

**Zeitschrift:** Horizons : le magazine suisse de la recherche scientifique  
**Herausgeber:** Fonds National Suisse de la Recherche Scientifique  
**Band:** 31 (2019)  
**Heft:** 120: Surprise! Place aux émotions : comment la science tente de saisir l'insaisissable

**Artikel:** Faut-il encore parler de "résultats statistiquement significatifs"?  
**Autor:** Fabbro, Thomas / Korner-Nievergelt, Fränzi  
**DOI:** <https://doi.org/10.5169/seals-866327>

### Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 13.01.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**



Oui,  
répond le statisticien  
Thomas Fabbro.

## Faut-il encore parler de «résultats statistiquement significatifs»?

Les résultats d'une étude doivent être statistiquement significatifs pour être considérés comme valables. Cette approche - le test d'hypothèse - est-elle vraiment désirable?



Non,  
répond la statisticienne  
Fränzi Korner-Nievergelt.

Photos: Valérie Chételat

La controverse autour des tests de signification est aussi ancienne que la méthode elle-même. La plupart des critiques se réduisent à cette alternative: les méthodes statistiques ne sont pas appliquées correctement ou leurs résultats sont mal interprétés. Bien réalisés, les tests d'hypothèse restent cependant utiles et importants.

La théorie de ces tests n'est certainement pas facile à comprendre. En revanche, les programmes informatiques rendent aujourd'hui leur réalisation très simple. Cela est utile pour de nombreux utilisateurs qui préfèrent se concentrer sur l'objet de leur recherche plutôt que sur des questions de méthodologie. Ils veulent pouvoir déterminer si l'effet observé est significatif ou non et témoignent alors une confiance aveugle à leur programme informatique. Malheureusement, «significatif» ne veut pas forcément dire «pertinent»... C'est ainsi que de très nombreuses études tirent des conclusions qui sont indéfendables du point de vue statistique. Les critiques ont donc raison sur certains points: il faudrait recourir moins fréquemment au test de significativité statistique basée sur la valeur-p. Il serait

souvent préférable de décrire l'effet estimé ainsi que sa précision.

Le test d'hypothèse reste néanmoins une méthode précieuse lorsqu'il faut prendre des décisions bien fondées et estimer précisément risques et bénéfices, par exemple lorsque l'on évalue un médicament ou une thérapie. Une bonne planification statistique est alors indispensable, ce qu'il faudrait fondamentalement améliorer dans de nombreux domaines.

## «Bien réalisés, les tests d'hypothèse restent utiles et importants.»

Par exemple, des scientifiques désirant montrer la différence entre deux traitements doivent absolument définir au préalable comment ils vont mesurer l'effet, et expliquer quelle intensité celui-ci devrait avoir pour avoir une pertinence concrète. Cette étape purement technique est indispensable durant la planification. On peut alors déterminer la bonne taille de

l'échantillon: s'il est trop petit, les effets pertinents ne pourront être mis en évidence, et s'il est trop grand, des effets non pertinents pourraient devenir statistiquement significatifs. De nombreux scientifiques n'ont pas idée du lien étroit entre la planification et l'interprétation des résultats d'un test d'hypothèse. C'est pourquoi je suis en faveur de la publication de ces informations avant même de réaliser une étude, comme le font certains journaux scientifiques. Et cette tendance va se renforcer.

Une bonne planification a de nombreux autres effets positifs sur le fonctionnement de la recherche. La discussion ne devrait donc pas se focaliser sur des questions isolées de méthode, mais sur le contexte global de la science.

Le biologiste Thomas Fabbro est responsable de l'infrastructure du Département de recherche clinique (Clinical Trial Unit) à l'Hôpital universitaire de Bâle.

### Résultats statistiquement significatifs, tests d'hypothèse et $p < 0,05$

Une étude tente de démontrer un effet à l'aide d'observations. Mais toute mesure produit naturellement des variations, et les différences observées pourraient ainsi être le fruit du hasard. La probabilité que ce soit le cas, appelée «valeur-p», peut être

calculée à l'aide de tests d'hypothèse. Si la valeur-p est suffisamment faible, les scientifiques partent de l'idée que l'effet est bien réel. Lorsqu'elle se situe au-dessous d'un seuil fixé à 5% ( $p < 0,05$ ), ils parlent de «résultats statistiquement significatifs».

Pour tirer des conclusions et prendre des décisions, on examine souvent si la valeur-p est inférieure à 5% ( $p < 0,05$ ). Malheureusement, le sens à donner à ce test s'avère si difficile à comprendre que la plupart des scientifiques ne l'interprètent pas correctement. L'idée qu'une valeur-p de plus de 5% prouve qu'il n'y a aucun effet est erroné, mais largement répandue.

Par exemple une étude canadienne a montré en 2017 que les enfants dont la mère avait pris un antidépresseur pendant la grossesse présentaient un risque d'autisme 1,6 fois plus grand que les autres. Mais la valeur-p se situait très légèrement au-dessus de 5% et l'effet n'était donc statistiquement pas significatif. Les auteurs ont alors faussement déduit que l'antidépresseur n'avait pas d'influence sur le risque d'autisme. Pourtant, même un risque 2,6 fois plus élevé aurait été compatible avec les données. De telles erreurs d'interprétation proviennent de la croyance que le résultat d'une étude s'inverse lorsque la valeur-p franchit la barre des 5%.

Les problèmes avec la valeur-p vont encore plus loin, et ceci même lorsqu'elle est correctement utilisée. Un biologiste

déclarant correctement qu'il n'a pas trouvé de lien significatif entre la durée d'ouverture de la chasse et la taille de la population de telle espèce, verra probablement un politicien en conclure qu'un tel lien n'existe pas. Au lieu d'utiliser la valeur-p, le biologiste aurait mieux fait de communiquer sur l'intensité de la corrélation. Cela permettrait d'estimer le développement futur de la population de l'espèce en fonction de la durée de l'ouverture de la chasse au mieux des connaissances actuelles. Des informations utiles pour trouver un compromis sur la durée de la chasse.

## «Réduire des résultats à une valeur-p ne permet pas de prendre une décision.»

Réduire des résultats scientifiques à une valeur-p en la comparant au seuil de 5% ne permet pas de prendre une décision, car elle fait disparaître les informations sur l'intensité des corrélations, et parce que l'issue de la comparaison est essentiellement

déterminée par la taille de l'échantillon. Or, la quantité de données réunies dans une étude ne change rien aux liens existant réellement dans la biologie.

Pour prendre une bonne décision, il faut l'adapter à la situation et se baser sur l'évaluation de différentes alternatives. Par exemple, des indices relativement faibles du recul d'une population animale suffisent à justifier des mesures de protection s'il s'agit d'une espèce fortement menacée, mais pas pour une espèce très répandue. Recourir de manière généralisée au même critère – comme le seuil  $p < 0,05$  conduit – à prendre des décisions arbitraires sans tenir compte des conséquences. C'est irresponsable.

La biologiste Fränzi Korner-Nierengelt est propriétaire du bureau de statistique Oikostat, chargée d'enseignement à l'ETH Zurich et collaboratrice de la Station ornithologique suisse de Sempach.