Chaque tentative de réforme devra combattre des habitudes fermement établies : la façon dont les statistiques sont enseignées dans les universités, celle dont les résultats des études sont exploités et interprétés et comment ils sont ensuite relayés dans les revues spécialisées. Mais au moins, de nombreux scientifiques ont admis l'existence d'un problème. Grâce à des chercheurs comme John Ioannidis, les préoccupations des statisticiens ne sont plus vues uniquement comme de la pure théorie.

**Il est pratiquement impossible d'arracher les auteurs à leur sacro-sainte valeur-*p***

Pour améliorer la situation, les statisticiens ont quelques outils à leur disposition. Par exemple, les chercheurs pourraient toujours publier également les tailles d'effet et les intervalles de confiance (*voir les Repères, page 6*) obtenus. Ces valeurs expriment ce que ne peut exprimer la valeur-*p* seule : la portée et l'importance relative de l'effet.

Beaucoup plaident aussi pour remplacer la valeur-*p* par des méthodes fondées sur la vision de Thomas Bayes, mathématicien britannique du XVIII^e^ siècle. Cette approche décrit une façon de penser les probabilités comme la plausibilité d'un résultat, plutôt que la fréquence potentielle de ce résultat. On introduit certes par là une subjectivité dans les statistiques, que les pionniers du début du XX^e^ siècle voulaient éviter à tout prix. Pourtant, la statistique bayésienne simplifie l'intégration des connaissances contextuelles sur le monde et permet de calculer comment les probabilités sont modifiées par des preuves nouvellement acquises.

D'autres soutiennent une approche plutôt pluraliste : les scientifiques devraient utiliser plusieurs méthodes sur un même ensemble de données. Lorsque les résultats différeront, les chercheurs devront être plus créatifs pour découvrir à quoi cela est dû. La compréhension de la réalité sous-jacente y gagnerait.

Aux yeux de Simonsohn, la meilleure protection pour un chercheur est de tout montrer. Les auteurs devraient garantir leur travail « sans *p*-hacking » en explicitant le choix de la taille de leurs échantillons, les données éventuellement mises de côté ainsi que les manipulations mises en œuvre. Aujourd'hui, aucune de ces informations n'est disponible ni vérifiable.

## DES ÉTUDES EN DEUX ACTES

À l'instar de la mention « Les auteurs n'ont pas d'intérêts financiers dépendants du contenu de cet article » encore courante aujourd'hui, ces informations feraient la différence entre faute scientifique involontaire ou délibérée. Lorsqu'une telle déclaration sera monnaie courante, le *p*-hacking sera éliminé. Ou au moins son absence sera-t-elle remarquée par le lecteur et lui permettra un jugement plus éclairé.

L'analyse en deux étapes ou « *preregistered replication* » (réplication préenregistrée) est une idée qui va dans ce sens, et elle gagne du terrain. Dans cette approche, les études exploratoires et confirmatoires sont abordées différemment et clairement identifiées. Au lieu, par exemple, de conduire quatre petites expériences et de réunir les résultats dans un article, les chercheurs balaieraient d'abord le domaine pour récolter des observations intéressantes avec deux petites études exploratoires, sans trop se préoccuper des fausses alertes. C'est seulement après, sur la base de ces données, qu'ils concevraient une étude qui confirmera, peut-être, leurs résultats, et préenregistreraient publiquement leurs intentions dans une banque de données publique, telle l'Open Science Framework. Ils publieraient ensuite leurs résultats, accompagnés de ceux de l'étude exploratoire dans un article d'une revue habituelle. Une telle approche laisse beaucoup de liberté, explique le politologue et statisticien Andrew Gelman, de l'université Columbia à New York. Mais elle est aussi suffisamment rigoureuse pour diminuer le nombre de fausses découvertes.

Plus généralement, l'heure est venue pour les scientifiques de prendre conscience des limites des statistiques conventionnelles. Avant tout, une estimation scientifique sérieuse de la plausibilité des résultats devrait être effectuée dès leur analyse : quels résultats ont été apportés par des recherches similaires ? Existe-t-il un mécanisme qui pourrait les expliquer ? Les résultats se recoupent-ils avec l'expérience clinique ? Voilà les questions décisives ! En y répondant dans de prochains travaux, Motyl peut encore espérer accéder au succès qu'il espère ! ■

## BIBLIOGRAPHIE

**D. BENJAMIN *ET AL.*,**
***Redefine statistical significance*, 2017**
**https://psyarxiv.com/mky9j**

**B. NOSEK *ET AL.*,**
**Restructuring incentives and practices to promote truth over publishability,**
***Perspect. Psychol. Sci.*,**
**vol. 7, pp. 615-631, 2012.**

**C. LAMBDIN**
**Significance tests as sorcery : Science is empirical – significance tests are not,**
***Theory & Psychology*,**
**vol. 2, pp. 67-90, 2012.**

Le *big data*, un flot de données venant de toutes parts.

© Shutterstock.com/Dmitriy Rybin

# LES DÉFIS DES BIG DATA

Comment faire face à ce déluge de données qui constitue les *big data* ? Comment les stocker, les exploiter, les analyser, les valoriser... ? Les algorithmes et les calculateurs à qui ces tâches incombent doivent être à la hauteur. La recherche sur ces deux aspects essentiels du traitement des *big data* est particulièrement active. Le développement des intelligences artificielles, « éduquées » par des quantités énormes de données, est une illustration de ces efforts. Ces systèmes sont parfois si performants que certains y voient une menace sur la façon de faire de la science.