Mathématiques

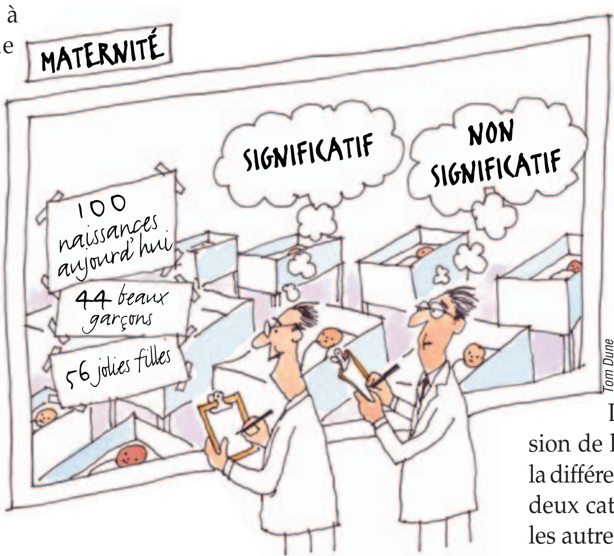
Le casse-tête des **petits effets**

Andrew Gelman et David Weakliem

On accorde parfois à de petites différences un sens qu'elles n'ont pas. Distinguer les petits effets qui ont une pertinence statistique et ceux qui relèvent du hasard est un défi pour les scientifiques.

Que penser de ces affirmations insolites parues récemment dans des revues scientifiques sérieuses ? Par exemple : « Les ingénieurs ont plus de garçons, les infirmières ont plus de filles », « Les hommes violents ont plus de garçons », ou encore « Les parents séduisants ont plus de filles ». On les doit à Satoshi Kanazawa, un chargé de cours en management et méthodologie de recherche à l'École d'économie de Londres.

Ses articles ont certes été publiés, mais les analyses statistiques sur lesquelles ils étaient fondés ont depuis été invalidées : l'une présentait des biais d'échantillonnage, l'autre des comparaisons multiples mal réalisées. Ainsi, les études rapportées par S. Kanazawa ne permettent pas d'affirmer qu'il existe une réelle différence entre la catégorie de parents jugés séduisants et les autres. Ces affirmations ne sont pas significatives et peuvent n'être que le fruit du hasard. Ces « découvertes » n'auraient pas été publiées si leur pertinence statistique avait été correctement évaluée. Mais leur mise en avant médiatique soulève une question : que penser des résultats qui ne sont pas statistiquement significatifs, mais qui attisent notre curiosité ? Après tout, n'importe qui muni de connaissances élémentaires en statistique peut fouiller dans des bases de données pour y trouver les « preuves » d'une hypothèse. Faudrait-il pour autant rejeter ces résultats ? Certainement pas. Imaginons que nous ayons trouvé que la probabilité d'avoir une fille soit supérieure



de cinq pour cent chez les parents séduisants, mais que l'écart-type – une mesure de l'incertitude, sur laquelle nous reviendrons – soit de quatre pour cent ; ce résultat serait-il pertinent ? Non, mais si nous devions donner notre avis, nous dirions sans doute que des parents séduisants ont plus de chances d'avoir une fille. D'ailleurs, l'hypothèse de S. Kanazawa n'est pas à rejeter, car elle s'appuie sur le modèle Trivers-Willard publié en 1973, et accepté par la communauté scientifique. Dans cet article, nous nous intéresserons à ces petites différences, que l'on nomme petits effets. Nous examinerons dans quelle mesure elles ont un sens statistique, nous donnerons des pistes pour interpréter des résultats non significatifs (que l'on cherche malgré tout à interpréter), et nous évoquerons les conséquences fâcheuses que peut avoir une interprétation erronée de ces petits effets.

Avant de poursuivre, voyons ce que signifient « statistiquement significatif », et « écart-type ». L'écart-type mesure la dispersion de la variable aléatoire étudiée (ici, la différence des proportions de filles entre deux catégories de parents : les beaux et les autres) autour de sa valeur moyenne. Plus l'écart-type est grand, plus la dispersion des valeurs trouvées est importante. Comment l'estimer ? L'écart-type est la racine carrée de la variance, laquelle se calcule à partir de la taille des groupes. Pour une proportion, elle est inversement proportionnelle à la taille de l'échantillon testé. On a donc tout intérêt à choi-

© ER Productions/Corbis

1. LES « BEAUX » COUPLES ont-ils plus de filles que de garçons ? Une étude l'a suggéré. Cependant, une analyse statistique rigoureuse des petits écarts, entre le nombre de filles et de garçons à la naissance, a montré que les résultats obtenus étaient non significatifs.

sir de grands échantillons pour réduire l'incertitude sur les mesures. Par exemple, pour un échantillon de 100 couples, l'écart-type de la proportion de filles est égal à cinq pour cent. Pour un échantillon de 3000 couples, l'écart-type n'est plus que de neuf pour mille.

Dans le cas qui nous intéresse, nous cherchons à vérifier l'hypothèse suivante: la proportion de filles est plus élevée dans le groupe de parents séduisants, que dans le groupe de parents jugés moins beaux. Quand peut-on dire que le résultat trouvé est statistiquement significatif ? Pour l'affirmer, il faut qu'il soit suffisamment éloigné d'un résultat plausible si les parents jugés beaux ont autant de chances d'avoir une fille que les autres parents (environ une chance sur deux). La différence des proportions entre deux groupes n'est pas statistiquement significative quand elle est probable, du seul fait des fluctuations d'échantillonnage, même s'il n'y a pas de réelle différence entre les deux groupes.

Considérons un exemple plus simple: une séquence de 20 lancers de pièce. Imaginons que nous obtenions 8 faces pour 12 piles. La proportion observée de faces serait de 40 pour cent, pour un écart-type égal à 11 pour cent. L'estimation obtenue n'étant pas assez éloignée de 50 pour cent – c'est-à-dire moitié faces moitié piles –, le résultat obtenu peut être attribué au hasard et n'est donc pas significatif. Dans les études dont il est question ici, on approximera la loi de la variable aléatoire par une loi gaussienne, et on se contentera de dire qu'une différence est significative si elle est supérieure à son écart-type multiplié par 1,96 *(voir la figure 2)*.

L'ESSENTIEL

✔ **Les facteurs qui influent sur la répartition des sexes à la naissance ont souvent été étudiés. Les effets attendus sont de l'ordre de un pour cent.**

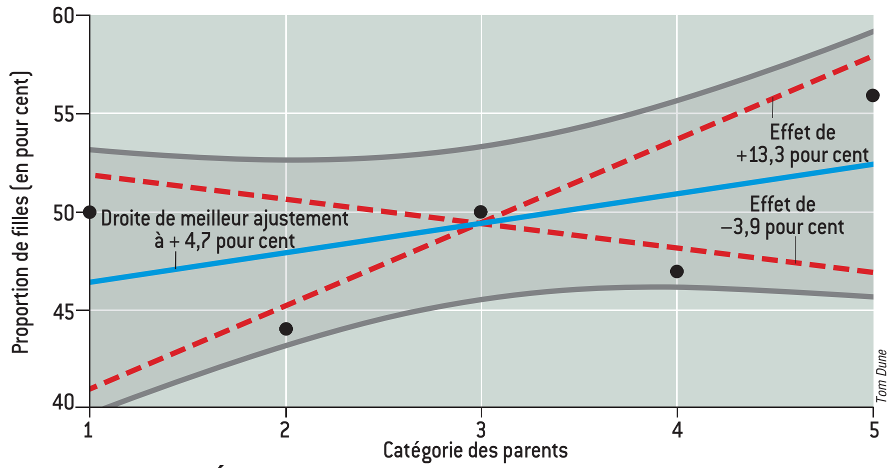
✔ **L'évaluation de la pertinence statistique d'un résultat repose sur l'estimation des incertitudes et sur le choix de l'analyse effectuée.**

✔ **Les analyses fréquentistes et bayésiennes sont deux méthodes possibles d'analyse.**

✔ **Une estimation des petits effets n'est pertinente que si la taille des échantillons testés est grande.**

Fille ou garçon ?

Voyons de plus près l'analyse de S. Kanazawa – les parents beaux ont plus de filles – fondée sur l'Étude américaine sur la santé des adolescents. La beauté des sujets fut évaluée sur une échelle de 1 à 5 et le sexe de leurs enfants répertorié. Pour les quelque 3000 parents étudiés, S. Kanazawa rapporta une différence des proportions de huit pour cent, paraissant significative: la proportion de filles était égale à 52 pour cent pour les parents les plus attirants, contre 44 pour cent pour la moyenne des quatre autres catégories (notées de 2 à 5). En fait, comme l'a montré l'un d'entre nous (Andrew Gelman), comparer la première catégorie aux quatre autres n'est qu'une des nombreuses approches possibles: on aurait pu, par exemple, comparer les deux groupes les plus beaux aux deux groupes les moins beaux. En tenant compte de cela, il s'avère que la significativité est perdue. Voilà un bel exemple de résultat sociologique suggestif, mais qui n'est pas statistiquement



2. UNE ANALYSE FRÉQUENTISTE des données détermine la courbe de meilleur ajustement *(la droite en bleu)* aux données de S. Kanazawa *(points en noir)*, pour un écart-type de 4,3 pour cent. Un intervalle de confiance à 95 pour cent montre que des effets aussi faibles que –3,9 pour cent et aussi grands que 13,3 pour cent sont compatibles avec les données, car ils encadrent l'estimation à 4,7 pour cent.

significatif : il pourrait tout à fait être le fruit du hasard. Pourtant, il semble étayer le modèle proposé. Quand nous faisons face à ce type de problème statistique, nous devons le traiter en tenant compte de l'amplitude des effets attendus. Comme nous le verrons, il est possible que des parents jugés beaux aient un pour cent de chances supplémentaires d'avoir une fille, mais il est peu probable que la différence atteigne cinq pour cent.

On s'attend à ce que les effets étudiés ici soient faibles, ce qu'attestent les multiples études sur les variations du rapport filles-garçons à la naissance. Ce rapport varie de un pour cent (la probabilité d'avoir une fille passant par exemple de 48,5 à 49,5 pour cent), selon divers facteurs : le groupe ethnique, l'âge des parents, le rang de naissance, le poids de la mère, le statut du couple et la saison de la naissance (l'effet de chacun de ces facteurs est de l'ordre de 0,3 pour cent). Les conditions socio-économiques, notamment la pauvreté et la sous-alimentation, ont une influence plus marquée, atteignant trois pour cent. Il n'est pas étonnant que des privations extrêmes augmentent la proportion de naissances de filles, car les fœtus mâles sont plus fragiles.

Compte tenu de ces données scientifiques, on s'attendrait à ce que l'effet de la beauté des parents sur le rapport filles-garçons à la naissance soit inférieur à un pour cent, comme les variations observées couramment. Vérifions si c'est bien le cas en nous fondant sur deux approches statistiques : l'analyse dite fréquentiste et l'analyse bayésienne. Dans la première approche, on se fixe des hypothèses, puis

Glossaire

✔ **L'écart-type, σ, mesure la dispersion de la variable aléatoire. Il est égal à la racine carrée de la variance.**

✔ **La variance, *v*, est la valeur moyenne du carré de l'écart à la moyenne. Pour une proportion dans un échantillon de taille *n*, *v* est égale à $p(1-p)/n$, où *p* est la probabilité de la catégorie étudiée.**

✔ **Variable aléatoire : résultat d'une expérience aléatoire à laquelle on affecte une probabilité (par exemple : résultat d'un jet de dé).**

✔ **Intervalle de confiance à *x* pour cent : il a une probabilité de x pour cent de contenir la valeur réelle du paramètre que l'on souhaite estimer.**

✔ **Distribution gaussienne *(voir la figure 3)* : la distribution d'une proportion dans un échantillon converge vers une gaussienne quand la taille *n* de l'échantillon augmente.**

on traite statistiquement les données pour savoir si elles sont plus compatibles avec l'une des hypothèses. Éventuellement, on décidera qu'une hypothèse est vraie. Dans la seconde, on tient compte d'une information *a priori*, (dans le cas qui nous intéresse : les effets de la beauté des parents sur le rapport des sexes à la naissance ne peuvent être grands) sous la forme d'une distribution des valeurs plausibles de l'effet. L'information que l'on tire de l'expérience est la loi conditionnée par l'observation, donc modifiée par elle, qui est appelée loi *a posteriori* : c'est une nouvelle distribution des effets plausibles.

L'analyse fréquentiste

En reprenant l'étude de S. Kanazawa, nous avons d'abord suivi une méthode d'analyse fréquentiste pour estimer la probabilité de naissance de filles en fonction de la beauté des parents. Au terme de cette étude, nous avons estimé une différence de probabilité de 4,7 pour cent entre les deux groupes, pour un écart-type de 4,3 pour cent, ce qui reste cohérent avec le résultat de S. Kanazawa *(voir la figure 2)*. Avec ces valeurs, on peut calculer l'intervalle de confiance qui contient la vraie valeur avec une probabilité de 95 pour cent, et qui est (en pourcentage) : [-3,9, 13,3]. Comment interpréter statistiquement l'intervalle de confiance ? Il contient la valeur zéro qui correspond à l'absence d'effet de la beauté des parents sur le rapport des sexes à la naissance. Notre estimation à 4,7 pour cent n'est donc pas significative et il faut poursuivre les analyses statistiques avant de conclure (si toutefois c'est possible !). Et en dehors des bornes de l'intervalle de confiance à 95 pour cent, que se passe-t-il ? Que représentent les cinq pour cent de probabilité restants ?

Pour que l'effet trouvé soit statistiquement significatif, il faudrait qu'il soit plus grand que 1,96 fois l'écart-type, c'est-à-dire au-dessus de 8,4 pour cent (l'intervalle de confiance ne contiendrait alors pas 0). Exprimé autrement : la probabilité de rejeter à tort la valeur nulle est de cinq pour cent.

Mais est-ce bien raisonnable d'envisager des effets significatifs aussi grands que 8,4 pour cent ? Non, bien sûr. C'est ce que l'on nomme une erreur de magnitude. L'étude est construite de façon telle que tout résultat statistiquement significatif surestime le véritable effet (qui ne peut dépasser un pour cent). S'y ajoutent des erreurs de signe quand l'estimation trouvée est de signe opposé au véritable effet (ici : « les parents beaux ont plus de garçons que de filles »). Ainsi, on distingue deux types d'effet : positif, si les parents « beaux » ont plus de filles que les autres, et négatif, s'ils en ont moins.

Pour illustrer les probabilités associées à ces erreurs, envisageons quatre scénarios fondés sur des écarts-type égaux à 4,3 pour cent *(voir l'encadré ci-dessous)*. Ils montrent qu'une étude à partir de cette taille d'échantillon (3 000 couples environ) n'est pas pertinente pour estimer des petits effets de l'ordre du pour cent. Cela est dû notamment à la valeur de l'écart-type qui est particulièrement élevée ici (4,3 pour cent). C'est pourquoi les études de la répartition des sexes des nouveau-nés utilisent des échantillons beaucoup plus grands, fondés sur de vastes bases de données démographiques, comptant plus d'un million d'individus.

L'analyse bayésienne

On résume tout ce que l'on sait sur l'effet à analyser, en utilisant des sources extérieures déjà connues, par une distribution *a priori*. Cette distribution sera modifiée (conditionnée) par le résultat de l'expérience pour donner la distribution *a posteriori*. Si l'on ne savait rien d'avance, on pourrait prendre une distribution *a priori* non informative et le résultat correspondrait à celui de l'approche fréquentiste. Ici, la distribution obtenue *a posteriori* serait alors approximativement une gaussienne, de moyenne 4,7 pour cent et d'écart-type 4,3 pour cent, ce qui correspondrait à une probabilité de 86 pour cent environ que l'effet réel soit positif. En général, plus la distribution *a priori* est concentrée autour de zéro (hypothèse selon laquelle l'effet réel sur la différence des sexes est faible), plus la probabilité *a posteriori* est proche de 50 pour cent.

Choisissons, par exemple, une distribution *a priori* gaussienne (distribution en forme de cloche), centrée sur zéro avec une forme telle que la différence réelle de probabilité d'avoir une fille (selon que les parents sont beaux ou non) soit proche de zéro, avec des probabilités 50 pour cent, 90 pour cent et 94 pour cent d'être respectivement dans les intervalles (en pourcentage) [–0,3, 0,3], [–1,1], et [–3,3].

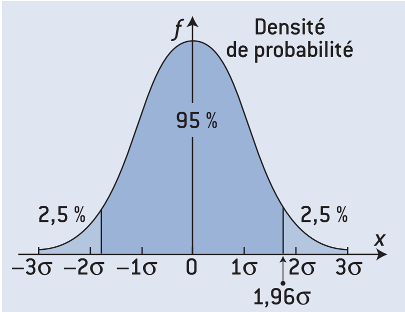
Pourquoi centrer la distribution sur zéro ? Parce que nous n'avons pas d'*a priori* sur le signe de la différence réelle de probabilité de naissance de fille en fonction de la beauté des parents.

À l'étape suivante, nous calculons, à partir de cette distribution *a priori* et des données, la distribution *a posteriori* de l'effet. Pour résumer, la distribution *a posteriori* donne une probabilité que l'effet soit positif (les parents beaux ont plus de filles) de seulement 58 pour cent, dont 45 pour cent que cette différence positive soit inférieure à un pour cent. Cette analyse dépend – mais peu – de la distribution *a priori* ; par exemple, si l'on décide d'élargir la courbe de distribution autour de zéro, la probabilité que l'effet réel soit positif n'augmente que de sept pour cent (pour atteindre 65 pour cent). Le fait de changer de famille de courbes de distribution a peu d'effets sur les résultats : les effets réels restent faibles, ce que confirment les données.

L'idéal scientifique dans l'inférence sur une quantité (ici, le lien entre la beauté des parents et le rapport des sexes à la naissance) est la description de l'incertitude

LES AUTEURS


Andrew GELMAN, professeur de statistiques et de sciences politiques, dirige le Centre de statistiques appliquées de l'Université Columbia, à New York. David WEAKLIEM est professeur de sociologie à l'Université du Connecticut.



3. L'INTERVALLE DE CONFIANCE à 95 pour cent est compris entre les bornes $-1{,}96\,\sigma$ et $1{,}96\,\sigma$. La probabilité associée est l'aire sous la courbe.

Zoom sur quatre scénarios

Supposons dans le scénario que nous nommerons « Différence exactement nulle » qu'il n'y a aucune corrélation entre la beauté des parents et le sexe de leur enfant (il n'y a pas de différence entre les deux groupes de parents). Il existe toujours une petite probabilité – cinq pour cent – que le résultat trouvé soit statistiquement significatif. Mais ce résultat sera quand même faux, on a cinq pour cent de chances de se tromper sur la valeur et ce quel que soit le signe de l'effet, positif ou négatif.

Dans le scénario dit de 0,3 pour cent exactement, les parents beaux ont 0,3 pour cent de probabilité supplémentaire d'avoir une fille – valeur plausible. On a toujours une probabilité de cinq pour cent de trouver un résultat statistiquement significatif à l'issue de l'étude. Comment se répartissent ces cinq pour cent de part et d'autre de la valeur réelle ? Dans ces conditions, on a une probabilité de trois pour cent d'observer un effet statistiquement significatif positif et deux pour cent d'observer un effet statistiquement significatif négatif. Mais dans chaque cas, l'effet estimé – au moins 8,4 pour cent pour qu'il soit significatif – sera beaucoup trop éloigné de la valeur réelle de 0,3 pour cent. Ainsi, l'erreur de magnitude sera importante. Par ailleurs, la probabilité d'aller dans le mauvais sens – erreur de signe – sera égale à 2/5. Imaginons que l'on se retrouve entre les bornes de l'intervalle de confiance à 95 pour cent, donc proches de la valeur réelle. Dans ce cas, l'erreur de magnitude serait faible, mais la probabilité de faire une erreur de signe n'est pas négligeable : 47,5 pour cent, proche des 50 pour cent observés en l'absence d'effet. Ainsi, le signe observé n'apporte pratiquement aucune information sur le sens d'une éventuelle différence.

Troisième scénario : si les parents séduisants ont une probabilité d'avoir une fille supérieure de un pour cent (ce qui semble être l'effet maximal possible), alors on a réciproquement des probabilités de quatre et un pour cent d'avoir des effets positif et négatif statistiquement significatifs. Globalement, il y a une probabilité de 40 pour cent de faire une erreur sur le signe de l'estimation ; là encore, l'estimation donne peu d'informations sur le signe ou l'ordre de grandeur de l'effet réel.

Enfin, envisageons le quatrième scénario : même si la différence réelle était de trois pour cent – valeur improbable –, il n'y aurait encore que dix pour cent de probabilité d'obtenir un résultat statistiquement significatif. Dans ce cas, nous aurions une probabilité de 24 pour cent de faire une erreur de signe. Ce type d'étude est donc peu informatif.

Tom Dune

✔ BIBLIOGRAPHIE

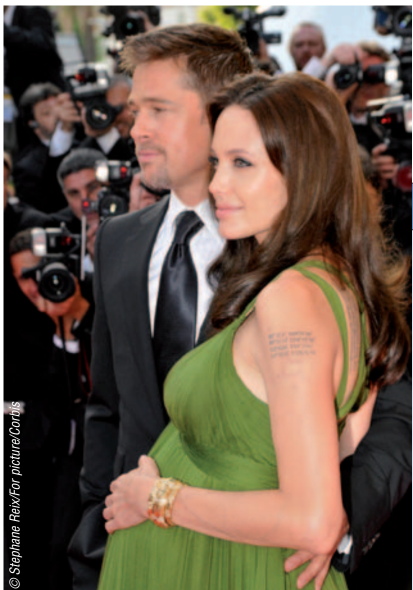
Andrew Gelman *et al.*, **Letter to the editor regarding some papers of Dr. Satoshi Kanazawa**, *Journal of Therorectical Biology*, vol. 245, pp. 597-599, 2007.

Satoshi 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.

Richard von Mises, **Probability, Statistics and Truth**, Dover, 2006.

Robert Trivers et Dan Willard, **Natural selection of parental ability to vary the sex ratio of offspring**, *Science*, vol. 179, pp. 90-92, 1973.

Nous remercions la revue *American Scientist* de nous avoir autorisés à publier cet article.



© Stephane Reix/For picture/Corbis

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 S. Kanazawa) et publier leurs résultats. Des méta-analyses 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 pour cent à 58 pour cent, puis peut-être redescendre à 38 pour cent et ainsi de suite.

Beauté et sexe des enfants : un lien ?

Comment reconnaître les données fiables ? En collectant toujours plus de données. Allons plus loin en collectant plus de données. Chaque année, le magazine américain *People* publie une liste des 50 célébrités mondiales les plus belles. Nous avons répertorié de 2005 jusqu'à 2007 le sexe de leurs enfants pour les numéros parus entre 1995 et 2000. Pour les personnes citées en 1995, par exemple, on a dénombré 32 naissances de filles pour 24 garçons, soit 57,1 pour cent de filles, ce qui correspond à 8,6 pour cent de plus que dans la population générale (48,5 pour cent). Un résultat en accord avec l'hypothèse de S. Kanazawa. Mais l'écart-type étant de 6,7 pour cent, l'estimation de 8,6 pour cent n'est pas statistiquement significative. Pour le confirmer, nous avons comparé ce résultat avec ceux des années antérieures *(voir le tableau ci-dessous)*. Nous en avons déduit que les plus belles personnes citées dans *People* entre 1995 et 2000 ont eu 157 filles sur un total de 329 enfants, soit 47,7 pour cent de filles (pour un écart-type de 2,8 pour cent), ce qui est seulement 0,8 pour cent inférieur au chiffre obtenu pour la population générale. On ne peut rien en conclure…

Alors pourquoi perdre notre temps à étudier des erreurs statistiques que personne n'a repérées ? À cela deux raisons. D'abord, les résultats qui semblent avoir un sens sans être statistiquement significatifs sont les plus problématiques. Deuxièmement, 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.

Ainsi, les résultats de S. Kanazawa ont tout de suite suscité l'intérêt des médias. On pouvait lire sur un blog du quotidien américain *The New York Times* : « Une nouvelle étude de Satoshi Kanazawa, psychologue évolutionniste de

Les 50 *people* les plus beaux

Le magazine *People* publie chaque année la liste des 50 célébrités mondiales les plus belles. On a répertorié entre 1995 et 2000 le sexe de leurs enfants, ainsi que l'écart-type des données obtenues. Les célébrités citées plusieurs fois ne sont prises en compte qu'une fois.

Année de publication	Nombre de filles	Nombre de garçons	Proportion de filles	Écart-type
1995	32	24	57,1 %	6,7 %
1996	45	35	56,2 %	5,6 %
1997	24	35	40,7 %	6,5 %
1998	21	25	45,7 %	7,4 %
1999	23	30	43,4 %	6,9 %
2000	29	25	53,7 %	6,8 %
1995-2000	157	172	47,7 %	2,8 %

l'École d'économie de Londres, suggère […] qu'il y a plus de belles femmes que d'hommes beaux dans le monde. Pourquoi ? Parce que les parents beaux ont 36 pour cent plus de chances d'avoir une fille comme premier enfant, ce qui suggère, en termes évolutionnistes, que la beauté est un caractère plus important pour les femmes que pour les hommes. L'étude a été réalisée auprès de 3000 Américains, provenant d'une Étude longitudinale américaine sur la santé des adolescents. »

Cette publication dans une revue à comité de lecture semblait être une caution suffisante balayant les doutes éventuels, ce qui est étonnant quand on sait que les blogueurs du *New York Times* sont plutôt qualifiés pour juger les recherches en sciences sociales.

Le téléphone arabe

Qui plus est, l'effet n'a cessé d'augmenter. Ainsi, l'estimation – statistiquement non significative – de 4,7 pour cent que nous avons faite est passée à huit pour cent dans l'analyse de S. Kanazawa (comparaison du groupe le plus beau à la moyenne des quatre groupes les moins séduisants), pour atteindre la valeur de 26 pour cent après une étude complémentaire introduisant d'autres corrections, avant de grimper à 36 pour cent pour des raisons encore floues ! Cette inflation nous surprit, ce chiffre étant 10 à 100 fois supérieur à tous les 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 (on l'a vu, l'écart-type, donc l'incertitude, diminue avec la taille de l'échantillon testé). Ainsi, si l'on veut affirmer quelque chose à propos d'effets de l'ordre de un pour cent, on a tout intérêt à partir de données pertinentes et à réaliser des tests qui exploitent bien les données. Cet exemple

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 échantillonnera 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

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 – c'est-à-dire l'incertitude – 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 S. 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 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 !

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 S. 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.

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 un pour cent 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. ■