

LOGIQUE & CALCUL

L'embarrassant paradoxe de Simpson

Les statistiques conduisent à des décisions fondées et rationnelles... à moins qu'un paradoxe ne s'en mêle, comme cela arrive avec certains jeux de données.

Jean-Paul DELAHAYE

Un bon paradoxe est un paradoxe dont on ne réussit jamais à se débarrasser. Quand vous croyez en avoir trouvé la clef, une remarque vous fait découvrir que rien n'est résolu. Les paradoxes de Zénon à propos de l'impossibilité du mouvement sont de tels paradoxes. Mais le plus élémentaire de tous est le paradoxe de Simpson dont on imagine des solutions... qui conduisent à d'autres paradoxes !

Sans cesse, des scientifiques et des utilisateurs de statistiques tombent dans les pièges qu'il tend. Chaque année, paraissent des articles qui tentent de déterminer son sens profond et la façon dont on doit le traiter. Malgré cette littérature abondante, il n'est pas certain que l'on détienne une solution entièrement satisfaisante pour se libérer de cette récalcitrante absurdité.

Imaginons la situation suivante. On mène des tests en double aveugle sur un nouveau médicament traitant la maladie grave MG. On a traité 160 patients, dont 80 ont reçu le médicament, et les 80 autres un placebo.

Le taux de guérison varie selon que l'on considère les malades ayant pris le médicament ou ceux ayant pris le placebo (voir le tableau 1 ci-contre). Parmi les 80 patients ayant pris le médicament, 40 ont été guéris (50 %). Parmi les 80 patients ayant reçu le placebo, seuls 32 ont été guéris (40 %). Ces résultats suggèrent que le médicament est efficace. Il faut donc le prescrire pour soigner les patients atteints de la maladie MG.

Mais en analysant plus en détail les données et en considérant le sexe des personnes

ayant participé aux tests, on a une surprise : parmi les hommes, le placebo réussit mieux que le médicament, et il en va de même parmi les femmes. La somme des deux tableaux 2 et 3 redonne bien le tableau 1.

Total	Guéri	Non guéri	Taux de guérison
Médicament	40	40	50 %
Placebo	32	48	40 %

Hommes	Guéri	Non guéri	Taux de guérison
Médicament	36	24	60 %
Placebo	14	6	70 %

Femmes	Guéri	Non guéri	Taux de guérison
Médicament	4	16	20 %
Placebo	18	42	30 %

Les trois tableaux sont compatibles, aucune erreur ne s'est produite. Les résultats sont sans appel : chez les hommes, le placebo est meilleur que le médicament ; chez les femmes, le placebo est meilleur que le médicament. Pourtant, en regroupant hommes et femmes, le médicament produit de meilleurs résultats que le placebo. On se trouve dans un cas du paradoxe de Simpson : la fusion de données concluant individuellement dans un sens – l'inutilité du médicament testé – donne des résultats concluant dans le sens inverse, l'utilité du médicament.

La plus lointaine mention d'un cas analogue remonte à 1899, où le mathématicien anglais Karl Pearson décrit des données équivalentes. Plus tard, en 1903, Undy Yule redécouvrit le phénomène, et le Britannique

Edward Simpson (né en 1922) écrivit en 1951 un article où cette singularité statistique était soigneusement étudiée et discutée. Assez injustement aujourd'hui, ce type de données paradoxales porte son nom.

De nombreux cas réels présentent cette inversion de résultat lorsqu'on regroupe plusieurs catégories complémentaires en une seule. Chaque année, on découvre de nouveaux exemples produisant étonnement et incrédulité. On a rencontré le paradoxe à propos des taux d'admission des filles dans divers départements de l'Université de Berkeley en 1973 : elles étaient meilleures que les garçons dans la plupart des départements de l'université, mais quand on fusionnait les résultats, leur taux général d'admission à l'université était inférieur à celui des garçons. De nombreux cas en médecine ont été rapportés. Le paradoxe a aussi été rencontré à propos de l'analyse des matchs de basket-ball, en démographie, dans l'étude des risques d'accidents, etc. (voir l'encadré 1).

Explication arithmétique

Avant de tenter de savoir s'il faut oui ou non prescrire le médicament testé dans le cas des tableaux 1-3, étudions l'arithmétique du paradoxe. Si l'on note A, B, C et D les quatre nombres du premier tableau, a, b, c et d les quatre nombres du second tableau, a', b', c' et d' les quatre nombres du troisième tableau, on a les relations suivantes :

$$A = a + a'; B = b + b'; C = c + c'; D = d + d'; \\ A/B > C/D; a/b < c/d; a'/b' < c'/d'.$$

L'étonnement vient de ce que l'on croit que $[a/b < c/d \text{ et } a'/b' < c'/d']$ entraîne $A/B < C/D$, ou encore avec seulement les petites lettres: $[a/b < c/d \text{ et } a'/b' < c'/d']$ entraîne $(a+a')/(b+b') < (c+c')/(d+d')$.

Or aucune démonstration n'établira cette implication entre inégalités arithmétiques, puisque justement les nombres des tableaux 1-3 vérifient les deux premières inégalités et pas la troisième: $36/24 < 14/6$ et $4/16 < 18/42$, mais $40/40 > 32/48$ (soit $1,5 < 2,33$ et $0,25 < 0,42$, mais $1 > 0,66$).

Sur le plan arithmétique, il n'y a pas le moindre paradoxe – seulement une légère surprise dont on se libère en analysant l'exemple qui nous montre que même si $a/b < c/d$ et $a'/b' < c'/d'$, il ne faut pas en conclure que $(a+a')/(b+b') < (c+c')/(d+d')$.

Une représentation graphique des données (voir l'encadré 2) illustre pourquoi il ne faut pas s'étonner de rencontrer des jeux de huit paramètres produisant le «paradoxe». Cette représentation conduit d'ailleurs à découvrir des cas encore plus simples de jeux de données présentant la prétendue anomalie. En imposant aux variables d'être des entiers entre 1 et 4, on trouve pour les huit paramètres quatre solutions qui sont: $\{1, 2, 3, 4, 4, 2, 3, 1\}$ $\{1, 3, 2, 4, 4, 3, 2, 1\}$ $\{4, 2, 3, 1, 1, 2, 3, 4\}$ $\{4, 3, 2, 1, 1, 3, 2, 4\}$.

Lorsqu'on impose aux variables d'être des entiers entre 1 et 5, le nombre de solutions explose: il y en a 232. Pour l'intervalle de 1 à 6, il y en a 1 370. Pour les intervalles suivants, le nombre de solutions est 8 126, puis 28 252, puis 86 140, etc. La propor-

tion de jeux de huit entiers pris entre 1 et n vérifiant les inégalités du paradoxe (ou les inégalités inverses qui sont aussi paradoxales) tend vers 0,9606 % quand n tend vers l'infini.

Comment raisonner ?

Tout n'est cependant pas réglé pour autant et, dans la réalité, un médecin face aux données des trois tableaux indiqués doit prendre une décision: oui ou non, faut-il prescrire le médicament qui semble efficace (d'après le tableau 1) et qui semble moins bon que le placebo (d'après les tableaux 2 et 3) ?

Imaginez-vous à la place du médecin. Comment allez-vous raisonner ? Plusieurs attitudes sont possibles.

1. Trois exemples réels

Pris parmi des dizaines d'autres, voici trois cas réels où le paradoxe de Simpson a produit des surprises et a contraint à analyser plus finement ce que des manipulations statistiques imprudentes faisaient apparaître.

Aptitudes en progrès pour tous, mais dont la moyenne est stable.

Aux États-Unis, au cours des années 1981-2002, la moyenne des notes obtenues par les élèves soumis au test linguistique SAT (*Scholastic Aptitude Test*) a été stable, avec un résultat toujours proche de 504 points. Pourtant, durant cette période, la moyenne pour le test mesuré séparément par groupe ethnique s'est accrue dans chaque groupe d'au moins huit points.

Gerald Bracey, de l'Université George Mason, a étudié cet étrange phénomène et a conclu qu'on était

là en présence d'un cas du paradoxe de Simpson.

La clef de l'explication se trouve dans le fait que pendant la période considérée, le nombre d'élèves blancs soumis au test a proportionnellement décliné, et que la proportion d'élèves appartenant à des minorités (réussissant en général moins bien le test linguistique) a augmenté, conduisant au total à une stagnation des résultats globaux qui, pourtant, progressaient dans chaque catégorie.

Mauvais taux de mortalité dans chaque tranche d'âge, mais globalement bons.

En 1986, Joel Cohen, de l'Université Rockefeller à New York, a identifié un cas réel de paradoxe de Simpson dans le domaine de la démographie. Il s'est intéressé à la mortalité au Costa Rica et en Suède (connue

pour son excellente espérance de vie). Sans surprise, il a trouvé qu'en 1960, le taux de mortalité des femmes dans toute tranche d'âge fixée était supérieur au Costa Rica, comparé au taux équivalent en Suède. Pourtant, le taux de mortalité général des femmes au Costa Rica était inférieur à celui de la Suède.

L'explication est encore liée au paradoxe de Simpson, et elle provient de la structure différente des populations. La population du Costa Rica est beaucoup plus jeune en moyenne que celle de la Suède, et donc les jeunes classes d'âge (qui ont un faible taux de mortalité) pèsent plus dans la moyenne pour le Costa Rica que pour la Suède, conduisant à un taux de mortalité global assez faible au Costa Rica, malgré un taux assez mauvais dans chaque tranche d'âge.

En roulant plus vite, vous aurez moins d'accidents.

Gary Davis, de l'Université du Minnesota, étudiait en 2004 la relation entre le nombre d'accidents piéton-véhicule et la vitesse moyenne des véhicules en divers endroits d'une ville.

Il proposa une modélisation de cette relation qu'il voulait utiliser pour évaluer l'intérêt de limiter plus sévèrement la vitesse autorisée des véhicules. De façon inattendue, le modèle montrait qu'en faisant passer la limite de 30 miles par heure à 25, le nombre d'accidents augmenterait.

Là encore, bien que pour chaque type de sites, le nombre d'accidents diminuait quand la vitesse limite était réduite, une agrégation malheureuse des données (qui ne prenait pas en compte que le nombre d'accidents était bien plus rare dans les zones résidentielles) engendrait une conclusion absurde.



Étudiants en Virginie



Rue au Costa Rica



Rue en Suède



Limitation de vitesse