QUELLE EST VOTRE DÉFINITION DE « POURCENTAGE » ? PROPOSITION POUR L'ENSEIGNEMENT

Jeanne FINE[1]

Dans un article récent du Monde Magazine (30 mars 2013), intitulé « Les maths, nouvelle langue morte ? », Maryline Baumard rappelle les difficultés dans lesquelles se trouvent nos hommes et femmes politiques face à des questions relevant de la « règle de trois », des pourcentages ou même simplement des tables de multiplication :

> « Un coup d'œil au niveau en maths de nos hommes politiques devrait cependant rassurer plus d'un parent sur les chances de carrière de ses enfants. Et inquiéter plus d'un citoyen quant à la gestion de nos finances publiques. *« Si 10 objets identiques coûtent 22 euros, combien coûtent 15 objets ? »*, demande un journaliste de RMC à Luc Chatel, en juin 2011. Bravant le sens commun qui veut que deux produits valent en général plus cher qu'un seul, le ministre de l'éducation nationale répond 16,5 euros (au lieu des 33 euros attendus) à cet exercice de CM2. Trois mois plus tôt, Valérie Pécresse, en charge de l'enseignement supérieur et diplômée de HEC – temple des forts en maths –, oublie devant les caméras que l'on n'additionne pas des pourcentages et explique avec aplomb que lorsqu'un département augmente ses impôts de 30% et que sa région alourdit les siens de 58%, la facture du contribuable est majorée de… 88%. Mais que la gauche ne se gausse pas trop devant tant de confusion à droite. L'innumérisme est aussi développé dans son camp. Au point que Didier Migaud, le premier président de la Cour des comptes, autant dire le grand vérificateur des dépenses publiques, s'est joliment illustré en répondant (toujours sur RMC) que 7 × 9 = 76. Quant à Olivier Besancenot, du Nouveau Parti anticapitaliste, il a refusé de multiplier 8 par 9. »

Dans la mesure où, en statistique, nous utilisons les pourcentages pour les fréquences et pour les probabilités, savoir interpréter et utiliser les pourcentages est le premier niveau de la *littératie statistique*, à prendre en charge dès le collège. Dans un article intitulé « Mathématiques et économie : je t'aime, moi non plus », Bonneval (2003) écrit :

> « Nous savons tous la difficulté que présente pour les élèves (et pour la majorité des adultes) une pratique intelligente des pourcentages. Pourquoi un tel pathos autour d'une notion très simple, qui a été étudiée au collège ? Je risque un diagnostic : c'est que nous l'enseignons mal.
> Les pourcentages ne sont rien d'autre qu'une façon d'écrire les nombres et ne méritent pas au lycée d'être étudiés pour eux-mêmes. Certes, ils sont utilisés couramment pour écrire les fréquences et les taux d'évolution. Mais ce sont les propriétés des fréquences et des taux d'évolution qui doivent être explicitées. Traiter ensemble ces deux notions très différentes, sous prétexte qu'elles s'écrivent toutes deux sous forme de pourcentage, conduit à des acrobaties de langage qui ne font qu'entretenir la confusion. »

[1] Statisticienne, Toulouse, jeanne.fine@gmail.com

Statistique et Enseignement, **3**(2), 87-97, http://www.statistique-et-enseignement.fr/

J'adhère totalement à ce propos : que ce soit au collège ou au lycée, la confusion entretenue autour des pourcentages est un obstacle majeur pour une formation en statistique (descriptive ou inférentielle). C'est l'objet de ce libre-propos.

1 Autour des définitions de « pourcentage »

1.1 Où est le problème ?

Le quart d'une population a moins de 20 ans. Quel est le « pourcentage » de la population ayant moins de 20 ans ?

a) Pour certains, le « pourcentage » est un abus de langage pour désigner la « proportion » ; on a les égalités :
$$\frac{1}{4} = 0{,}25 = \frac{25}{100} = 25\%.$$
La réponse est donc 25%.

b) Pour d'autres, le « pourcentage » est le nombre *p* tel que
$$\frac{p}{100} = \frac{1}{4}.$$
La réponse est donc 25.

1.2 Difficultés pour les élèves

Ce flou est gênant pour les élèves car ils peuvent être sanctionnés par des professeurs ayant des « définitions » différentes.

Ainsi, suite à un travail interdisciplinaire maths/SES, Boyer et Simon-Couineau (2006) écrivent :

> « En cours de maths, le calcul est : $\frac{14{,}1}{25{,}6} \sim 0{,}55 = 55\%$, résultat sanctionné en SES.
> En cours de SES, le calcul est : $\frac{14{,}1}{25{,}6} \times 100 \sim 55 = 55\%$, résultat sanctionné en maths.
> Le compromis trouvé est d'écrire pour le calcul :
> soit $\frac{14{,}1}{25{,}6} \sim 0{,}55$, soit $\frac{14{,}1}{25{,}6} \times 100 \sim 55$,
> avec la phrase de conclusion :
> le résultat est 55%. »

1.3 Un lien avec l'écriture des « grandeurs »

Pour la défense de ce professeur de SES, on peut constater qu'il utilise le signe % comme une unité de mesure :

« le rapport est, en pourcentage, $\frac{14{,}1}{25{,}6} \times 100 \sim 55 = 55\%$ ».

On peut noter en effet l'analogie avec l'écriture suivante : si un cycliste a parcouru 30 kilomètres en 2 heures,

« la vitesse, en *km/h*, est $\frac{30}{2} = 15\ km/h$ ».

Cette écriture, bien qu'incorrecte, ne présente pas d'ambigüité ; elle est incorrecte car il aurait fallu choisir entre les deux écritures suivantes :

- « la mesure de la vitesse, exprimée en *km/h*, est $\frac{30}{2} = 15$ », ou bien
- « la vitesse est $\frac{30\ km}{2\ h} = 15\ km/h$ ».

En effet, il faut choisir de travailler sur les mesures (nombres sans unités) ou sur les grandeurs, mais ne pas mélanger les deux registres. À propos de la « grandeur » vitesse et de sa « mesure », on lit en introduction du document ressource de collège sur « grandeurs et mesures», voir en référence Eduscol (2007) :

> « Ainsi, par exemple, la formule $v = \frac{d}{t}$ peut s'interpréter de deux manières. Ou bien *d*, *v* et *t* désignent des mesures (avec des unités convenables) des grandeurs que ces lettres évoquent directement : il s'agit alors d'un calcul purement numérique ; ou bien, comme c'est le cas dans de nombreuses disciplines et dans l'enseignement des mathématiques dans des pays voisins, les lettres désignent les grandeurs elles-mêmes et la formule $v = \frac{d}{t}$ constitue une définition de la vitesse, *indépendante des unités choisies*. Par exemple, la vitesse d'une balle de tennis lors du service d'un joueur est de 197 *km/h*. La formule $v = \frac{d}{t}$ permet d'obtenir facilement la conversion de cette vitesse en *m/s*, à l'aide du calcul suivant :
>
> $$v\ = 197\,\frac{km}{h} = \frac{197\ km}{1\ h} = \frac{197000\ m}{3600\ s} = \frac{197000}{3600}\ m/s,$$
>
> soit environ 54,7 *m/s*, résultat beaucoup plus significatif pour le spectateur. »

1.4 Rapprochement entre les deux définitions

Dans le cas de notre exemple introductif, nous n'avons pas d'unité de mesure de longueur ou de durée. Nous sommes en présence de « grandeurs discrètes », selon le vocabulaire utilisé par les auteurs du document ressource (p. 22). Pour ces « grandeurs discrètes » nous pouvons utiliser la mesure de dénombrement (ou d'effectif) et la mesure de fréquence : de même que « l'effectif » d'une population donnée peut être exprimé *en millions*, par exemple 25,6 millions, de même la « fréquence » de la population ayant moins de 20 ans peut être exprimée *en centième*, $25 \times 1/100$, l'écriture *en centième* étant équivalente à l'écriture *en pourcentage*, $25 \times \frac{1}{100} = 25\%$. On revient, par le biais de la mesure, à une écriture dans une certaine unité de mesure.

En résumé, pour notre exemple, *la fréquence* est 25 %, *la mesure de la fréquence en pourcentage* est 25, et *le pourcentage* est un abus de langage qui désigne pour les uns la fréquence (comprise entre 0 et 1) et pour les autres sa mesure en centième (comprise entre 0 et 100).

Dans un très intéressant *Que sais-je ?* intitulé « *Pourcentages et tableaux statistiques* », Novi (1998) écrit page 16, pour une fréquence de 16,3% : « la convention qui en découle est d'écrire 16,3 dans le cas descriptif et 16,3% ou 0,163 dans le cas opératoire, utilisant alors indifféremment le terme de pourcentage ou de proportion ». C'est bien cette imprécision que nous dénonçons. Nous proposons de parler de la proportion en pourcentage 16,3% dans le langage courant et dans la présentation des résultats mais d'utiliser l'écriture décimale de la proportion 0,163 dans les calculs.

Dans le paragraphe 2, nous présentons quelques notions, utilisées en statistique, qui s'expriment en pourcentage : fréquences, proportions, probabilités, taux, ainsi que leurs premières propriétés mais, auparavant, nous évoquons quelques problèmes de vocabulaire.

1.5 Des problèmes de vocabulaire

Proportion

Notons dès à présent que le mot « *proportion* » désigne deux notions différentes en mathématiques :

- en *statistique,* une « *proportion* », appelée aussi « *fréquence* », est un rapport d'*effectifs* d'un ensemble A par rapport à un ensemble de référence E contenant A : $f_E(A) = \frac{n(A)}{n(E)}$;
- dans l'étude de la *proportionnalité en mathématiques,* une « *proportion* » est l'égalité de deux rapports (recherche de la quatrième proportionnelle, produit en croix des termes d'une proportion…).

Même dans les programmes sur la proportionnalité, le mot *proportion* désigne indifféremment « un seul rapport » et « égalité de deux rapports ».

Taux

De même, lorsqu'il s'agit de dire que le salaire moyen d'une entreprise A est 20% supérieur à celui d'une entreprise B, comment nommer ce « 20% » : *taux d'évolution* ? *taux de variation* ? *variation relative* (à différencier de *variation absolue*) ?

Taux d'évolution ne convient pas puisqu'il ne s'agit pas d'évolution, *taux de variation* non plus car il a une autre définition en mathématiques, *variation relative* est couramment utilisée mais la disparition du mot taux est gênante. C'est la raison pour laquelle nous proposons de parler de *taux de comparaison* ; bien sûr, il faut préciser par rapport à quelle valeur de référence est pris ce taux. Le taux d'évolution (dans le temps) devient un cas particulier de taux de comparaison avec, par convention, la valeur de l'instant initial comme référence.

Pourcentages

Dans les programmes, ces différentes notions sont appelées « pourcentages ». Les fréquences ou proportions sont parfois appelées « pourcentages de la partie au tout », ou « pourcentages instantanés » sans doute pour les différencier des « pourcentages d'évolution ». Ces notions ne sont introduites que dans quelques sections de lycées… et ne sont donc pas considérées comme des connaissances minimales du futur citoyen, à acquérir au collège.

Il s'agit donc de nommer les concepts sans ambiguïté afin de réduire les difficultés de leur apprentissage. Le vocabulaire ainsi précisé devra être évidemment confronté à son utilisation dans le langage courant ; par exemple, le « taux » de chômage correspond en fait à une « proportion » (rapport du nombre d'actifs sans emploi sur le nombre total d'actifs).

2 Proposition pour l'enseignement au collège

2.1 Effectifs, fréquences, probabilités et leurs propriétés

Un même paragraphe, intitulé « Notations et raisonnement mathématiques (objectifs pour le lycée) » a été récemment introduit dans les programmes de lycée de tous les niveaux. Il me semble nécessaire de travailler dès le collège quelques éléments sur la logique et les ensembles.

Nous supposons dans ce qui suit comme prérequis une introduction à la logique et ses opérations : *et*, *ou*, *non*. Ce travail de redéfinition de mots du langage courant dans un contexte mathématique pourrait être réalisé en interdisciplinarité avec les professeurs de français. Nous supposons également comme prérequis une introduction aux ensembles et aux opérations : *intersection*, *union*, *complémentation*. Il n'est pas question d'introduire au collège le formalisme des « mathématiques modernes » mais d'introduire ces notions, essentielles en statistique, à partir d'exemples relevant de la statistique et en utilisant les diagrammes de Venn, les tableaux d'effectifs croisés et les arbres pondérés.

Dans ce qui suit, nous présentons de façon formelle les éléments qui sous-tendent le travail sur la statistique descriptive et les probabilités, nous revenons ensuite à l'utilisation des pourcentages. Cette présentation s'adresse essentiellement aux professeurs.

Vocabulaire de la statistique

Une *population* (statistique) est un ensemble fini E dont les éléments sont appelés *individus* (statistiques) ou *unités statistiques*.

Le nombre d'éléments (ou cardinal) de toute partie A de E est appelé *effectif* (ou taille) de A ; on le notera ici $n(A)$.

La *fréquence* (ou *proportion*) de A (relativement à E) est définie par : $f(A) = \frac{n(A)}{n(E)}$.

Si A est non vide, la fréquence relativement à A de B, ou *fréquence conditionnelle à A de B*, est définie par : $f_A(B) = \frac{n(A \cap B)}{n(A)} = \frac{f(A \cap B)}{f(A)}$.

Propriétés de la fréquence

On a les propriétés suivantes :

a) pour toute partie A de E, on a : $0 \leq f(A) \leq 1$;
b) $f(E) = 1$;
c) si A et B sont deux parties disjointes de E, alors $f(A \cup B) = f(A) + f(B)$ (additivité).

On reconnaît ici la définition axiomatique de la probabilité sur un ensemble fini.

Définition axiomatique de la probabilité sur un ensemble fini

Si E est un ensemble fini, une *probabilité* sur E est, par définition, une application définie sur l'ensemble des parties de E vérifiant les propriétés a), b) et c).

La *fréquence* (ou *proportion*) considérée comme application définie sur les parties de E est un cas particulier de *probabilité.* Dans le cadre de la modélisation probabiliste, les fréquences par rapport à une population ou les fréquences de réalisation d'une expérience sont couramment utilisées comme estimations des probabilités.

Bien d'autres propriétés sont vérifiées ; en fait, les propriétés a), b) et c) sont des axiomes à partir desquels sont déduites les autres propriétés. Cette définition de la probabilité est exclue des programmes de collège car considérée comme trop difficile. Pourtant, les approches « laplacienne » et « fréquentiste » des probabilités proposées au collège présentent des difficultés plus importantes que nous avons déjà évoquées dans un article précédent (Fine, 2010, p. 15).

Propriétés de la fréquence conditionnelle

Si A est non vide, la *fréquence conditionnelle à A* vérifie les propriétés a), b) et c). Il s'agit donc d'une probabilité sur E, appelée *probabilité conditionnelle à A*.

Si A et B sont deux parties quelconques de E (non vides et strictement incluses dans E), on a :

d) $f(B) = f(A \cap B) + f(\bar{A} \cap B)$, où $\bar{A}$ désigne le complémentaire de A ;
e) $f(A \cap B) = f(A) f_A(B)$;
f) $f(B) = f(A \cap B) + f(\bar{A} \cap B) = f(A) f_A(B) + f(\bar{A}) f_{\bar{A}}(B)$;
g) $f_B(A) = \frac{f(A \cap B)}{f(B)} = \frac{f(A) f_A(B)}{f(A) f_A(B) + f(\bar{A}) f_{\bar{A}}(B)}$ (formule de Bayes pour les fréquences).

Il est ainsi proposé d'introduire au collège les distributions d'effectifs et de fréquences (conjointes, marginales et conditionnelles) de deux variables qualitatives ; il s'agit d'un programme minimal de littératie statistique qui faciliterait considérablement l'introduction des probabilités. Il est essentiel de distinguer les fréquences $f(A \cap B)$, $f_A(B)$ et $f_B(A)$, et de travailler les différentes formulations en français de ces nombres, y compris leurs ambiguïtés. Nous donnons des exemples dans le paragraphe 3. Les opérations sur les ensembles et les applications numériques définies sur des ensembles sont des obstacles qui peuvent être travaillés sur les effectifs et fréquences avant l'introduction des probabilités.

2.2 Taux de comparaison, taux d'évolution et leurs propriétés

Définitions

On appelle *taux de comparaison* d'une quantité positive x entre deux situations A et B, relativement à A, le rapport (on suppose x_A non nul) :

$$\frac{x_B - x_A}{x_A}.$$

On appelle *taux d'évolution* d'une quantité positive x évoluant dans le temps entre les instants 0 et 1 (on suppose x_0 non nul) :

$$\frac{x_1 - x_0}{x_0}$$

(il s'agit d'un cas particulier du taux de comparaison où, *par convention*, la référence est l'instant initial).

Si t désigne le taux, alors le nombre m défini par $m = 1 + t$ est appelé *multiplicateur associé au taux t*.

Propriétés des taux de comparaison et taux d'évolution

Ces taux sont compris entre –1 et $+\infty$; les multiplicateurs associés sont donc compris entre 0 et $+\infty$.

Si une quantité évolue au taux t entre les instants 0 et 1 et au taux t' entre les instants 1 et 2, alors le taux d'évolution t'' entre les instants 0 et 2 vérifie : $1 + t'' = (1 + t)(1 + t')$. Si on note m, m' et m'' les multiplicateurs associés aux taux t, t' et t'', alors $m'' = m\,m'$. On n'additionne pas les taux, on multiplie les multiplicateurs associés.

Nous ne développons pas davantage ici l'application de cette notion au *taux d'intérêt d'un emprunt* et aux calculs financiers, mais nous ne résistons pas à donner ici un petit exemple d'application.

Dans la préface de la convention interministérielle pour la promotion de l'égalité des chances entre les filles et les garçons dans le système éducatif (BO n° 10 du 9/02/00), on lit :

> « Cette inégalité de carrière est soulignée par un écart moyen de rémunération entre hommes et femmes d'environ 25%. »

Comment faut-il entendre cette phrase ? Nous avons quatre interprétations différentes :

- Pour 100 euros de salaire moyen des hommes, les femmes gagnent en moyenne 25% en plus.
- Pour 100 euros de salaire moyen des femmes, les hommes gagnent en moyenne 25% en moins.
- Pour 100 euros de salaire moyen des femmes, les hommes gagnent en moyenne 25% en plus.
- Pour 100 euros de salaire moyen des hommes, les femmes gagnent en moyenne 25% de moins.

Nous éliminons les deux premières interprétations car nous savons que les hommes gagnent en moyenne davantage que les femmes ; nous éliminons la troisième interprétation car, *par convention*, les hommes sont pris en référence. C'est donc la quatrième interprétation que nous retenons implicitement avec raison mais, pour mieux faire apparaître les inégalités, un énoncé équivalent pourrait être proposé, en prenant les femmes en référence : « Cette inégalité de carrière est soulignée par une rémunération moyenne des hommes supérieure d'environ 33% à celle des femmes. ».

2.3 Où sont passés les pourcentages ?

Les pourcentages sont utilisés pour exprimer des fréquences, proportions, probabilités, taux de comparaison, taux d'évolution. Nous avons donc défini ces notions et donné quelques propriétés. Comme déjà dit, nous conseillons de faire les preuves et les calculs sur les nombres écrits sous formes fractionnaire ou décimale et de n'utiliser les pourcentages que dans la présentation des résultats. Pour les taux négatifs, on indiquera dans la présentation du résultat « en moins », ou « baisse de », ou « diminution »…

Nous proposons de n'enseigner aucune règle de calcul sur les pourcentages. C'est en effet le formalisme sur les pourcentages qui a conduit aux difficultés que nous connaissons.

Les notations « p % » ou « f % » ou « t % » pour des *proportions* ou *probabilités*, pour des *fréquences* ou pour des *taux* induisent en erreur car elles laissent penser que la *proportion* ou la *probabilité*, la *fréquence* ou le *taux* sont les numérateurs des rapports sur 100.

Même la notation « a % » me semble parfois discutable. Présenter le pourcentage comme étant la recherche du numérateur du rapport sur 100 ramène aux difficultés de la proportionnalité alors que présenter l'écriture en pourcentage comme un changement d'unité de mesure du nombre décimal correspondant est relativement simple. Cela nécessite évidemment de travailler avec les élèves sur les différentes écritures d'un même nombre, en y incluant bien sûr le rapport sur 100 :

$$\frac{1}{4} = 0{,}25 = \frac{25}{100} = 25\%.$$

Néanmoins, certains professeurs préfèrent insister sur ce passage à un dénominateur égal à 100 ; c'est une façon de donner du sens à la nécessité de passer, par exemple, des effectifs aux fréquences. C'est la pratique en sociologie : dans un encadré, page 12, du livre de Novi (1998) déjà cité, on lit : « Calculer des pourcentages, c'est raisonner toutes choses égales par ailleurs en ce qui concerne les effectifs totaux. En les ramenant à 100, c'est faire comme si ces effectifs totaux n'avaient pas varié ».

Pour revenir aux programmes du secondaire, dans le programme de 6$^{\text{ème}}$, on lit en page 13, dans le paragraphe *pourcentage*, « appliquer un taux de pourcentage » et en page 15, dans le paragraphe *fraction*, « prendre une fraction d'une quantité ». Prendre 25% ou le 1/4 d'une quantité n'a pas à faire l'objet de deux paragraphes différents. Qui plus est, lorsque les taux sont introduits au lycée, apparaît l'expression « pourcentage du taux ». C'est à ce niveau que l'on étudie comment calculer un pourcentage de pourcentage mais que pour certains pourcentages on n'ajoute pas les pourcentages, on multiplie les multiplicateurs associés. Ce sont ces expressions obscures et ces procédures calculatoires confuses qui sont sources de difficultés.

3 Pourquoi ce libre-propos sur les pourcentages ?

Étant particulièrement vigilante sur cette question, j'ai mon florilège d'exemples marquants sur les pourcentages ; je me limiterai à deux expériences qui sont à l'origine de ce libre-propos.

Exemple 1

L'exercice suivant a été posé dès la rentrée à des étudiants préparant le concours de professeurs de mathématiques (ces étudiants ont une licence de mathématiques ou un diplôme équivalent) :

> « Les répartitions des candidats reçus au baccalauréat général dans l'académie de Toulouse en 2000 sont présentées dans les deux tableaux suivants : pour chaque sexe selon les séries et pour chaque série selon le sexe. Calculez la proportion de filles parmi les candidats reçus. »

Série \ Sexe	**Filles**	**Garçons**
Econom. et Social	27%	22%
Littéraire	32%	9%
Scientifique	41%	69%
Ensemble	*100%*	*100%*

Série \ Sexe	**Filles**	**Garçons**	***Ensemble***
Econom. et Social	64%	36%	*100%*
Littéraire	83%	17%	*100%*
Scientifique	46%	54%	*100%*

Les étudiants ont été très décontenancés ; ils ne parvenaient pas à poser correctement le problème. En cours de recherche, une aide a été apportée : il est possible de répondre à la question en utilisant seulement les données concernant la série ES (soit les trois données 27%, 22% et 64%. Sur les soixante étudiants présents, un seul (un ingénieur de formation) a donné la réponse correcte après des calculs assez lourds sur des effectifs rapportés à une population fictive de 10 000 candidats reçus.

Voici une proposition de solution.

L'ensemble de référence est l'ensemble des candidats reçus. On note p la proportion de filles. La proportion des ES s'écrit alors $0{,}27p + 0{,}22(1-p)$ en fonction de p. De même, la proportion relativement aux ES des filles s'écrit $\frac{0.27p}{0{,}27p+0{,}22(1-p)}$ et cette proportion est égale à 0,64. On en déduit que la proportion de filles parmi les reçus est 59%.

Commentaires.

La réponse tient en quatre lignes ; il suffisait de poser l'inconnue et de trouver une équation pour obtenir la proportion cherchée. Pour leur défense, ces étudiants de mathématiques n'ont jamais eu, durant leurs études du secondaire et du supérieur, à travailler sur les tableaux d'effectifs croisés, mais ils savent interpréter les deux tableaux présentés et ils savent formaliser et résoudre des problèmes bien plus difficiles. Il faut donc trouver une autre explication. Lors d'une discussion à ce sujet, une étudiante a parlé de la présentation confuse et contradictoire des pourcentages dans le secondaire. Elle préférait de loin les mathématiques formelles et était complètement déstabilisée dès qu'apparaissait un pourcentage dans un texte !

C'est sur cet exemple que nous proposons de donner quelques formulations de ces fréquences dans le langage courant (on préfèrera du reste parler de proportions plutôt que de fréquences) : 27% des filles reçues sont en ES, 64% des candidats reçus de ES sont des filles ; la proportion de filles parmi les candidats reçus est de 59% d'après nos calculs. On en déduit que la proportion de filles de ES parmi l'ensemble des candidats reçus est d'environ 16%. Nous avons ici trois variables catégorielles, « sexe » (F, H), « série » (ES, L, S) et « résultat » (reçu, collé) et il est possible à partir des données de reconstituer le tableau des effectifs croisés « sexe » × « série » de l'ensemble des candidats, puis le tableau similaire du sous-ensemble des candidats reçus.

Exemple 2

Le deuxième exercice proposé repose sur une expérience encore plus stupéfiante.

> « Un laboratoire pharmaceutique souhaite mettre sur le marché un nouveau traitement contre l'acné. Le traitement concurrent, utilisé depuis de nombreuses années, amène une réduction du nombre de boutons après un mois de traitement de 30% (observée par les dermatologues selon un même protocole). Le nouveau traitement est expérimenté sur 330 individus et amène une réduction du nombre de boutons après un mois de traitement de 35% (réduction observée dans les mêmes conditions). Le nouveau traitement est significativement meilleur que l'ancien au seuil de 5% car :
>
> $$\frac{0{,}35-0{,}30}{\sqrt{\frac{0{,}30\times 0{,}70}{330}}} = 1{,}98 > 1{,}96.$$
>
> Qu'en pensez-vous ? »

Réponse : ce raisonnement est incorrect car c'est un test de comparaison d'une *proportion* à une valeur donnée qui est proposé ici alors que la statistique observée est un *taux d'évolution* du nombre de boutons.

Voici le contexte dans lequel j'ai été amenée à découvrir ce problème. Un collègue biophysicien me soumet trois articles publiés dans des revues médicales ayant pour objectif de comparer deux molécules pour le traitement dermatologique de l'acné. Le critère de comparaison est le « taux de diminution » du nombre de boutons sur une partie déterminée du visage. Dès la lecture du premier article, je soupçonne l'erreur mais les explications méthodologiques sont évasives et ne permettent pas de vérifier quel test a été utilisé. Ce n'est que dans le dernier article qu'est indiquée la valeur de la statistique du test. Il s'agit bien, comme je le soupçonnais, du test de comparaison d'une proportion à une valeur donnée alors que l'on doit comparer ici des taux.

La référence en « statistique médicale et biologique » en France est le très bon manuel de Lazar et Schwartz (1964), plusieurs fois réédité depuis 1964. Dans ce livre, les auteurs présentent les fluctuations d'échantillonnage d'un *pourcentage*, la précision d'un *pourcentage*, la comparaison d'un *pourcentage* observé à un *pourcentage* théorique… Il s'agit en fait de *proportions* mais les auteurs des trois articles sur l'acné ont considéré qu'ils avaient bien à comparer leur *pourcentage* à une valeur donnée et que ce test leur permettait de le faire.

Les articles sont signés de nombreux collaborateurs ; aucun des membres de ces trois équipes ni même les experts chargés de rapporter sur les articles n'ont vu le problème. Qui plus est, les données permettaient (en faisant des hypothèses assez faibles sur les variances non disponibles) de conclure à une supériorité significative d'une molécule sur l'autre. Les enjeux financiers étant énormes, il est stupéfiant de trouver des erreurs méthodologiques aussi importantes.

Les débats sur des erreurs méthodologiques en statistique qui émergent parfois dans les médias ont déjà été évoqués en fin de la troisième contribution de ce numéro (cf. Fine, 2012).

En conclusion :

- diffusons les ambiguïtés du mot « pourcentage » et précisons le cadre dans lequel il est employé : ce peut être une proportion, une fréquence ou une probabilité (comprise entre 0 ou 1 et vérifiant les propriétés d'une mesure), ce peut être un taux de comparaison ou d'évolution (jamais inférieur à –1 mais pouvant être supérieur à 1 et vérifiant d'autres propriétés), ce peut être aussi la mesure en pourcentage d'un tel nombre, c'est-à-dire 100 fois ce nombre ;
- introduisons de façon cohérente dès le collège les notions d'effectifs, fréquences, probabilités, taux de comparaison et d'évolution.

Références

[1] Bonneval, L.-M. (2003), Mathématiques et économie : je t'aime, moi non plus, *Repères-IREM*, *Topiques Editions*, 52, 5-28.
En ligne : http://www.univ-irem.fr/commissions/reperes/consulter/52bonneval.pdf

[2] Boyer, E. et C. Simon-Couineau (2006), Une expérience de travail interdisciplinaire Maths-SES en lycée, *Bulletin de l'APMEP (Association des Professeurs de Mathématiques de l'Enseignement Public)*, 462, 33-50.

[3] Eduscol, Mathématiques, Ressources pour faire la classe au collège et au lycée, http://eduscol.education.fr/cid45766/ressources-pour-faire-la-classe.html#lien1

[4] Eduscol (2007), Ressources pour le collège, « Grandeurs et mesures », http://cache.media.eduscol.education.fr/file/Programmes/16/9/doc_acc_clg_grandeurs_109169.pdf

[5] Fine, J. (2010), Probabilités et statistique inférentielle : approche sondage versus approche modèle, *Statistique et Enseignement*, **1**(2), 5-21.

[6] Fine, J. (2012), Statistique, informatique, mathématiques et interdisciplinarité, *Statistique et Enseignement*, **3**(2), 33-59.

[7] Novi, M. (1998), *Pourcentages et tableaux statistiques*, PUF, Coll. Que sais-je ?, n° 3337.

[8] Lazar, P. et D. Schwartz (1964), *Éléments de probabilités et statistique à l'usage des étudiants en biologie humaine et générale*, Flammarion (1re édition).