

L'ESSENTIEL

● Certains résultats fracassants en matière de statistiques sont parfois fondés sur de petits effets difficiles à interpréter.

● La pertinence statistique de tels résultats requiert une estimation précise des incertitudes et une méthode d'analyse adaptée.

● Les analyses fréquentistes et bayésiennes sont parmi les méthodes les plus fiables.

● Autre précaution nécessaire : une estimation des petits effets n'a de sens que si la taille des échantillons est grande.

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.

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

LE CASSE-TÊTE des petits effets

On accorde parfois à de petites différences un sens qu'elles n'ont pas. Distinguer les petits effets statistiquement pertinents de ceux relevant du hasard est un défi !

« **L**

es ingénieurs ont plus de garçons, les infirmières plus de filles », « Les hommes violents ont plus de garçons », « Les individus séduisants ont plus de filles »... On doit ces phrases définitives à Satoshi Kanazawa, psychologue évolutionniste à l'École d'économie de Londres. L'auteur est certes controversé, mais il a néanmoins publié ces « conclusions » pour le moins insolites dans des revues scientifiques sérieuses. L'un de ses derniers articles, parus en 2017, est intitulé « Les individus sont d'autant moins

séduisants que leur père était âgé au moment de la conception ». Qu'en pensez-vous ?

Les spécialistes ont rendu leur avis : les analyses statistiques sur lesquelles ces articles étaient fondés ont été invalidées, par exemple à cause de biais d'échantillonnage. Ainsi, ces « découvertes » ne sont pas significatives et peuvent n'être que le fruit du hasard : elles n'auraient jamais été publiées si leur pertinence statistique avait été correctement évaluée.

Faut-il pour autant rejeter en bloc ces travaux ? Non, et d'ailleurs, certaines hypothèses de Satoshi Kanazawa s'appuient sur des théories reconnues par la communauté scientifique. Cependant, le bulldozer de la médiatisation a écrasé toute nuance et précaution pourtant indispensables à l'interprétation des petites différences statistiques relevées par le psychologue. Elles relèvent de ce que l'on nomme les petits effets. Dans quelle mesure ces derniers ont-ils un sens statistique ? Comment interpréter des résultats non significatifs (et que l'on cherche malgré tout à interpréter) ? Quelles peuvent être les conséquences d'une interprétation erronée de ces petits effets ?

Avant de répondre, voyons ce que signifient «statistiquement significatif», et «écart-type», deux notions centrales dans l'étude des petits effets. Nous utiliserons l'exemple de l'étude montrant que «Les individus séduisants ont plus de filles». L'écart-type mesure la dispersion de la variable aléatoire étudiée (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 choisir 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 5%. Pour un échantillon de 3000 couples, l'écart-type n'est plus que de 0,9%.

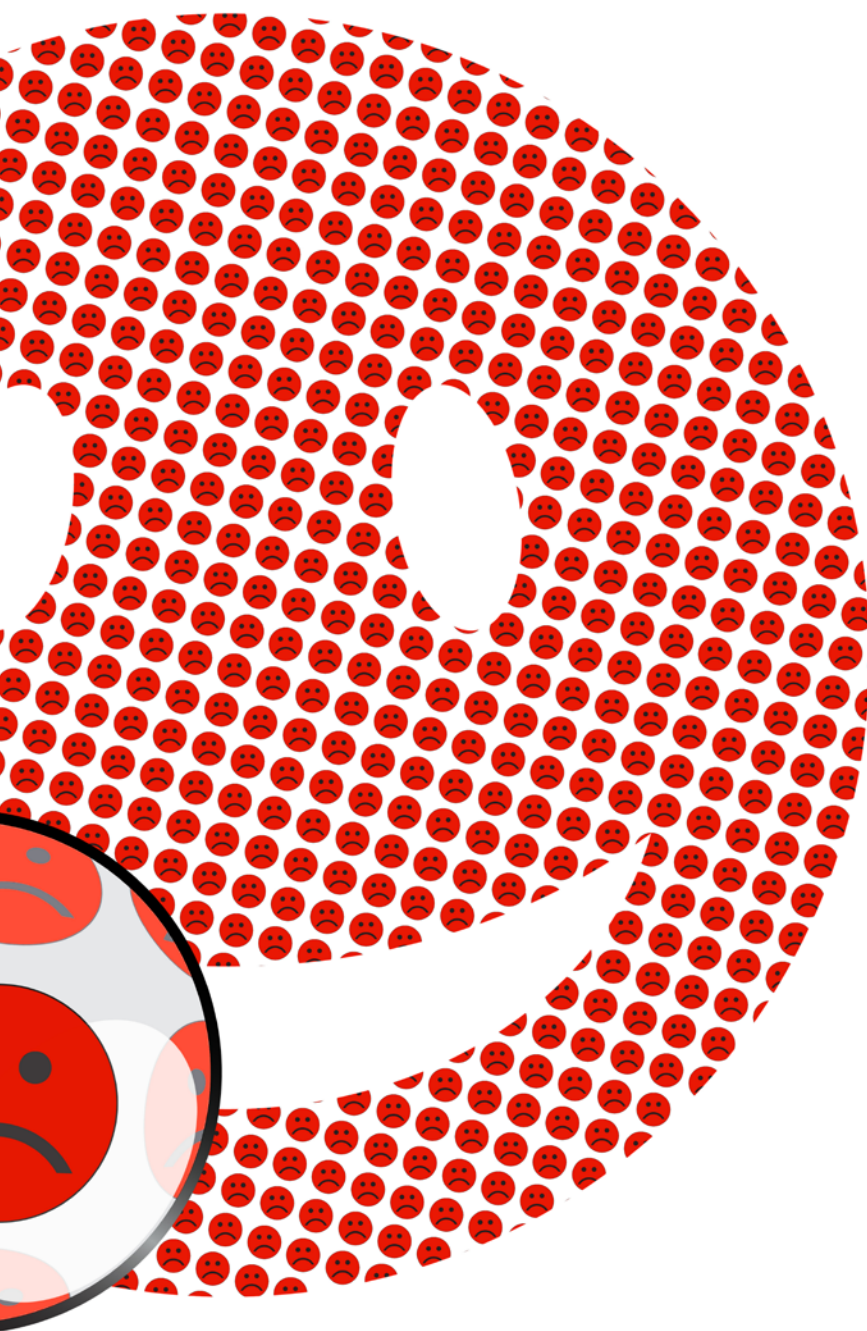
Dans le cas qui nous intéresse, nous cherchons à vérifier l'hypothèse suivante: la proportion de filles est plus élevée dans le groupe de parents séduisants, que dans le groupe de parents jugés moins beaux. Quand peut-on dire que le résultat trouvé est statistiquement significatif? Pour l'affirmer, il 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.

Prenons l'exemple plus simple d'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 un écart-type de 11%. L'estimation obtenue n'étant pas assez éloignée de 50% (moitié faces et 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 approximerait la loi de la variable aléatoire par une loi gaussienne, et on se contenterait de dire, sans détailler le calcul, qu'une différence est significative si elle est supérieure à son écart-type multiplié par 1,96.

FILLE OU GARÇON?

Voyons de plus près l'analyse de Kanazawa. La beauté des sujets fut évaluée sur une échelle de 1 à 5 et le sexe de leurs enfants répertorié. Pour les quelque 3 000 parents étudiés, Kanazawa rapporta une différence des proportions de 8%, paraissant significative: la proportion de filles était égale à 52% pour les parents les plus attirants (notés 1), contre 44% pour la moyenne des quatre autres catégories (de 2 à 5). En fait, 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: cette fois, la significativité disparaît.

Voilà un bel exemple de résultat sociologique suggestif, mais qui n'est pas statistiquement significatif: il pourrait tout à fait être le fruit du hasard. Pourtant, il semble étayer le modèle proposé. Face à ce type de problème statistique, nous devons tenir compte de l'amplitude des effets attendus.





BEAUTÉ ET DESCENDANCE

Le magazine *People* publie chaque année la liste des cinquante 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. Difficile d'en déduire que plus on est beau, plus on a de filles.

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 %

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 1% (la probabilité d'avoir une fille passant par exemple de 48,5 à 49,5%), 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. Les conditions socioéconomiques, notamment la pauvreté et la sous-alimentation, ont une influence plus marquée, atteignant 3%, 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 1%, comparable aux 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.

L'ANALYSE FRÉQUENTISTE

Dans la première approche, on se fixe des hypothèses, puis on traite statistiquement les données pour savoir si elles sont plus compatibles avec l'une des hypothèses: celle-ci sera alors éventuellement déclarée vraie. Dans la seconde, on tient compte d'une information *a priori*, (ici, 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.

En reprenant l'étude de 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. Nous avons estimé une différence de probabilité de 4,7% entre les deux groupes, pour un écart-type de 4,3%, ce qui reste cohérent avec le résultat de Kanazawa (voir la figure page 32). Avec ces valeurs, on peut calculer l'intervalle de confiance qui contient la vraie valeur avec une probabilité de 95%, et qui est, en pourcentages: [-3,9; 13,3]. Comment interpréter statistiquement l'intervalle de confiance?

**Comment reconnaître les données fiables ?
En collectant toujours plus de données !**

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% n'est donc pas significative et il faut poursuivre les analyses statistiques avant de conclure (si c'est possible!). Et en dehors des bornes de l'intervalle de confiance à 95%, que se passe-t-il? Que représentent les 5% de probabilité restants?