

20. März 2019, 19:06 Uhr Statistik

Signifikanter Unfug

Die statistische Signifikanz, gemessen mit dem sogenannten p-Wert, hat in der Wissenschaft eine geradezu götzenhafte Bedeutung erlangt. 800 Forscher beklagen Fehler und fordern ein Umdenken.

Von Patrick Illinger

Ist eine fragwürdige Münze gezinkt oder nicht? Das ist ein klarer Fall für Empirie. Man plane ein Experiment, welches die Hypothese "gezinkt" belegt oder widerlegt. In diesem Fall würde man die Münze wohl mehrmals werfen und notieren, wie oft sie Kopf oder Zahl zeigt.

Nehmen wir an, zwei Forschergruppen gehen ans Werk. Die erste Gruppe wirft die Münze zwanzigmal. Sie fällt zwölfmal auf Kopf. Ist sie also gezinkt? Die Intuition sagt ebenso wie die Formelwerke der Statistik: Nein, die Daten dieses Experiments sind nicht signifikant. Sie sind mit der Nullhypothese konsistent, also der These, dass das Geldstück nicht gezinkt ist. Die zweite Forschergruppe wirft die Münze 200-mal und sie fällt 120-mal auf Kopf. Nun ist laut Statistik evident: Mit der Münze ist etwas faul. Und schon gibt es einen herzhaften wissenschaftlichen Diskurs. Zwei Studien kommen zu widersprüchlichen Ergebnissen! Und das, obwohl die Münze in beiden Fällen mit gleicher relativer Häufigkeit Kopf zeigte.

Ist ein derart vereinfachtes Beispiel geeignet, um einen Missstand in der echten Wissenschaft zu illustrieren? Leider ja, wie ein Kommentar in der aktuellen Ausgabe der Zeitschrift Nature zeigt. Mehr als 800 Forscher haben den Appell unterschrieben. Die Verfasser, der Zoologe Valentin Amrhein von der Uni Basel sowie die amerikanischen Medizin-Statistiker Sander Greenland und Blake McShane, prangern den Missbrauch statistischer Kenngrößen an, insbesondere das Prinzip der "Signifikanz". Ihre Empfehlung lautet, die Signifikanz komplett abzuschaffen und die Ergebnisse von Experimenten nicht mehr in einfache Ja-Nein-Schemen zu pressen.

Fehlende statistische Signifikanz ist kein Gegenbeweis

Die Autoren zitieren ein frappierendes Beispiel aus der Medizin, das große Ähnlichkeit mit dem eben genannten Münzwurf-Beispiel hat. In zwei Studien wurden Hinweise auf schädliche Nebenwirkungen eines Medikaments überprüft. Es zeigte sich, dass die Patienten beide Male 20 Prozent häufiger an Nebenwirkungen erkrankten. Aber in einem Fall (den die Pharmaindustrie bevorzugte) blieb die Kenngröße für Signifikanz unter einer kritischen Schwelle. Die Studienautoren zogen daraus den Schluss: Seht her, es gibt keine Nebenwirkungen.

Die Unterzeichner des Kommentars in Nature fordern, derartige Schlüsse in Zukunft nicht mehr zu ziehen. Insbesondere sei ein statistisch nicht signifikantes Ergebnis keine Evidenz dafür, dass eine Hypothese widerlegt ist.

Dies mag für Laien wie eine fachinterne Diskussion klingen. Doch muss man sich klarmachen, dass die statistische Signifikanz, meist gemessen mit dem sogenannten p-Wert, eine geradezu götzenhafte Bedeutung in allen empirisch arbeitenden Wissenschaften erlangt hat. Ein p-Wert unter fünf Prozent (mitunter auch ein Prozent oder weniger) gilt in den allermeisten Studien, von den Sozialwissenschaften bis zur Pharmakologie, als Beweis für einen Hypothese. Fatalerweise wird ein höherer p-Wert als das Gegenteil angesehen, als Beweis der sogenannten Nullhypothese. Dieser Dichotomie, diesem Schwarz-Weiß-Denken möchten die Unterzeichner des Nature-Kommentars Einhalt gebieten.

"Wir sind es buchstäblich leid, all diese unsinnigen Beweise einer Nullhypothese zu sehen", schreiben Amrhein und seine Kollegen. Sie verweisen auf eine Analyse von 791 Fachpublikationen, die zeigt: Fast die Hälfte der untersuchten Studien kamen fälschlicherweise zu dem Schluss, dass fehlende Signifikanz gleichzusetzen sei mit einer Widerlegung der infrage stehenden Hypothese. "Wir müssen lernen, mehr Unsicherheit zu mögen", fordert der Nature-Kommentar.

Quelle:

<https://www.sueddeutsche.de/wissen/statistik-p-wert-signifikanz-hypothese-nullhypothese-1.4375636>