | | ID Article | Zahid-Ref.12-ajira-021019 |

RESUME

Contexte: La théorie des tests statistiques est la pierre angulaire des recherches empiriques dans des domaines aussi variés que l'éducation, la psychologie, la sociologie, la médecine, l'agronomie, l'écologie, le droit, etc. Malheureusement, ce qui est généralement présenté aujourd'hui en matière d'enseignement de cette théorie, est un amalgame entre les *tests de signification* de Fisher et les *tests d'hypothèses* (ou de décision) de Neyman-Pearson. **Objectifs:** L'objectif de cet article est de mettre en évidence l'occurrence de quelques aspects de cet amalgame chez les étudiants. **Méthodes:** Afin de réussir ce projet, nous avons choisi, d'une manière aléatoire, un échantillon constitué de 70 étudiants, auquel nous avons administré un questionnaire composé de trois tâches à réaliser, impliquant tels aspects. **Résultats:** Nos étudiants sont en situation d'amalgame entre la théorie de Fisher et celle de Neyman-Pearson. **Conclusions:** Nous pensons que la mise en place d'une ingénierie didactique en matière d'enseignement et d'apprentissage de la théorie des tests statistiques s'avère nécessaire et urgente. Une telle ingénierie doit prendre en considération le fait de ne pas présenter les tests statistiques comme un amalgame entre les tests de signification et les tests d'hypothèses, mais comme deux théories totalement différentes, dont chacune a ses propres caractéristiques.
Mots clés: *Tests de signification, tests d'hypothèses, approche bayésienne, logique hybride, niveau de signification.*

ABSTRACT

Background: Statistical testing theory is the cornerstone of empirical researches in many fields such as: education, psychology, sociology, medicine, agronomy, ecology, law, etc. Unfortunately, what is commonly presented in matter of teaching this theory is an amalgam between the significance testing of Fisher and hypothesis testing of Neyman-Pearson. **Objectives:** The aim of this paper is to underline the existence of some aspects of this amalgam among students. **Methods:** In order to successfully carry out this project, we chose a random sample constituted of 70 university students, to whom we administered a questionary including tree tasks involving such aspects. **Results:** Our students are in an amalgam of Fisher's theory with that of Neyman-Pearson. **Conclusions:** We believe that the implementation of didactic engineering in teaching and learning the theory of statistical tests is necessary and urgent. Such engineering must take into account not presenting statistical tests as an amalgam between significance tests and hypothesis tests, but as two totally different theories, each of which has its own characteristics.
Keywords: *Significance testing, hypothesis testing, bayesian approach, hybrid logic, significance level.*

1. INTRODUCTION

L'un des principaux objets de la discipline de statistique est d'établir des inférences relatives à des populations sur la base d'observations collectées à partir d'échantillons issus de telles populations, et auxquels on soumet à une réalisation (ou épreuve) singulière. Les résultats de ces inférences, qui sont toujours à caractère probable et non déterministe, pourraient être obtenus par le biais d'une masse importante de théories relevant de la boite à outils de la statistique inférentielle [1, 2]. De telles théories sont souvent traitées comme étant des alternatives, mais en fait elles pourraient se compléter dans le but de mettre en évidence les erreurs de fluctuations d'échantillonnage et celles de mesures expérimentales, fort probable d'être commises lors du passage du connu (échantillons) à l'inconnu (populations).

En outre, les tests de signification systématisés par Fisher (1925, 1935, 1956) [3, 4, 5]et les tests d'hypothèses (ou de décision) formalisés par Neyman et Pearson (1928, 1933) [6, 7] sont deux théories alternatives, utilisant les mêmes éléments - à savoir les hypothèses statistiques H_0 et H_1 relatives aux populations et les données observées D_0, issues des échantillons soumis à une épreuve singulière w_0- afin de jouer des rôles strictement différents, tout en procédant de manières non identiques, et en véhiculant des concepts statistiques ainsi que des raisonnements propres à chacun d'eux.

Malheureusement, ce qui est généralement présenté aujourd'hui en matière d'enseignement du concept de test statistique, et ce dont traitent la plupart des ouvrages spécialisés en statistiques, est un amalgame, appelé «*logique hybride de l'inférence scientifique*» [8], entre les *tests de signification* de Fisher et les *tests d'hypothèses* (ou de décision) de Neyman-Pearson. Le dénominateur commun entre ces deux théories - et aussi entre toutes les théories de l'inférence statistique - est que les résultats qu'elles pourraient générer toutes les deux sont à caractère relatif et non absolu, probable et non déterministe, et comportent souvent des risques d'erreurs [9, 10].

De son côté, Denis (2004) a annoncé que les tests statistiques, comme ils sont traités aujourd'hui, sont des méthodes hybrides qui diffèrent du modèle proposé par Fisher en 1925 [11]. Il a montré qu'associer le nom de Fisher à ces méthodes non seulement est incorrect, mais encore adresse à Fisher des reproches injustifiés au sujet des profondes faiblesses méthodologiques et philosophiques des tests statistiques. Il a essayé de distinguer entre la méthodologie originale de Fisher et l'hybride actuelle. Il a ainsi conclu que les chercheurs en sciences sociales utilisent aujourd'hui un défectueux et déplaisant mélange des ingrédients dus à Fisher, Neyman-Pearson et Bayes.

Ainsi, des méthodes bayésiennes pour la plupart des situations rencontrées dans l'analyse des données expérimentales ont été développées. Elles peuvent être utilisées et enseignées facilement et ouvrent de nouvelles voies prometteuses dans la méthodologie statistique [12].

Nous allons présenter dans cet article une étude exploratoire sur quelques aspects d'amalgame que les étudiants à l'université pouvant commettre lors de leur traitement « des procédures » relatives à l'application des théories sous-jacentes aux tests statistiques. Mais avant d'y arriver, nous allons, dans un premier temps, donner brièvement un aperçu global sur les procédures d'application des tests de signification de Fisher et des tests d'hypothèses de Neyman-Pearson, en se basant pour cela sur une situation de test statistique usuellement employée.

2. TESTS DE SIGNIFICATION ET TESTS D'HYPOTHÈSES

Fisher (1925, 1935, 1956) [3, 4, 5] et Neyman et Pearson (1928, 1933) [6, 7] sont à l'origine d'ouvrages et d'articles fondateurs des théories relatives aux tests de signification et aux tests d'hypothèses. Plus récemment, Carver (1978) [13] et Poitevineau (1998) [14] ont présenté les tests statistiques d'une manière plus simple et facile à comprendre.

Néanmoins, nous allons ci-après nous baser sur la situation de tests statistiques (dite de *conformité*) pour esquisser un bref aperçu sur les procédures sous-jacentes à ces deux types de tests. La situation est la suivante:

Tableau 1: illustre la situation du test de conformité

- Soit X une variable aléatoire qui suit une loi normale de moyenne μ et d'écart type connu σ : $X \sim N(\mu, \sigma^2)$
- On considère les deux hypothèses statistiques suivantes H_0 et H_1 relatives à la moyenne μ de X, et auxquelles on veut appliquer ***un test statistique***[1] :

$$\begin{cases} H_0 : \text{“}\mu = \mu_0\text{”} \\ H_1 : \text{“}\mu > \mu_0\text{”} \end{cases}$$

- Pour cela, on a recueilli à l'aide d'un échantillon $(X_1, \ldots, X_{n_0})$ de taille n_0 et d'une épreuve (ou réalisation) w_0, les données D_0 : $(x_1^0, \ldots, x_{n_0}^0)$ telles que $\overline{x}_{ob} = \frac{x_1^0 + \ldots + x_{n0}^0}{n_0}$ (moyenne *ob*servée).
- $\overline{X} = \frac{X_1 + \ldots + X_{n_0}}{n_0}$ désigne la moyenne d'échantillon $(X_1, \ldots, X_{n_0})$ issu de la variable aléatoire X.

2.1 Tests de signification (de Fisher)

Le rôle d'un test de signification est de *conclure,* avec une forte probabilité, si les données observées D_0 rejettent l'hypothèse nulle H_0. Les deux conclusions possibles sont «Oui» ou «Non»:

- Oui: H_0 est fausse (ou Rejeter H_0). On dit aussi que les données observées D_0 sont statistiquement significatives.
- Non: Echec dans le rejet de H_0 (ou les données D_0 sont statistiquement non significatives).

Pour ce faire, on détermine la probabilité $P(\overline{X} \geq \overline{x}_{ob} \mid H_0)$[2], appelée «p-value». Elle est égale à $P\left[N(0,1) \geq \frac{\sqrt{n_0}(\overline{x}_{ob} - \mu_0)}{\sigma}\right]$ puisque $\frac{\sqrt{n_0}(\overline{X} - \mu_0)}{\sigma}$ suit la loi N(0,1). On la compare à un seuil de signification, noté α, choisi a posteriori (c'est-à-dire après avoir recueilli les données). Si cette probabilité est inférieure ou égale à α, on conclut que H_0 est fausse. Si, au contraire, elle est supérieure à α, on conclut qu'on a échoué dans le rejet de H_0.

2.2 Tests d'hypothèses (de Neyman-Pearson)

[1] C'est-à-dire: appliquer un test de signification et un test d'hypothèses.
[2] C'est la probabilité d'obtenir les données observées D_0 ou des données plus extrêmes que celles observées, sachant que H_0 est vraie.

Le rôle d'un test d'hypothèses est de *décider*, moyennant un risque d'erreur, entre H_0 et H_1. Les deux décisions auxquelles on peut aboutir sont : «Rejeter H_0 et Accepter H_1» *ou bien* «Accepter H_0 et Rejeter H_1».

La procédure adoptée par Neyman-Pearson consiste à élaborer une règle de décision, en fonction de la taille n_0 de l'échantillon et d'une valeur[3] α appartenant à l'intervalle]0,1[. Cette règle de décision donne lieu à la construction d'une région[4] de rejet de H_0 et d'acceptation de H_1. Cette région, qu'on peut noter RC, est caractérisée par :

$$\begin{cases} P(RC|H_0) \text{ ne dépasse pas } \alpha \\ \text{et} \\ P(RC|H_1) \text{ est maximale} \end{cases}$$

Dans le cas de la situation qu'on est en train de traiter:

$RC=\left\{(X_1,\dots,X_{n_0})/\overline{X}\in\left]\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}},+\infty\right[\right\}$, avec Φ_α est le quantile d'ordre (1-α) de la loi normale centrée réduite N(0,1).

Si $\overline{x}_{ob}\in\left]\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}},+\infty\right[$ alors on décide de rejeter H_0 et d'accepter H_1 : le test est dit statistiquement significatif. Si $\overline{x}_{ob}\in\left]-\infty,\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}}\right]$ alors on décide d'accepter H_0 et de rejeter H_1 : le test est dit statistiquement non significatif.

Le risque de 1ère espèce est la probabilité $P\left(\overline{X}\in\left]\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}},+\infty\right[\,\middle|\,H_0\right)$, et celui de 2ème espèce est la probabilité $P\left(\overline{X}\in\left]-\infty,\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}}\right]\,\middle|\,H_1\right)$. La puissance du test est $P\left(\overline{X}\in\left]\mu_0+\Phi_\alpha\frac{\sigma}{\sqrt{n_0}},+\infty\right[\,\middle|\,H_1\right)$: c'est le complément à 1 du risque de 2ème espèce.

2.3 Quelques similitudes et différences entre les deux théories

A part le caractère relatif et probable des résultats fournis, partagé par tous les outils de la statistique inférentielle pour passer du connu (échantillon) vers l'inconnu (population), les tests de signification et les tests d'hypothèses font partis de deux théories totalement différentes et incompatibles dont leurs seuls points communs sont d'une part le fait qu'elles se basent toutes les deux sur la probabilité conditionnelle des données[5] D sachant les hypothèses H (P(D|H)), afin d'atteindre leurs objectifs, au lieu d'utiliser, au contraire, la probabilité conditionnelle inverse P(H|D), adoptée par d'autres théories telle que la statistique bayésienne, et d'autre part les résultats auxquels elles pourraient aboutir toutes les deux sont de type dichotomique:

- Les données observées D_0 rejettent-elles H_0 ? Oui ou Non (*cas des tests de signification*)
- Laquelle des deux hypothèses peut-on décider: H_0 ou H_1? (*cas des tests d'hypothèses*)

Pourtant, les différences entre les tests de signification et les tests d'hypothèses sont très nombreuses. Nous présentons, à titre d'exemples, quelques unes dans le tableau 2 ci-dessous:

Tableau 2: le tableau introduit quelques différences entre les tests de signification et les tests d'hypothèses.

Elément de différence	Tests de signification	Tests d'hypothèses
Domaine d'application	La recherche scientifique	La décision scientifique
Rôle	Conclure, avec une forte probabilité, si les données observées rejettent H_0	Décider, moyennant un risque de se tromper, entre H_0 et H_1
Type du raisonnement adopté	Déductif (qui permet de passer de la population de l'ensemble des échantillons possibles vers l'échantillon particulier duquel les données observées sont issues)	Comportement inductif (dans la mesure où ce qui nous intéresse n'est pas l'hypothèse H_i, mais l'action sous-jacente)
Signification de l'hypothèse H_0	Effet nul	Pas d'action
Signification de l'hypothèse	Effet strictement positif	Action nouvelle

3 Appelée «niveau de signification» du test.
4 Appelée aussi région critique du test.
5 D sont les données observées ou des données plus extrêmes que celles-ci dans le cas de Fisher ; et la région de rejet dans le cas de Neyman-Pearson.

H_1		
Hypothèse d'intérêt	H_1	H_0
Hypothèse(s) impliquée(s)	H_0	H_0 et H_1
Nombre d'expériences (réelles ou imaginaires) considérées dans une procédure	Unique	En nombre infini
Données observées	Elles caractérisent le test	Elles ne jouent qu'un rôle intermédiaire : ce n'est qu'après construction de la région de rejet que nous faisons appel aux données observées, pour décider
Localisation du niveau de signification	Après la collecte des données	Avant la collecte des données
Interprétation du niveau de signification	Caractéristique d'une expérience unique	Taux d'erreurs (*fréquence*) dans un échantillonnage répété
Détermination de la statistique test	D'une manière intuitive	A partir du rapport de vraisemblance
Résultat impliqué	Rejeter H_0 ou Echec de rejeter H_0	Rejeter H_0 ou Accepter H_0
Rejeter H_0	H_1 est acceptable	Accepter H_1
Population à partir de laquelle les données observées sont considérées issues	Population purement hypothétique, infinie et inconnue (une population imaginaire).	L'ensemble des répétitions imaginaires, à l'infini, de l'expérience réalisée

2.4 Quelques aspects d'amalgame entre les tests de signification et ceux d'hypothèses

Beaucoup de ceux qui lisent les ouvrages s'intéressant aux méthodes statistiques, se sont inclinés de croire les tests statistiques comme étant une théorie unifiée sans controverses, cherchant à rejeter l'hypothèse nulle afin de fournir une preuve de la fiabilité de l'hypothèse alternative [15]. Une p-value et un niveau alpha (α) sont alors fournis pour déterminer la probabilité de preuve que le résultat obtenu soit dû à la chance ou aux erreurs de fluctuations d'échantillonnage et de mesures expérimentales. Aussi, admettent-ils le fait qu'il y a au moins deux types d'erreurs qui peuvent être commises dans ce processus: si l'on rejette l'hypothèse nulle, une erreur de 1ère espèce, ou un faux résultat positif, peut avoir lieu ; et de même si l'on ne rejette pas l'hypothèse nulle, une erreur de 2ème espèce, ou un faux résultat négatif, peut avoir lieu.

Aussi, les chercheurs et même les statisticiens ne distinguent-ils pas clairement les tests de signification des tests d'hypothèses, ils utilisent souvent des aspects d'amalgames très divers, tout en combinant les éléments de différences cités dans le tableau 2 ci-dessus. De tels amalgames ont été dénoncés par plusieurs auteurs, à titre d'exemples: Morrison et Henkel (1970) [16], et Gigerenzer (1993) [8] qui parle dans ce sens de «logique hybride de l'inférence scientifique» pour faire référence à cette théorie unifiée d'amalgame qui vient d'être mise en place et qui est non seulement traitée par beaucoup d'ouvrages de statistiques, mais aussi appliquée par une masse importante de chercheurs et de praticiens expérimentés. Pire encore, des auteurs qui entendent distinguer la théorie de Fisher de celle de Neyman-Pearson, paraissent parfois victimes de ces amalgames [14].

Ainsi, les deux aspects d'amalgame les plus habituels sont les suivants:

- Les chercheurs se basent, le plus souvent, sur la notion de p-value (Fisher), *mais* ils font aussi appel, occasionnellement, à la notion de risque de 2ème espèce (ou puissance) (Neyman-Pearson).
- Les chercheurs utilisent souvent une statistique test ne dérivant pas d'un rapport de vraisemblance (Fisher), *et au même temps,* ils prennent des décisions qui ne pourraient être impliquées que par application de la théorie de Neyman-Pearson, à savoir : Accepter H_1, Accepter H_0.

Il est à noter aussi que dans certaines situations de tests statistiques, les données observées D_0 pourraient être un déterminant de la théorie à adopter : Fisherienne ou Neyman-Pearsonienne, dans la mesure où de telles données favoriseraient la vérité de l'une des deux hypothèses : H_0 ou H_1. Par exemple, la nature des données D_0 des situations de tests que nous présentons ci-après, exige l'adoption de la théorie des tests de signification et non celle des tests d'hypothèses:

- Le cas du test de Kolmogorov, interpellé pour jouer un rôle de test *secondaire* (d'un test *principal*) testant la normalité d'une distribution, alors que dans le test *principal* on a priori fait tout ce qu'il faut pour que les données observées soient issues d'une distribution normale, afin de pouvoir l'appliquer.
- Le cas du test de Signe ou celui de Wilcoxon, du fait que les valeurs ex-æquo, supportant d'une manière directe l'hypothèse H_0, sont écartées des données observées.

3. ETUDE EXPLORATOIRE AUPRÈS DES ÉTUDIANTS

3.1 Objectif de l'étude et expérimentation

Nous avons conduit une expérimentation auprès de 70 étudiants en 3ème année de formation (38 étudiants de la Faculté des Sciences de Rabat-Agdal et 32 élèves ingénieurs de l'Institut Agronomique et Vétérinaire de Rabat), auxquels nous avons administré durant une demi heure un questionnaire en vrai - faux, avec une demande de justification écrite des réponses. Les étudiants interrogés ont tous suivi un enseignement classique de probabilités – statistiques, d'un volume horaire de 70 heures, portant sur la théorie de base des probabilités (espaces probabilisés, variables aléatoires et lois de probabilités, lois des grands nombres,…), l'estimation ponctuelle et par intervalles de confiance, ainsi que les tests statistiques habituellement étudiés dans les cas paramétriques et non paramétriques.

Le questionnaire a été construit sur la base d'une situation relevant d'un test de *conformité* (cf. Tableau 1), généralement présentée aux étudiants dans leur cours sur les tests statistiques. Il est composé de 3 items- numérotés A, B et C – recouvrant quelques aspects d'amalgame.

Avant de procéder à l'analyse proprement dite des productions des étudiant, nous allons consacrer le prochain paragraphe à l'analyse a priori du questionnaire, afin de mieux cerner l'interprétation (statistique) des réponses des étudiants.

3.2 Analyse a priori du questionnaire

Les trois aspects (items numérotés de A à C) qui composent le questionnaire présenté aux étudiants, sont tous relatifs à la situation de test de *conformité* présenté dans le tableau 1 ci-dessus:

(A) A votre avis, l'application d'un test statistique dans cette situation sert à:

(1) Conclure, avec une forte probabilité, que les données observées D_0 rejettent l'hypothèse H_0

□ **Vrai** □ **Faux**

(2) Décider, moyennant un risque de se tromper, entre les hypothèses H_0:" $\mu = \mu_0$" et H_1:"$\mu = \mu_1$"

□ **Vrai** □ **Faux**

Analyse *a priori* de l'item A

La modalité (1) est vraie. Elle indique le rôle joué par un test statistique selon la théorie de Fisher. Rappelons que ce rôle est de conclure, avec une forte probabilité, si les données observées D_0 rejettent ou non l'hypothèse nulle H_0. La modalité (2) est aussi vraie. Elle indique le rôle joué par un test statistique selon la théorie de Neyman-Pearson. Cette dernière permettra de décider entre H_0 et H_1, moyennant un risque de se tromper dans cette décision.

(B) Une procédure de test statistique se base sur le concept de seuil de signification souvent noté α (ou ε). A votre avis, le seuil de signification α:

(1) Permet de construire, après avoir fixé la taille n_0 de l'échantillon, une règle de décision pour laquelle, au vu des données observées D_0 on décide de:

- rejeter l'hypothèse H_0 (et accepter l'hypothèse H_1)

Ou

- accepter l'hypothèse H_0 (et rejeter l'hypothèse H_1)

□ **Vrai** □ **Faux**

(2) Est directement comparé à la probabilité P $\left[N(0,1) \geq \frac{\sqrt{n_0}\left(\bar{x}_{ob} - \mu_0\right)}{\sigma} \right]$ pour conclure au rejet de H_0, ou à l'échec dans le rejet de H_0

□ **Vrai** □ **Faux**

Analyse *a priori* de l'item B

Les modalités (1) et (2) de cette question sont toutes les deux vraies. Elles permettent d'expliciter le rôle que peut jouer le concept de seuil de signification dans une procédure de test statistique selon la théorie de Neyman-Pearson (modalité (1)) et selon la théorie de Fisher (modalité (2)).

(C) Après application d'un test statistique, la (les) conclusion(s) qu'on peut tirer est (sont):

(1) Rejeter H_0 OU Accepter H_0

□ **Vrai** □ **Faux**

(2) Rejeter H_0 OU Echouer de rejeter H_0

☐ **Vrai** ☐ **Faux**

Analyse *a priori* de l'item C

Les modalités (1) et (2) sont toutes les deux vraies:

- si le test statistique en question est vu selon la théorie de Neyman-Pearson,
- si le test statistique en question est vu selon la théorie de Fisher.

Cependant, en nuançant les réponses, si nous sommes face à un test statistique selon la théorie de Fisher, la modalité (1) devient fausse, du fait que la conclusion «Accepter H_0» est différente «d'échouer de rejeter H_0». De même que, si nous sommes face un test statistique selon la théorie de Neyman-Pearson, la modalité (2) devient fausse, du fait que la conclusion «Echouer de rejeter H_0» n'implique pas «Accepter H_0». En fait, il y a deux raisons pour cela:

- La première est qu'il peut y avoir d'autres raisons que celles explicitées dans l'hypothèse H_0 pour que les données soient conformes aux prédictions de H_0, en particulier, dans le cas où il y aurait un manque de spécificité de H_0: des hypothèses H_i, autres que H_0, pourraient être elles qui soient vraies (et non H_0).

- La seconde est que le test peut manquer de puissance. Il peut arriver en effet que l'hypothèse H_1 soit vraie, alors que le dispositif expérimental ou le plan échantillonnal ne permettent pas de le détecter.

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

Nous avons décidé de retenir un codage en bimodalité «Réussite - Echec» pour toute modalité de chaque item du questionnaire.

Dans le tableau 3 suivant, nous résumons le codage des modalités des items du questionnaire, les libellés des items, ainsi que la signification associée à chaque modalité:

Tableau 3: le tableau présente le codage des modalités de chaque item du questionnaire.

Libellé	Codage	Signification	RR[6]	RE[7]
Rôle d'un test statistique	A1	- *Conception Fishérienne.*	A1R	A1E
	A2	- *Conception Neyman-Pearsonienne.*	A2R	A2E
Localisation du seuil de signification, α, dans une procédure de tests statistiques	B1	- *Avant la collecte des données (théorie de Neyman-Pearson).*	B1R	B1E
	B2	- *Après la collecte des données (théorie de Fisher).*	B2R	B2E
Résultat d'un test statistique	C1	- *Résultats en conformité avec la théorie de Neyman-Pearson.*	C1R	C1E
	C2	- *Résultats en conformité avec la théorie de Fisher.*	C2R	C2E

3.4 Etude de l'amalgame

3.4.1 Indicateur de mesure de l'amalgame

Nous disons qu'un individu est en situation d'amalgame entre la théorie de Fisher et celle de Neyman-Pearson par rapport aux deux caractéristiques K_1 et K_2 s'il adopte[8] une conception fishérienne[9] pour K_1 et une conception neyman-pearsonienne[10] pour K_2.
Par exemple, si K_1 est : «Rôle du test statistique» et K_2 est: «Localisation du niveau de signification α dans une procédure de test statistique», l'individu est en situation d'amalgame par rapport à K_1 et K_2 s'il affirme que d'une part le rôle d'un test statistique est de décider, moyennant un risque de se tromper entre H_0 et H_1, et d'autre part le niveau de signification α doit être fixé après collecte des données, dans une procédure de test statistique.

Par ailleurs, pour un ensemble donné d'individus, nous pouvons mesurer l'amalgame relative aux deux caractéristiques K_1 et K_2, par le biais du coefficient de corrélation ρ entre K_1 et K_2 ($\rho(K_1,K_2)$), où K_j (j=1 ou 2) pourrait être considérée comme étant une variable dichotomique de valeurs: F ou N , définie de la façon suivante:

$$K_j(i) = \begin{cases} F & \text{(si l'individu i adopte une conception fishérienne pour } K_j\text{)} \\ N & \text{(si l'individu i adopte une conception neyman-pearsonienne pour } K_j\text{)} \end{cases}$$

---

[6] **R**éponse **R**éussie.
[7] **R**éponse **E**chouée (y compris les **N**on **R**éponses).
[8] La réciproque est aussi vraie.
[9] et non neyman-pearsonienne.
[10] et non fishérienne.

Ainsi, plus $\rho(K_1,K_2)$[11] est faible, plus nous sommes convaincus que les individus, *d'une manière générale*, soient en situation d'amalgame entre la théorie de Fisher et celle de Neyman-Pearson par rapport aux deux caractéristiques: K_1 et K_2.

3.4.2 Application à la présente étude

Les caractéristiques évoquées dans notre étude et qui sont susceptibles d'impliquer un amalgame entre la théorie de Fisher et celle de Neyman-Pearson sont: K_1: «Rôle d'un test statistique» (Rôle), K_2: «Localisation du niveau de signification α dans une procédure de test statistique» (Loc(α)) et K_3: «Résultat d'un test statistique» (Résultat). Elles sont respectivement identifiées par les couples d'items (A1, A2), (B2, B1) et (C2, C1) du questionnaire.

Afin de mettre en évidence un tel amalgame par rapport à deux caractéristiques quelconques K_j et K_m (j et m $\in \{1,2,3\}$ et $j\neq m$), nous sélectionnons tout d'abord les individus[12] caractérisés par les modalités suivantes: $[(X_jR, Y_jE)$ ou $(X_jE, Y_jR)]$ et $[(X_mR, Y_mE)$ ou $(X_mE, Y_mR)]$[13], avec (X_j, Y_j) et (X_m, Y_m) sont respectivement les couples d'items du questionnaire identifiant les caractéristiques K_j et K_m. Nous pouvons définir alors les variables dichotomiques K_j et K_m de la manière suivante:

$$K_j(i) = \begin{cases} F & \text{(si l'individu i correspond aux modalités } (X_jR, Y_jE)) \\ N & \text{(si l'individu i correspond aux modalités } (X_jE, Y_jR)) \end{cases}$$

$$K_m(i) = \begin{cases} F & \text{(si l'individu i correspond aux modalités } (X_mR, Y_mE)) \\ N & \text{(si l'individu i correspond aux modalités } (X_mE, Y_mR)) \end{cases}$$

Le tableau 4 suivant présente les valeurs (F ou N), les individus et les coefficients de corrélation relatifs aux couples de caractères (Rôle, Loc(α)), (Rôle, Résultat) et (Loc(α), Résultat), impliqués par notre étude.

Tableau 4: le tableau résume les réponses des sujets et coefficients de corrélation relatifs à l'amalgame.

Individus	Rôle	Loc(α)
V4	N	N
V9	F	N
V11	N	F
V17	F	F
V19	F	N
V22	N	N
V25	N	N
V28	F	N
V30	N	N
V31	F	N
R35	N	F
R36	N	N
R39	F	N
R40	N	N
R41	N	F
R42	F	N
R45	N	N
R48	N	N
R49	N	N
R64	N	N
R65	N	F
R67	N	N
R69	N	F
R70	N	F

ρ (Rôle, Loc(α)) = **- 0,21**

Individus	Rôle	Résultat
V2	F	F
V6	F	N
V9	F	N
V11	N	N
V17	F	N
V18	F	N
V19	F	N
V24	N	N
V25	N	N
V28	F	N
V30	N	N
R34	N	N
R35	N	N
R36	N	N
R37	F	N
R38	F	N
R39	F	N
R40	N	N
R41	N	N
R45	N	N
R48	N	N
R49	N	N
R52	N	N
R60	F	F
R62	N	N
R67	N	N
R69	N	N
R70	N	N

ρ (Rôle, Résultat) = **0,345**

Individus	Loc(α)	Résultat
V1	F	F
V7	F	N
V9	N	N
V10	F	N
V11	F	N
V14	F	N
V16	N	N
V17	F	N
V19	N	N
V20	N	N
V21	N	N
V25	N	N
V27	N	N
V28	N	N
V30	N	N
V32	N	N
R33	F	N
R35	F	N
R36	N	N
R39	N	N
R40	N	N
R41	F	N
R43	N	F
R45	N	N
R48	N	N
R49	N	N
R50	N	N
R51	F	N
R59	N	N

[11] Vous pouvez coder les modalités F et N respectivement par 1 et 0, afin de calculer ce coefficient de corrélation.
[12] Dans la présente étude, les 32 élèves ingénieurs de l'Institut Vétérinaire et Agronomique de Rabat sont codés de V1 à V32 et les 38 étudiants de la Faculté des Sciences de Rabat-Agdal sont codés de R33 à R70.
[13] (X_jR, Y_jE) indique que l'individu a correctement répondu à l'item X_j et a échoué de répondre à l'item Y_j.

R67	N	N
R68	N	N
R69	F	N
R70	F	N

ρ (Loc(α), Résultat) = **0,072**

Nous remarquons que ces trois coefficients de corrélation sont significativement faibles. Par conséquent, nous pouvons conclure, d'une manière générale, que nos sujets sont en situation d'amalgame entre la théorie de Fisher et celle de Neyman-Pearson, par rapport aux trois caractéristiques suivantes: «Rôle d'un test statistique», «Localisation du niveau de signification α dans une procédure de test statistique» et «Résultat d'un test statistique».

4. CONCLUSIONS

Plusieurs auteurs ont rapporté que la base théorique des tests statistiques est largement male comprise. Ils ont argumenté que les conceptions des étudiants relatives à l'inférence statistique sont incohérentes parce qu'elles dérivent de deux théories mutuellement incompatibles, d'une part celle de Fisher et de l'autre celle de Neyman- Pearson. Bien que la théorie de Neyman-Pearson se soit développée en dehors de celle de Fisher, il n'en reste pas moins qu'il y a eu beaucoup de débat entre ceux qui appuient la théorie de Fisher et ceux qui appuient celle de Neyman-Pearson autour de la nature même de l'inférence statistique. Ses conclusions étaient que ce qui est devenu institutionnel aujourd'hui ce n'était plus les statistiques fishériennes, mais un amalgame incohérent d'idées de Fisher et de Neyman-Pearson.

Ainsi, contrairement aux autres outils de l'inférence statistique, la théorie des tests statistiques a fait l'objet de multiples controverses et critiques et a suscité beaucoup de débats [15]. Ces controverses et critiques ont porté sur sa procédure d'utilisation par les chercheurs et praticiens, et sur l'interprétation des résultats. Face à cette situation, plusieurs auteurs et associations pédagogiques ont recommandé que cette théorie soit exclue de l'enseignement, ou au moins être accompagné d'autres méthodes statistiques telles les intervalles de confiances, la taille de l'effet, la réplication, la statistique bayésienne, etc. Cependant, et malgré les prétendus défauts décelés dans l'application de cette théorie, elle a été largement utilisée par le passé, continue de l'être aujourd'hui, et pourrait être utilisée pour plusieurs années encore.
Il serait toutefois injuste de controverser l'utilisation de cette théorie et de recommander son exclusion pour la simple raison que des scientifiques sont incapables de comprendre -ou ne veulent pas comprendre- proprement le rôle joué par cette théorie dans le processus global d'inférence scientifique, sa logique, ses limites ainsi que l'interprétation des résultats qu'elle pourrait apporter.

Néanmoins, nous pensons que la construction d'une ingénierie didactique en matière d'enseignement et d'apprentissage de la théorie des tests statistiques s'avère nécessaire. Elle doit introduire cette théorie comme une pièce intégrée dans une structure de méthodologie de recherche plus générale qui implique un plan expérimental adéquat, et prenant en considération le fait de ne pas présenter les tests statistiques comme un amalgame entre les tests de signification de Fisher et les tests d'hypothèses de Neyman-Pearson, mais comme deux théories totalement différentes, dont chacune a ses propres caractéristiques.

La mise en place de cette ingénierie didactique pourrait être, sans doute, un support de base pour éviter non seulement les controverses et les critiques non fondées contre l'utilisation de cette théorie, mais aussi les conceptions erronées et les difficultés d'apprentissage que rencontrent les étudiants et les chercheurs à tout âge et à tout niveau d'expertise.

5. REFERENCES

[1] Elm'hamedi Z. (2010), *Contribution à une ingénierie didactique pour l'enseignement et l'apprentissage des tests statistiques à l'université*. Doctorat. Université Sidi Mohammed Ben Abdellah- Faculté des Sciences Dhar El Mehraz, Fès, Maroc.
[2] Elm'hamedi, Z. (2014), Effets d'un apprentissage empirique sur la compréhension du concept de moyenne arithmétique. *Annales de Didactique et de Sciences Cognitives*, IREM de Strasbourg, 19, p 129 - 169.
[3] Fisher, R.(1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.
[4] Fisher, R. (1935), *The Design of Experiments*, Edinburgh: Oliver & Boyd.
[5] Fisher, R. (1956), *Statistical Methods and Scientific Inference*, Edinburgh: Oliver & Boyd.
[6] Neyman, J. et Pearson, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference, Part I, *Biometrika*. 1928; 20A: 175-240.
[7] Neyman, J. et Pearson, E.S. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*. 1933; Series A 231: 289-337.
[8] Gigerenzer, G. (1993), *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.
[9] Zaki, M. et Elm'hamedi, Z. (2006), Mesures de support facilitant une meilleure appréhension du concept de «tests statistiques». *Actes du Symposium International Forapeval_st2006.* Université Sidi Mohammed Ben Abdellah- Faculté des Sciences Dhar El Mehraz. Fès, 23-24 novembre 2006.
[10] Zaki, M. et Elm'hamedi, Z. Eléments de mesure pour un enseignement des tests statistiques. *Annales de Didactique et de Sciences Cognitives*, IREM de Strasbourg. 2009 ; 14 : 153-194.
[11] Denis, D. J. The Modern Hypothesis Testing Hybrid: R. A. Fisher's Fading Influence''. With Discussion by Michel Armatte, Bernard Bru, Michael Friendly, Jeff Gill, Ernest Kwan, Bruno Lecoutre, Marie-Paul Lecoutre, Jacques Poitevineau and Stephen Stigler, *Journal de la Société Française de Statistique*. 2004; 145 4: 5-68.

[12] Rouanet, H. (2000), Statistical practice revisited. *In* H. Rouanet, J.-M. Bernard, M.-C. Bert, B. Lecoutre, M.-P. Lecoutre & B. Le Roux, *New ways in statistical methodology: From significance tests to Bayesian inference* (2nd edition), 29-64, Peter Lang, Bern, SW.
[13] Carver, R. P. The case against significance testing, *Harvard Educational Review*. 1978; 48: 378-399.
[14]Poitevineau, J. *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, 1998.
[15] Zaki, M. et Elm'hamedi, Z. Aspects de quelques critiques non fondées de la théorie des tests statistiques. *Annales de Didactique et de Sciences Cognitives*, IREM de Strasbourg. 2013 ; 18 : 139-171.
[16] Morrisson, D. E. et Henkel, R. E. (Eds). *The Significance Tests Controversy*, A reader. Chicago: Aldine, 1970.

Cite this article: Zahid El M'hamedi. LA LOGIQUE HYBRIDE DANS L'UTILISATION DES TESTS STATISTIQUES. *Am. J. innov. res. appl. sci. 2019; 9(5): 343*-351.