

de ce qu'il veut prouver, par exemple que les individus qui ont reçu un médicament ne sont pas en meilleure santé que ceux ayant reçu un placebo. Le chercheur suppose ensuite que l'hypothèse nulle est vraie et calcule sous cette condition préalable la probabilité d'obtenir des résultats au moins aussi marqués que ceux qu'il a réellement observés. Cette probabilité est la valeur- p . Plus elle est petite, d'après Fisher, plus grande est la probabilité que l'hypothèse nulle soit fausse, et donc qu'elle doive être rejetée, ce que le chercheur souhaitait au début, car l'effet étudié est ainsi confirmé.

On peut calculer la valeur- p avec autant de chiffres après la virgule que l'on souhaite, mais cette précision apparente est trompeuse, car les hypothèses étudiées sont loin d'être aussi fines. Pour Fisher, la valeur- p ne représentait qu'un maillon dans une chaîne qui relie les observations et les connaissances de base pour parvenir à une conclusion scientifique.

Ces précautions ont été balayées par un mouvement dont le but était de rendre la prise de décision fondée sur les preuves aussi rigoureuse et objective que possible. Les pionniers de ce revirement étaient, à la fin des années 1920, le mathématicien polonais Jerzy Neyman et le statisticien britannique Egon Pearson, les rivaux les plus acharnés de Fisher. Leur système inclut des concepts comme «faux positif», «faux négatif» (voir les Repères,

page 6) et «puissance» d'un test statistique, que l'on trouve aujourd'hui dans tous les cours de statistiques pour débutants. Neyman et Pearson mirent cependant volontairement de côté la valeur- p .

Pendant que les représentants des deux camps s'écharpaient, d'autres chercheurs perdirent patience et rédigèrent eux-mêmes des manuels de statistiques. Malheureusement, beaucoup de ces auteurs n'étaient pas suffisamment versés en la matière pour apprécier les subtilités philosophiques des deux visions. Ils intégrèrent donc la valeur- p de Fisher, facile à calculer, dans le système régulé, rigoureux et sécurisant de Neyman et Pearson. Depuis, du haut de son piédestal, la valeur- p trône toujours. La valeur de 0,05 devint le seuil séparant le significatif de ce qui ne l'est pas. Mais ce n'était pas son rôle.

GUEULE DE BOIS ET TUMEUR

En conséquence, il règne une grande confusion aujourd'hui sur ce que la valeur- p exprime réellement. L'étude de Motyl en est une bonne illustration. La plupart des scientifiques interpréteraient sa première valeur- p de 0,01 comme une probabilité de fausse alerte de 1 %. C'est pourtant faux. La valeur- p ne livre qu'un résumé sommaire des données en considérant comme vraie une hypothèse nulle spécifique.

PLAUSIBLE OU NON ?

Avant l'expérience, la plausibilité de l'hypothèse testée (la probabilité qu'elle soit vraie) peut être estimée à partir d'études antérieures, de mécanismes supposés... Ici, on envisage trois cas.

Une valeur- p égale à 0,05 traduit conventionnellement un résultat « statistiquement significatif ». À 0,01, c'est « très significatif ».

Après l'expérience, une faible valeur- p peut rendre l'hypothèse plus plausible, mais pas nécessairement de façon... significative.

La valeur- p mesure la probabilité qu'un résultat observé soit dû au hasard. Mais elle néglige la question essentielle : quelle est la probabilité que l'hypothèse testée soit correcte ? La réponse dépend de la robustesse du résultat observé et, plus encore, de la plausibilité de l'hypothèse testée.

