

➤ description de l'incertitude qui est résumée ici par une distribution de probabilités. Des chercheurs peuvent collecter des données ou analyser des données qui ont déjà été publiées de façon créative (comme l'a fait Kanazawa) et publier leurs résultats. Des métaanalyses peuvent être faites pour revoir tous ces résultats conjointement; elles lisseront les variations qui sont inhérentes à ces études sur de petits échantillons dans lesquelles la probabilité d'un effet positif peut passer de 50% à 58%, puis peut-être redescendre à 38% et ainsi de suite.

Comment reconnaître les données fiables? En collectant toujours plus de données. Chaque année, le magazine américain *People* publie une liste des 50 plus belles célébrités mondiales. Nous avons répertorié le sexe de leurs enfants pour les numéros parus entre 1995 et 2000 (voir l'encadré page 30). En 1995, par exemple, on a dénombré 32 naissances de filles pour 24 garçons, soit 57,1% de filles, ce qui correspond à 8,6% de plus que dans la population générale (48,5%). Un résultat en accord avec l'hypothèse de Kanazawa. Mais l'écart-type étant de 6,7%, l'estimation de 8,6% n'est pas statistiquement significative. Pour le confirmer, nous avons comparé ce résultat avec ceux des années suivantes. Entre 1995 et 2000, les plus belles personnes selon *People* ont eu 157 filles sur un total de 329 enfants, soit 47,7% de filles (pour un écart-type de 2,8%), ce qui est seulement 0,8% inférieur au chiffre obtenu pour la population générale. On ne peut rien en conclure...

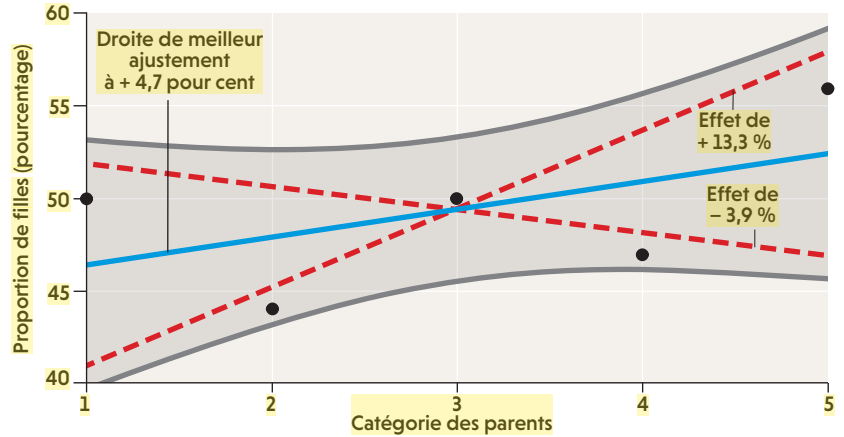
Pourquoi perdre notre temps à étudier des erreurs statistiques que personne n'a repérées? Pour deux raisons. D'abord, les résultats qui semblent avoir un sens sans être statistiquement significatifs sont les plus problématiques. Ensuite, certains médias et diverses publications scientifiques, par leur intérêt pour certains sujets de sociologie et leur sélection des résultats, biaisent la recherche en sciences sociales.

De fait, les résultats de Kanazawa ont tout de suite suscité l'intérêt des médias, jusqu'à des blogs du *New York Times*. Cette publication dans une revue à comité de lecture semblait être une caution suffisante balayant les doutes éventuels.

UN BRUIT ASSOUDISSANT

Qui plus est, l'effet n'a cessé d'augmenter. Ainsi, l'estimation – statistiquement non significative – de 4,7% que nous avons faite est passée à 8% dans l'analyse de Kanazawa (comparaison du groupe le plus beau à la moyenne des quatre groupes les moins séduisants), pour atteindre la valeur de 26% après une étude complémentaire introduisant d'autres corrections, avant de grimper à 36% pour des raisons encore floues!

Cette inflation nous surprend, ce chiffre étant 10 à 100 fois supérieur à tous les



Une analyse fréquentiste des données détermine la courbe de meilleur ajustement (la droite bleue) aux données de S. Kanazawa (points en noir), pour un écart-type de 4,3%. Un intervalle de confiance à 95% montre que des effets aussi faibles que -3,9% et aussi grands que 13,3% sont compatibles avec les données, car ils encadrent l'estimation à 4,7%.

rapports filles-garçons publiés dans la littérature. Nous en avons conclu que, dans cette étude, le bruit (les données parasites) était supérieur au signal pertinent. La puissance statistique désigne la capacité de détecter une différence lorsqu'elle existe. Les études avec des échantillons plus importants ont toujours plus de puissance. Ainsi, si l'on veut affirmer quelque chose à propos d'effets de l'ordre de 1%, on a tout intérêt à partir de données pertinentes et à réaliser des tests qui exploitent bien les données. Cet exemple illustre bien le fait que les études qui manquent de puissance statistique ont peu de chances de parvenir à une pertinence statistique et, plus important encore, elles surestiment la taille des effets. Autrement dit, avec ces études, le bruit devient plus fort que le signal, c'est-à-dire l'effet observé.

Comment échapper à ce type de problèmes en sociologie? Aujourd'hui, la plupart des sujets de sociologie ont été passés au crible, et les chercheurs en sont donc réduits à étudier les petits effets. L'étude du rapport des sexes à la naissance est un sujet de société proche de nos préoccupations. Présenté sous forme d'une «vérité politiquement incorrecte», le résultat de Kanazawa, parce qu'il concerne les naissances, touche à des questions sensibles telles que l'avortement, le congé parental, le rôle de l'homme et de la femme dans la société.

On a vu que les études dont la pertinence statistique est insuffisante produisent des résultats aléatoires, parfois statistiquement significatifs, mais le plus souvent intuitifs. C'est un des points faibles de la psychologie évolutionniste: elle interprète des résultats aléatoires sans reconnaître la fragilité des explications qu'elle donne. Par exemple: les personnes jugées séduisantes auraient plus de chances d'être en bonne santé, riches et issues de groupes ethniques dominants, et plus

DES PETITS EFFETS PASSÉS AU CRIBLE

Il est difficile de mettre en évidence un petit effet, car les données n'apportent toujours qu'une information limitée. Plus l'effet que l'on cherche à estimer est petit, plus la quantité de données nécessaire pour le mettre en évidence est grande pour qu'il ressorte des fluctuations dues au hasard. N'y a-t-il pas d'autre issue qu'une simple augmentation de la quantité de données, coûteuse et parfois impossible ? Autrement dit : « Comment exploiter au mieux un ensemble de données ? » Donnons quelques exemples. Dans un essai thérapeutique concernant une nouvelle préparation que l'on souhaite comparer à une ancienne, les patients sont souvent très hétérogènes. On constitue des paires de patients aussi proches que possible pour toutes les caractéristiques susceptibles d'influer sur le résultat (sexe, âge, catégorie socioprofessionnelle, origine ethnique...).

Pour éviter tout biais, on choisit au hasard dans chaque paire le patient qui reçoit le traitement à tester et celui qui reçoit l'ancien. La comparaison des deux traitements se fonde ainsi sur un ensemble de comparaisons élémentaires beaucoup plus efficaces que si l'on avait choisi les patients sans tenir compte de leur statut. Dans les sondages d'opinion, on sait que l'âge, la catégorie socioprofessionnelle influencent le résultat. On répartit la population à sonder en différents groupes plus homogènes que l'on échantillonne séparément (échantillonnage stratifié). Si l'effectif des groupes est déjà connu, on montre que l'on peut ainsi améliorer la précision par rapport à un échantillonnage aléatoire simple de même taille. On dispose en plus d'une information beaucoup plus riche sur la répartition des opinions. En outre, les statisticiens ont développé la théorie des plans d'expérience pour améliorer la collecte des données. Il s'agit, dans les exemples cités, de concevoir, par exemple, l'allocation des mesures ou le choix des unités expérimentales, de façon à ce que pour un coût donné la précision de l'analyse sur les quantités

d'intérêt soit la meilleure possible. Cette optimisation repose sur le modèle d'analyse et en exploite les propriétés. Une analyse statistique n'est jamais menée sans une connaissance minimale du problème, et la prise en compte, dans le modèle statistique, de toute information pertinente déjà disponible, permet aussi un gain de précision. Les statisticiens ont développé beaucoup d'outils pour permettre une inférence efficace, mais sans jamais lever le caractère intrinsèquement probabiliste de la démarche. Il reste toujours une incertitude : statistique n'est pas divination. Un exemple historique est celui des élections présidentielles américaines de 1948. Les sondeurs d'opinion, rendus trop confiants par leur précédent succès, annoncent la victoire de Dewey, mais Truman l'emporte et le jour de son investiture, les sénateurs – déçus – de l'Indiana observeront une minute de silence « à la mémoire du Dr Gallup », le fondateur de l'institut de sondage qui porte son nom !

François Rodolphe
Laboratoire de mathématique, informatique
et génome (INRA), Jouy-en-Josas

généralement de présenter des caractéristiques valorisées par la société. Elles auraient du pouvoir, ce qui d'après certaines théories sociologiques, serait plus bénéfique pour les hommes que pour les femmes. Il serait donc « naturel » que des parents attrayants aient plus de garçons. Nous ne prétendons pas que c'est vrai ; nous disons simplement que cela pourrait l'être, mais qu'on pourrait tout aussi bien imaginer une argumentation aboutissant au fait qu'ils ont plus de filles... Ce qui n'est pas sans poser quelques difficultés !

TROUVER DES DIFFÉRENCES LÀ OÙ IL N'Y EN A PAS

Comparez ces deux affirmations : « Les parents beaux ont plus de filles » et « Il n'est pas prouvé que des parents beaux aient plus ou moins de filles ». Nul doute que la première, sensationnelle, ferait davantage les gros titres ! Les éditeurs des revues sérieuses où l'affirmation de Kanazawa a été publiée ont-ils eux-mêmes été influencés ? Sans doute, et à cela deux raisons possibles. D'une part, les erreurs statistiques sont parfois difficilement détectables, même par des spécialistes. D'autre part, la signification statistique n'est pas directement liée à la taille des échantillons quand les effets testés sont petits. Avec un échantillon suffisamment grand, on peut presque toujours trouver de petits effets statistiquement significatifs. Mais quand les effets étudiés sont infimes, les études faites sur des cohortes trop petites conduisent à des interprétations abusives.

BIBLIOGRAPHIE

- A. GELMAN ET AL.
Letter to the editor regarding some papers of Dr. Satoshi Kanazawa,
Journal of Theoretical Biology,
vol. 245, pp. 597-599, 2007.
- S. KANAZAWA
Beautiful parents have more daughters: a further implication of the generalized Trivers-Willard hypothesis,
Journal of Theoretical Biology,
vol. 244, pp. 133-140, 2007.
- R. VON MISES
Probability, Statistics and Truth,
Dover, 2006.

L'étude du rapport des sexes à la naissance n'est pas neuve. Par exemple, dans son ouvrage *Probabilité, statistiques et vérité*, publié en 1957, Richard von Mises étudia ce rapport pour les naissances de 1907 et 1908 à Vienne, et trouva moins de variations qu'on en attendrait du simple hasard. Il l'attribua à des répartitions des sexes différentes selon les groupes ethniques. Pourtant, l'incertitude obtenue sur les mesures n'était ni plus ni moins que celle d'un hasard pur. Que faire face à cette volonté de trouver des différences là où il n'y en a pas ? Pour éviter ces biais, il faut montrer que les résultats observés représentent des effets réels indépendants de la sélection des échantillons, et trouver un argument biologique confirmant que des effets de l'ordre de 1 % sont importants.

Lorsque nous devons estimer des petits effets, les statisticiens doivent garder un regard critique sur les estimations obtenues. Mais les méthodes d'analyses ne sont pas exemptes de failles méthodologiques : les calculs fréquentistes ne tiennent pas compte des tailles des effets ; les analyses bayésiennes ne sont pas souvent meilleures en utilisant surtout des distributions *a priori* gaussiennes, et ignorent souvent les problèmes de puissance statistique. D'où l'importance d'estimer correctement les incertitudes, notamment en calculant les statistiques sur le signe des effets et sur leur amplitude. Nul doute que l'échange de méthodes et d'idées ouvrira la voie à une meilleure estimation des petits effets. ■