

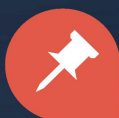
Le bulletin de l'APMEP - N° 546

AU FIL DES MATHS

de la maternelle à l'université...

Édition Octobre, Novembre, Décembre 2022

Maths et élèves à besoins particuliers (2)



APMEP

Association des Professeurs de Mathématiques de l'Enseignement Public

ASSOCIATION DES PROFESSEURS DE MATHÉMATIQUES DE L'ENSEIGNEMENT PUBLIC

26 rue Duméril, 75013 Paris

Tél. : 01 43 31 34 05

Courriel : secretariat-apmep@orange.fr - Site : <https://www.apmep.fr>

Présidente d'honneur : Christiane ZEHREN



Au fil des maths, c'est aussi une revue numérique augmentée :
<https://afdm.apmep.fr>

version réservée aux adhérents. Pour y accéder connectez-vous à votre compte via l'onglet **Au fil des maths** (page d'accueil du site) ou via le QRcode, ou suivez les logos ▶.

Si vous désirez rejoindre l'équipe d'**Au fil des maths** ou bien proposer un article, écrivez à
aufildesmaths@apmep.fr

Annonces : pour toute demande de publicité, contactez Mireille GÉNIN mcgenin@wanadoo.fr

ÉQUIPE DE RÉDACTION

Directrice de publication : Claire PIOLTI-LAMORTHE.

Responsable coordinatrice de l'équipe : Cécile KERBOUL.

Rédacteurs : Vincent BECK, François BOUCHER, Richard CABASSUT, Séverine CHASSAGNE-LAMBERT, Frédéric DE LIGT, Mireille GÉNIN, Cécile KERBOUL, Valérie LAROSE, Alexane LUCAS, Lise MALRIEU, Marie-Line MOUREAU, Serge PETIT, Daniel VAGOST, Thomas VILLEMONTAIX, Christine ZELTY.

« **Fils rouges** » **numériques** : Gwenaëlle CLÉMENT, Nada DRAGOVIC, Fanny DUHAMEL, Laure ÉTÉVEZ, Marianne FABRE, Robert FERRÉOL, Cédric GROLLEAU, Louise GROLLEAU, Yann JEANRENAUD, Armand LACHAND, Agnès VEYRON.

Illustrateurs : Pol LE GALL, Olivier LONGUET, Sixtine MARÉCHAL.

Équipe T_EXnique : François COUTURIER, Isabelle FLAVIER, Philippe PAUL, François PÉTIARD, Guillaume SEGUIN, Sébastien SOUCAZE, Sophie SUCHARD, Michel SUQUET.

Maquette : Olivier REBOUX.

Correspondant Publimath : François PÉTIARD.

Votre adhésion à l'APMEP vous abonne automatiquement à Au fil des maths.

Pour les établissements, le prix de l'abonnement est de 60 € par an.

La revue peut être achetée au numéro au prix de 15 € sur la boutique en ligne de l'APMEP.

Mise en page : François PÉTIARD

Dépôt légal : Décembre 2022. ISSN : 2608-9297.

Impression : Imprimerie Corlet

ZI, rue Maximilien Vox BP 86, 14110 Condé-sur-Noireau



Paradoxe de Simpson et estimateurs biaisés

Quoi ? Des filles meilleures en arithmétique et géométrie que les garçons mais moins bonnes en maths ? Comment est-ce possible ? C'est le paradoxe de Simpson : méfiance quand on regroupe des catégories pour les pourcentages !

Pierre Carriquiry

Un exemple

Le paradoxe de Simpson¹  peut apparaître lorsque l'on utilise des statistiques relatives à des sous-populations pour obtenir des résultats sur l'ensemble de la population. En voici un exemple tiré de l'article « L'embarrassant paradoxe de Simpson » de Jean-Paul Delahaye paru dans la revue *Pour la Science* n° 429 .

80 hommes et 80 femmes ont tous la même maladie. On choisit 60 hommes et 20 femmes qui prendront le même médicament et on donne un placebo à tous les autres. On obtient les résultats suivants :

Hommes	Guéris	Non guéris	Taux de guérison
Médicament	36	24	60 %
Placebo	14	6	70 %

Femmes	Guéries	Non guéries	Taux de guérison
Médicament	4	16	20 %
Placebo	18	42	30 %

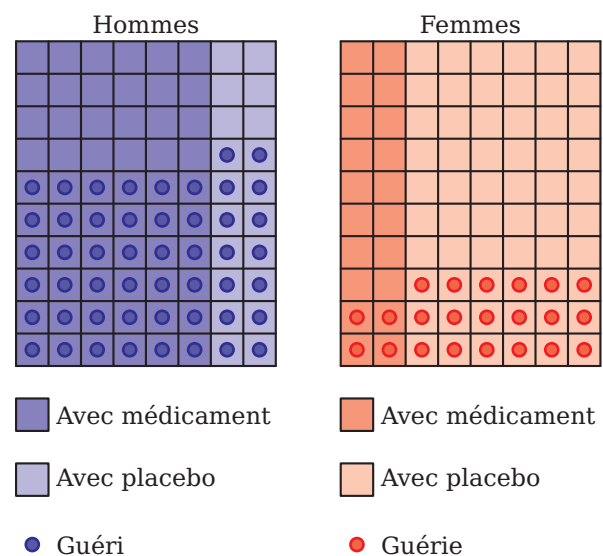
On constate que le placebo est plus efficace que le médicament chez les hommes (70 % contre 60 %) et chez les femmes (30 % contre 20 %).

Si on observe les résultats dans l'ensemble de la population, on obtient :

Ensemble	Guéris	Non guéris	Taux de guérison
Médicament	40	40	50 %
Placebo	32	48	40 %

Le médicament est plus efficace que le placebo sur l'ensemble de la population (50 % contre 40 %).

Les résultats sur les sous-populations et sur la population sont contradictoires. On a un paradoxe de Simpson. Cet exemple est peut-être virtuel (tous les pourcentages tombent juste) mais les lecteurs trouveront plusieurs exemples réels dans l'article de Jean-Paul Delahaye.



1. Edward Simpson est un statisticien britannique né le 10 décembre 1922 et mort le 5 février 2019 . À ne pas confondre avec Bart Simpson , ni avec Robert Simson (14 octobre 1687 — 1^{er} octobre 1768) .



Tentative d'explication

Les paradoxes sont parfois dus à des définitions imprécises. Par exemple, les paradoxes sur la théorie des ensembles sont probablement dus à une définition imprécise du mot ensemble. Essayons alors de préciser la signification de la proposition : *Le médicament est plus efficace que le placebo sur une population de n malades*. L'administration d'un médicament ou d'un placebo peut être considérée comme une expérience aléatoire qui a deux résultats : le malade guérit ou ne guérit pas. On va donc supposer que l'on peut attribuer deux probabilités à chaque malade : la probabilité de guérir en prenant le médicament et la probabilité de guérir en prenant le placebo. **Pour un malade, le médicament est plus efficace que le placebo si la probabilité de guérir en prenant le médicament est supérieure à la probabilité de guérir en prenant le placebo et, pour une population de n malades, on dira que le médicament est plus efficace que le placebo si la moyenne des probabilités de guérir en prenant le médicament est supérieure à la moyenne des probabilités de guérir en prenant le placebo.**

Revenons à l'expérience et notons $h_1, \dots, h_{80}$ les probabilités de guérir des 80 hommes lorsqu'ils prennent le médicament, $g_1, \dots, g_{80}$ les probabilités de guérir des 80 hommes lorsqu'ils prennent le placebo, $f_1, \dots, f_{80}$ les probabilités de guérir des 80 femmes lorsqu'elles prennent le médicament et $w_1, \dots, w_{80}$ les probabilités de guérir des 80 femmes lorsqu'elles prennent le placebo.

Notons $h = \frac{1}{80} \sum_{i=1}^{80} h_i$, $g = \frac{1}{80} \sum_{i=1}^{80} g_i$, $f = \frac{1}{80} \sum_{i=1}^{80} f_i$

et $w = \frac{1}{80} \sum_{i=1}^{80} w_i$ les moyennes de ces probabilités.

En appliquant la définition donnée ci-dessus, le médicament est plus efficace que le placebo chez les hommes si $h > g$, chez les femmes si $f > w$ et dans l'ensemble de la population si

$$\frac{1}{160} \left[\sum_{i=1}^{80} h_i + \sum_{i=1}^{80} f_i \right] > \frac{1}{160} \left[\sum_{i=1}^{80} g_i + \sum_{i=1}^{80} w_i \right]$$

ce qui équivaut à : $0,5(h + f) > 0,5(g + w)$.

On voit que si $h > g$ et $f > w$, on a toujours $0,5(h + f) > 0,5(g + w)$. Le paradoxe de Simpson est impossible dans ce modèle mathématique. Malheureusement, ces 320 probabilités sont inconnues et l'expérience aléatoire nous permet seulement d'estimer certains paramètres. On pourrait alors se contenter de dire que le paradoxe s'explique par l'imprécision des estimations mais on va voir qu'il s'explique surtout par les formules de calcul des pourcentages.

Estimateurs

Étudions l'expérience aléatoire qui consiste à choisir 60 hommes parmi 80, 20 femmes parmi 80, à donner le médicament à ces 80 malades et le placebo aux 80 autres. On définit les variables aléatoires :

- $H_i = 1$ si l'homme i reçoit le médicament et guérit, 0 sinon ;
- $F_i = 1$ si la femme i reçoit le médicament et guérit, 0 sinon ;
- $G_i = 1$ si l'homme i reçoit le placebo et guérit, 0 sinon ;
- $W_i = 1$ si la femme i reçoit le placebo et guérit, 0 sinon.

La variable aléatoire $H = \frac{1}{60} \sum_{i=1}^{80} H_i$ représente la proportion d'hommes qui ont guéri par rapport à l'ensemble des hommes qui ont reçu le médicament. On définit de même :

$$F = \frac{1}{20} \sum_{i=1}^{80} F_i, G = \frac{1}{20} \sum_{i=1}^{80} G_i \text{ et } W = \frac{1}{60} \sum_{i=1}^{80} W_i.$$

Calculons les espérances mathématiques de ces variables :

$$E(H) = \frac{1}{60} \sum_{i=1}^{80} E(H_i) ;$$

$$E(H_i) = P(H_i = 1) = 0,75h_i$$

car la probabilité que l'homme i reçoive le médicament est $\frac{60}{80}$ et la probabilité qu'il guérisse sachant qu'il a reçu le médicament est h_i .



$$\text{D'où } E(H) = \frac{1}{60} \sum_{i=1}^{80} 0,75h_i = \frac{1}{80} \sum_{i=1}^{80} h_i = h.$$

L'espérance de la variable aléatoire H est égale au paramètre h que l'on veut estimer. On dit que H est un estimateur sans biais de h . On trouve de même que $E(F) = f$, $E(G) = g$, $E(W) = w$.

La variable aléatoire qui représente la proportion de guérisons dans l'ensemble des personnes qui ont reçu le médicament est

$$T = \frac{1}{80} \left[\sum_{i=1}^{80} H_i + \sum_{i=1}^{80} F_i \right].$$

Son espérance est

$$\begin{aligned} E(T) &= \frac{1}{80} E[60H + 20F] \\ &= 0,75E(H) + 0,25E(F) \\ &= 0,75h + 0,25f \end{aligned}$$

qui est différent du paramètre $0,5(h + f)$ que l'on veut estimer (sauf cas particuliers). On dit que T est un estimateur biaisé de $0,5(h + f)$.

De même la variable aléatoire

$$S = \frac{1}{80} \left[\sum_{i=1}^{80} G_i + \sum_{i=1}^{80} W_i \right]$$

représente la proportion de guérisons dans l'ensemble des personnes qui ont reçu le placebo. Son espérance est

$$\begin{aligned} E(S) &= \frac{1}{80} E[20G + 60W] \\ &= 0,25g + 0,75w \end{aligned}$$

qui est encore différent du paramètre $0,5(g + w)$ que l'on veut estimer (sauf cas particuliers). Ainsi, S est un estimateur biaisé de $0,5(g + w)$.

Si on applique ces résultats à l'expérience aléatoire on peut dire que :

- 0,6 et 0,2 sont des estimations non biaisées de h et f respectivement ;
- 0,7 et 0,3 sont des estimations non biaisées de g et w respectivement ;
- 0,5 est une estimation biaisée de $0,5(h + f)$;
- 0,4 est une estimation biaisée de $0,5(g + w)$.

On voit donc qu'il est dangereux de conclure que $0,5(h + f) > 0,5(g + w)$ en utilisant des estimations biaisées de ces paramètres car cela reviendrait à dire : « l'estimation (non biaisée) de $0,75h + 0,25f$ étant supérieure à celle de $0,25g + 0,75w$, on en déduit que $0,5h + 0,5f$ est supérieur à $0,5g + 0,5w$ ».

On remarque au passage que les proportions dans la population (0,5 et 0,4) sont des moyennes pondérées des proportions dans les sous-populations mais avec des coefficients de pondération différents :

$$\begin{aligned} 0,5 &= 0,75 \times 0,6 + 0,25 \times 0,2 \\ \text{et } 0,4 &= 0,25 \times 0,7 + 0,75 \times 0,3. \end{aligned}$$

Ce qui donne une explication non probabiliste du paradoxe : on compare des moyennes pondérées avec des coefficients de pondération différents.

On remarque aussi que si 50 % des hommes et 50 % des femmes avaient pris le médicament, il n'y aurait pas de paradoxe de Simpson.

En conclusion, si on adopte ce modèle mathématique, les hommes doivent utiliser les statistiques relatives aux hommes, les femmes doivent utiliser les statistiques relatives aux femmes et ne pas considérer les résultats globaux s'ils sont biaisés.

Si on effectue des partitions de chaque sous-population (voir les exemples de Double-Simpson dans l'article de Jean-Paul Delahaye où il distingue hommes aux yeux clairs, hommes aux yeux foncés, femmes aux yeux clairs, femmes aux yeux foncés), la règle de décision consiste à retenir la partition la plus fine. Par exemple, les hommes aux yeux clairs utiliseront les statistiques relatives aux hommes aux yeux clairs et non pas les statistiques relatives aux hommes ou celles relatives aux yeux clairs si elles sont biaisées.

Si les autorités sanitaires doivent prendre une décision binaire, c'est-à-dire autoriser le médicament pour l'ensemble de la population ou ne pas l'autoriser, on comprend qu'elles puissent hésiter en présence de paradoxes de Simpson. Mais il est possible d'autoriser un médicament



pour certains groupes de population : il existe des médicaments déconseillés voire interdits aux enfants, aux femmes enceintes, aux personnes âgées, etc.

Le paradoxe de Simpson ne se manifeste pas seulement dans le domaine médical. On peut aussi le rencontrer dans l'évaluation des élèves comme dans l'exemple (fictif) suivant.

Dans un collège de 80 garçons et 80 filles, on choisit 60 garçons et 20 filles auxquels on pose une question d'arithmétique, et on pose une question de géométrie à tous les autres. Pour chaque question il n'y a que deux possibilités : l'élève réussit ou échoue. Les résultats sont donnés dans les tableaux suivants :

Arithmétique	Réussite	Échec	Taux de réussite
Garçons	36	24	60 %
Filles	14	6	70 %

Géométrie	Réussite	Échec	Taux de réussite
Garçons	4	16	20 %
Filles	18	42	30 %

On constate que les filles ont de meilleurs résultats en arithmétique (70 % contre 60 %) et en

géométrie (30 % contre 20 %) mais si on considère l'ensemble des épreuves, les garçons ont de meilleurs résultats que les filles (50 % contre 40 %). Le paradoxe de Simpson nous amène à dire que les filles sont meilleures en arithmétique et en géométrie et que les garçons sont meilleurs en mathématiques (si l'arithmétique et la géométrie sont les seules matières enseignées).

Si on adopte le modèle mathématique précédent, on doit conclure que les filles sont meilleures que les garçons puisque les résultats d'ensemble correspondent à des estimations biaisées. Si on avait posé une question d'arithmétique et une question d'histoire on pourrait trouver bizarre de faire une statistique sur l'ensemble des deux matières.

Ainsi un modèle mathématique nous permet d'expliquer le paradoxe et de prendre des décisions qui ne contredisent pas l'intuition.

Les paradoxes de la théorie des ensembles ne nous ont pas chassés du paradis de Cantor et le paradoxe de Simpson ne nous chassera pas du paradis de Kolmogorov².



Pierre Carriquiry est aujourd'hui à la retraite ; il a enseigné à l'École Nationale de Commerce. Il est membre de l'APMEP depuis 35 ans.

carriquiry@free.fr

© APMEP Septembre 2022

2. Cette phrase fait écho à celle prononcée par David Hilbert : « Nul ne doit nous exclure du paradis que Cantor a créé. » et au fait que Kolmogorov est l'un des fondateurs de la théorie moderne des probabilités.



Sommaire du n° 546



Maths et élèves à besoins particuliers (2)

Éditorial	1	Ouvertures	49
Opinions	3	✦ Géométrie et élèves dyspraxiques — Ludivine Hanssen	49
L'âme vive de l'APMEP — Claire Piolti-Lamorthé et le bureau national de l'APMEP	3	Preuves visuelles — François Boucher	54
✦ Difficultés d'apprentissage en mathématiques ou dyscalculie ? — Marie-Line Gardes	5	Écart à l'indépendance d'événements : un encadrement remarquable — Jean-Baptiste Hiriart-Urruty	59
✦ Pour des élèves à HPI, comment soutenir le goût d'apprendre en mathématiques ? — Line Massé, Marie-France Nadeau & Claudia Verret	12	Paradoxe de Simpson et estimateurs biaisés — Pierre Carriquiry	62
Avec les élèves	19	Récréations	66
La résolution de problèmes au cœur des apprentissages — Marie-France Guissard, Valérie Henry, Pauline Lambrecht, Marie-Françoise Van Troeye & Isabelle Wettendorff (CREM)	19	Au fil des problèmes — Frédéric de Ligt	66
✦ Soutenons l'utilisation des doigts en mathématiques — Benoît Bonnet, Nathalie Bonneton-Botté, Hélène Hili, Sonia Jarry, Claire Labesse, Fanny Ollivier, Nolwenn Quelaudren & Nadège Saliot	28	JEUX-Écollège 5, une pépite de géométrie — Sophie Roubin	69
✦ Une séquence sur les angles en ULIS-collège — Claire Chantreuil	39	Au fil du temps	72
Zelliges pythagoriciens — Sébastien Reb	45	Charlotte Angas Scott — Roger Mansuy	72
		Des décimaux avant les décimaux ? — Michel Sarrouy	76
		Le CDI de Marie-Ange — Marie-Ange Ballereau	86
		Une ressource pour l'option mathématiques complémentaires — Charlotte Derouet	88
		Matériaux pour une documentation	91
		Courrier des lecteurs	95



CultureMATH

