

Zeitschrift: Horizonte : Schweizer Forschungsmagazin
Herausgeber: Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen
Forschung
Band: 32 [i.e. 31] (2019)
Heft: 120: Überraschung! Wir zeigen Gefühle : Emotionen im nüchternen
Blick der Wissenschaft

Artikel: Braucht es statistisch signifikante Resultate noch?
Autor: Fabbro, Thomas / Korner-Nievergelt, Fränzi
DOI: <https://doi.org/10.5169/seals-866210>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. 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. [Siehe Rechtliche Hinweise.](#)

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. [Voir Informations légales.](#)

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. [See Legal notice.](#)

Download PDF: 14.03.2025

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



Ja

antwortet der Statistiker
Thomas Fabbro.

Braucht es statistisch signifikante Resultate noch?

In der Wissenschaft müssen
Resultate statistisch signifikant
sein. Das schafft Klarheit, steht
aber zunehmend in der Kritik.
Sind Hypothesentests noch
zeitgemäss?



Bilder: Valérie Chételat

Nein

antwortet die Statistikerin
Fränzi Korner-Nievergelt.

Die Kontroverse um Signifikanztests ist so alt wie die Methode selbst. Die meisten Kritikpunkte laufen auf das gleiche hinaus: Die statistischen Methoden werden falsch angewendet oder die Resultate daraus falsch interpretiert. Richtig durchgeführt, ist das Testen von Hypothesen jedoch weiterhin nützlich und wichtig.

Die Theorie dahinter ist sicher nicht einfach zu verstehen. Im Gegensatz dazu ist die Anwendung dank heutigen Computerprogrammen simpel. Das kommt vielen Anwendern sehr entgegen, da sie sich primär mit ihrem Forschungsthema und nicht mit der Methodik auseinandersetzen möchten. Sie möchten unterscheiden, ob ein Effekt signifikant ist oder nicht, und vertrauen dem Computerprogramm blind. Leider bedeutet signifikant nicht automatisch relevant. So werden aus sehr vielen Untersuchungen Schlussfolgerungen gezogen, die statistisch unhaltbar sind. Gewisse Argumente der Kritiker sind also richtig: Die statistische Signifikanz mit dem p-Wert sollte weniger zum Zug kommen. Häufig wäre es besser, den geschätzten Effekt und dessen Genauigkeit zu beschreiben.

Das Hypothesentesten bleibt jedoch eine sehr wertvolle Methode, wenn es darum geht, informierte Entscheidungen zu treffen und deren Nutzen und Risiken genau abzuwägen, zum Beispiel, wenn ein Medikament oder eine Therapie getestet wird. Eine unverzichtbare Grundlage für eine solche Anwendung ist jedoch eine gute statistische Planung, und diese muss in vielen Fachgebieten wesentlich verbessert werden.

«Wie eng die Planung mit der Interpretation der Resultate verbunden ist, ist vielen nicht bewusst.»

Möchten Forschende zum Beispiel zeigen, dass sich zwei Behandlungen unterscheiden, müssen sie zwingend im Vorfeld festlegen, wie sie den Effekt messen. Zusätzlich müssen sie begründen, wie gross der gesuchte Effekt mindestens sein soll, damit er inhaltlich relevant ist. Dieser rein fachliche Schritt ist für die Planung

unerlässlich. Basierend darauf kann auch die Stichprobengrösse berechnet werden: Ist sie zu klein, kann ein relevanter Effekt nicht gezeigt werden; ist sie zu gross, können irrelevante Effekte signifikant werden. Wie eng die Planung mit der Interpretation der Resultate eines Hypothesentests verbunden ist, ist vielen nicht bewusst. Um das Bewusstsein dafür zu schärfen, schlage ich deshalb vor, diese Informationen schon vor der Durchführung einer Studie zu veröffentlichen. Schon heute gibt es wissenschaftliche Fachmagazine, die solche Studienprotokolle vorab veröffentlichen. Der Trend wird sich fortsetzen.

Neben den Hypothesentests hat eine gute Planung noch viele weitere positive Auswirkungen auf den ganzen Wissenschaftsbetrieb. Der Fokus der Diskussion sollte deshalb weg von isolierten methodischen Aspekten hin zum wissenschaftlichen Gesamtkontext gelenkt werden.

Thomas Fabbro ist Biologe, Dozent an der Universität Basel und Leiter klinische Forschungsinfrastruktur am Departement Klinische Forschung (Clinical Trial Unit) des Universitätsspitals Basel.

p-Wert und statistisch signifikante Resultate

Eine Studie zeigt einen Effekt. Aber sind die beobachteten Unterschiede vielleicht zufällig? Messwerte variieren ja immer. Forschende berechnen deswegen die Wahrscheinlichkeit, dass die Unterschiede rein zufällig

entstanden sind: den p-Wert. Ist er klein genug, gehen sie davon aus, dass der Effekt echt ist. Liegt er unter der Schwelle von fünf Prozent ($p < 0,05$), sagen sie: «Die Resultate sind statistisch signifikant.»

Um Schlüsse zu ziehen und Entscheidungen zu treffen, wird oft getestet, ob der p-Wert unter fünf Prozent liegt ($p < 0,05$). Leider ist die Bedeutung des p-Werts so schwierig zu verstehen, dass die meisten Forschenden falsche Schlüsse daraus ziehen. Weit verbreitet ist die falsche Annahme, dass ein p-Wert über 5 Prozent beweise, dass kein Effekt vorhanden ist.

Zum Beispiel fand 2017 eine kanadische Studie bei Kindern, deren Mütter in der Schwangerschaft ein Antidepressivum eingenommen haben, ein 1,6 Mal höheres Risiko für Autismus. Der p-Wert lag aber ganz knapp über fünf Prozent, womit der Effekt statistisch nicht signifikant war. Daraus folgerten die Autoren fälschlicherweise, das Antidepressivum habe keinen Einfluss auf das Risiko für Autismus. Dabei wäre sogar ein 2,6 Mal erhöhtes Risiko immer noch mit den Daten vereinbar. Zu solchen Fehlschlüssen kommt es, weil der Glaube so verbreitet ist, ab einem p-Wert von fünf Prozent kehre sich die Schlussfolgerung einer Studie um.

Die Probleme mit dem p-Wert gehen noch weiter: Selbst wenn Forschende den p-Wert korrekt verwenden, kann es in der

Praxis schiefgehen. Die von Biologen richtig formulierte Aussage «wir fanden keinen signifikanten Zusammenhang zwischen der Dauer der Jagdzeit und der Population einer Tierart» ist für einige Politiker Beweis für das Fehlen eines solchen Zusammenhangs. Anstelle des p-Werts hätten die Biologen besser die Stärke des Zusammenhangs gezeigt. Das würde erlauben abzuschätzen, wie sich die zukünftige Population entwickelt, je nach Jagddauer, unter Berücksichtigung des aktuellen Wissensstands. Aufgrund dieser Information könnten sich die Interessengruppen auf eine Jagddauer einigen.

«Nur weil wir mehr Daten erheben, verändert sich ein biologischer Zusammenhang nicht.»

Eine datenbasierte Entscheidung ist nicht möglich, wenn das Resultat auf einen p-Wert grösser oder kleiner als fünf Prozent reduziert wird. Denn ob der Schwellenwert

unterschritten wird, darüber entscheidet vor allem die Stichprobengrösse. Nur weil wir mehr Daten erheben, verändert sich ein biologischer Zusammenhang aber nicht. Durch die Reduktion auf den p-Wert geht die Information über die Stärke des Zusammenhangs verloren.

Eine gute Entscheidung ist der Situation angepasst und basiert auf Abwägung verschiedener Alternativen. Bei einer stark gefährdeten Tierart genügen schwache Hinweise auf einen Populationsrückgang, um Schutzmassnahmen zu rechtfertigen. Hingegen werden wir bei einer weit verbreiteten Art erst handeln, wenn starke Hinweise auf eine Abnahme vorliegen. Ein für alle Anwendungen gleiches Kriterium, wie es $p < 0,05$ ist, führt zu willkürlichen Entscheidungen ohne Berücksichtigung der Konsequenzen. Das ist verantwortungslos.

Fränzi Korner-Nievergelt ist Biologin, Inhaberin des Statistikbüros Oikostat, Dozentin an der ETH Zürich und Mitarbeiterin der Schweizerischen Vogelwarte Sempach.