



DATENANALYSE

Tücken der **Interpretation**

Signifikanz

Schafft den p-Wert ab!

Fehlurteile

Wahrscheinlichkeiten richtig deuten

Metaanalysen

Schlechter als ihr Ruf



Antje Findeklee
E-Mail: findeklee@spektrum.de

Liebe Leserin, lieber Leser,
wir leben in einer Welt der Daten – aus Umfragen, Messprogrammen, Analysen und auch Berechnungen auf der Basis von anderen Werten. Die daraus erstellten Interpretationen sind Grundlage für manchmal weitreichende Entscheidungen – von der ureigenen persönlichen Ebene bis hin zu globaler Bedeutung.

Dementsprechend wichtig ist es, dass solche Daten zum einen korrekt erhoben und zum anderen mit den richtigen Werkzeugen ausgewertet werden – und das funktioniert nicht immer. Die Gründe dafür reichen von Nichtwissen bis hin zu absichtlicher Manipulation.

Eine interessante Lektüre wünscht

Erscheinungsdatum dieser Ausgabe: 30.01.2023

Folgen Sie uns:



CHEFREDAKTION: Dr. Daniel Lingenhöhl (v.i.S.d.P.)

CREATIVE DIRECTOR: Marc Grove

LAYOUT: Oliver Gabriel, Marina Männle

SCHLUSSREDAKTION: Christina Meyberg (Ltg.),

Sigrid Spies, Katharina Werle

BILDREDAKTION: Alice Krüßmann (Ltg.), Anke Lingg, Gabriela Rabe

REDAKTION: Antje Findeklee, Dr. Michaela Maya-Mrschtik

VERLAG: Spektrum der Wissenschaft Verlagsgesellschaft mbH,

Tiergartenstr. 15–17, 69121 Heidelberg, Tel.: 06221 9126-600,

Fax: 06221 9126-751; Amtsgericht Mannheim, HRB 338114,

USt-IdNr.: DE229038528

GESCHÄFTSLEITUNG: Markus Bossle

ASSISTENZ GESCHÄFTSLEITUNG: Stefanie Lacher

MARKETING UND VERTRIEB: Annette Baumbusch (Ltg.),

Michaela Knappe (Digital)

LESER- UND BESTELLSERVICE: Helga Emmerich, Sabine Häusser,

Tel.: 06221 9126-743, E-Mail: service@spektrum.de

BEZUGSPREIS: Einzelausgabe € 4,99 inkl. Umsatzsteuer

ANZEIGEN: Wenn Sie an Anzeigen in unseren Digitalpublikationen interessiert sind, schreiben Sie bitte eine E-Mail an anzeigen@spektrum.de.

Sämtliche Nutzungsrechte an dem vorliegenden Werk liegen bei der Spektrum der Wissenschaft Verlagsgesellschaft mbH. Jegliche Nutzung des Werks, insbesondere die Vervielfältigung, Verbreitung, öffentliche Wiedergabe oder öffentliche Zugänglichmachung, ist ohne die vorherige schriftliche Einwilligung des Verlags unzulässig. Jegliche unautorisierte Nutzung des Werks berechtigt den Verlag zum Schadensersatz gegen den oder die jeweiligen Nutzer. Bei jeder autorisierten (oder gesetzlich gestatteten) Nutzung des Werks ist die folgende Quellenangabe an branchenüblicher Stelle vorzunehmen: © 2023 (Autor), Spektrum der Wissenschaft Verlagsgesellschaft mbH, Heidelberg. Jegliche Nutzung ohne die Quellenangabe in der vorstehenden Form berechtigt die Spektrum der Wissenschaft Verlagsgesellschaft mbH zum Schadensersatz gegen den oder die jeweiligen Nutzer. Für unaufgefordert eingesandte Manuskripte und Bücher übernimmt die Redaktion keine Haftung; sie behält sich vor, Leserbriefe zu kürzen.

SEITE
04

FORSCHUNGSPRAKTIKEN
Wie sieht eine gute
psychologische
Studie aus?



ANYABERKUTZ / GETTY IMAGES / ISTOCK

SEITE
16

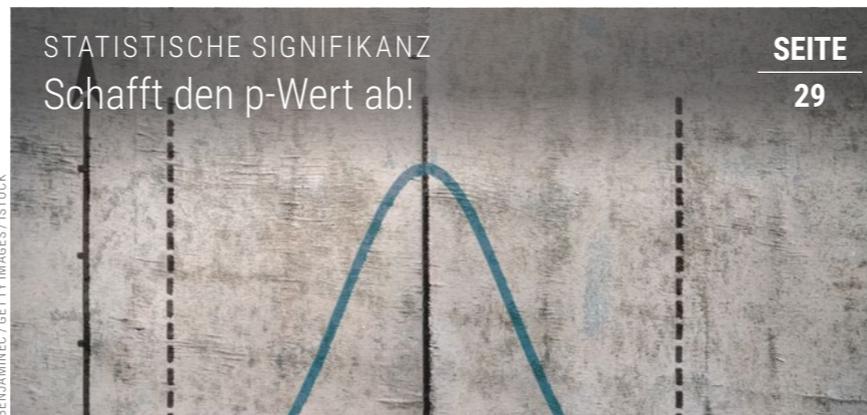
FEHLURTEILE
Wie Sie
Wahrscheinlich-
keiten richtig
deuten



MATERIO / GETTY IMAGES / ISTOCK

STATISTISCHE SIGNIFIKANZ
Schafft den p-Wert ab!

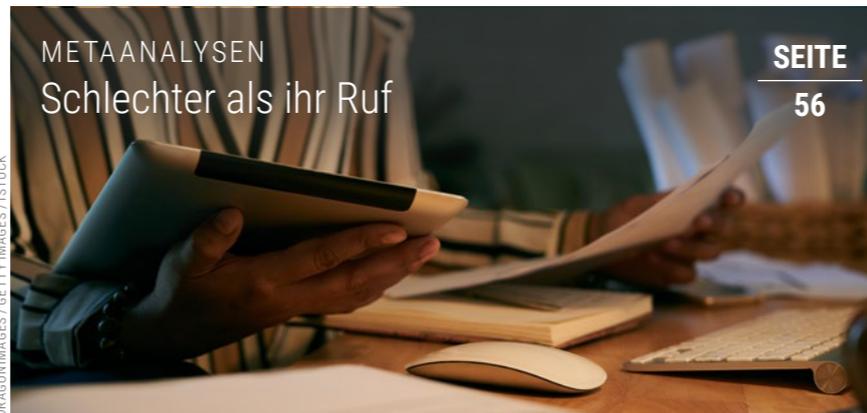
SEITE
29



BENJAMINEC / GETTY IMAGES / ISTOCK

METAANALYSEN
Schlechter als ihr Ruf

SEITE
56



DRAGONIMAGES / GETTY IMAGES / ISTOCK

- 10 METHODENKRITIK
Was psychologische Studien
wirklich aussagen
- 23 STATISTIK
»Ich habe gelernt, nicht auf meine
Intuition zu vertrauen«
- 28 DATENANALYSE
Wissenschaft im Datenraum
- 39 BLINDE FLECKEN IN DER FORSCHUNG
Schluss mit dem Selbstbetrug
- 44 FAKT UND FIKTION
Ich weiß weniger, als ich denke
- 47 IMPUTATION
Der Umgang mit fehlenden Werten
- 52 DIE FABELHAFTE WELT DER MATHEMATIK
Gegen jede Intuition
- 65 DETEKTOR.FM
Agnotologie – die Tricks der
Fake-Wissenschaft
- 66 »FAKT ODER FAKE?«
Statistische Kompetenz erlernen

FORSCHUNGSPRAKTIKEN

WIE SIEHT EINE GUTE PSYCHOLOGISCHE STUDIE AUS?



von Jan Schwenkenbecher

»P-Hacking«, »cherry picking«: Nur wer solche Schummeleien erkennt, kann den Wert einer empirischen Studie richtig einschätzen. Wie auch Laien gute von schlechten Praktiken unterscheiden können – ein Leitfaden.

Vor rund zehn Jahren geriet die Psychologie in die Schlagzeilen: Die Fälschungen des niederländischen Sozialpsychologen Diederik Stapel flogen auf und Daryl Bems vermeintliche Belege für »Präkognition«, eine Art übersinnliche Vorahnung, ließen sich vielfach nicht bestätigen. In der Folge scheiterten zahlreiche weitere Versuche, klassische Studien zu wiederholen, ihre Befunde zu »replizieren«. Die so genannte Replikationskrise erfasste auch andere Fächer. Eine der Hauptursachen: schlechte Forschungspraktiken. Und die waren schon vor den beiden Skandalen weit verbreitet, wie mehrere einflussreiche Arbeiten deutlich machten.

Damit kam Bewegung in die Psychologie. Immer mehr Forscherinnen und Forscher verzichteten seitdem auf zweifelhafte Methoden wie »p-Hacking« oder »cherry picking« und setzen stattdessen auf »Registered Reports« und »Open Data«. Was hinter diesen Begriffen steckt, erläu-

Jan Schwenkenbecher ist studierter Psychologe und arbeitet als Wissenschaftsjournalist im Rhein-Main-Gebiet.

tert der folgende Leitfaden. Er schildert gute Wissenschaft in vier Akten – von der Planung über Design und Auswertung bis zur Veröffentlichung.

1. Akt: Die Vorarbeit

Gutes wissenschaftliches Arbeiten beginnt schon mit der Formulierung der Forschungsfrage. Die erste Entscheidung steht an: »confirmatorisch« oder »explorativ« forschen? Wer confirmatorisch vorgeht, hat eine Hypothese und möchte sie überprüfen, im besten Fall: bestätigen (lateinisch: confirmare = bestätigen). Eine explorative Arbeit hingegen erkundet unbekannte Gewässer (lateinisch: explorare = erkunden) und entwickelt aus den Entdeckungen im Nachgang Hypothesen, die es dann in weiteren Studien zu prüfen gilt.

Entscheidend bei der confirmatorischen Hypothesen-Entwicklung: erst die Literatur lesen, dann daraus Hypothesen ableiten. Die Reihenfolge macht einen Unterschied, denn eine gut klingende Hypothese ist schnell erdacht und die passende Literatur gezielt herausgepickt, um die schöne Hypothese zu rechtfertigen. Dabei besteht die Gefahr, anders lautende Befunde beiseitezuschieben. Eine gute

AUF EINEN BLICK

Dos und Don'ts in Psycho-Studien

01 Planung und Design: Am Anfang steht meist die Hypothese, und die sollte sich aus der Fachliteratur ableiten – nicht erst im Nachhinein aus den Daten (»HARKing«). Wer die Studie vorab registriert oder sogar das Design im Peer-Review prüfen lässt, sorgt für Transparenz. Wie viele Versuchspersonen nötig sind, zeigt eine Poweranalyse.

02 Auswertung: Geschummelt wird vorzugsweise mit »p-Hacking«. Dazu zählt unter anderem, Versuchspersonen aus der Analyse auszuschließen, die nicht zu den erhofften Ergebnissen beitragen, oder die Erhebung dann zu stoppen, wenn die Daten das gewünschte Ergebnis bringen. Weil der p-Wert, das Maß für die statistische Signifikanz der Ergebnisse, mit vielen Problemen verbunden ist, arbeiten Fachleute heute zunehmend mit anderen Maßen wie Effektstärken und Konfidenzintervallen.

03 Veröffentlichung: Gute Forschung heißt auch, die Studie in jeder Hinsicht transparent darzustellen. Das bedeutet zum einen, dass die Daten auffindbar, zugänglich, verknüpfbar und wiederverwendbar sein sollten. Zum anderen soll der Artikel auch unbequeme Ergebnisse und etwaige Interessenkonflikte offenlegen.

Hypothese muss zwar nicht zur gesamten verfügbaren Literatur passen, sie aber vollständig berücksichtigen.

Eine zunehmend beliebte Variante des konfirmatorischen Arbeitens ist, die Studien von anderen zu replizieren, also zu wiederholen. Wozu? Jeder Befund gilt zunächst nur für den Ort, den Zeitpunkt und das Setting der betreffenden Studie. Doch je öfter eine Studie mit dem gleichen Ergebnis wiederholt wird, umso wahrscheinlicher wird es, dass es auch wirklich stimmt und auf andere Kontexte übertragen werden kann.

Und schließlich kann man bei der Studienplanung noch auf die Team-Zusammenstellung achten: Wer programmieren muss, der zieht einen Programmierer oder eine Programmiererin hinzu. Wer höhere Statistik braucht, sucht sich einen Statistikprofi. Ohnehin gilt: Interdisziplinäres Arbeiten, Großprojekte mit weltweiten Erhebungen und der Aufbau gemeinsamer Datenbanken liegen im Trend. Das mag zwar aufwändiger sein als eine Online-Fragebogenstudie mit 100 Versuchspersonen aus dem Familien- und Freundeskreis, doch dafür sind die Ergebnisse belastbarer.

2. Akt: Das Studiendesign

Steht die Forschungsfrage, leitet sich daraus oft schon das grobe Gerüst für das Studiendesign ab. Dennoch gibt es einige Punkte, die Forschende berücksichtigen können. Einer davon: nicht unnötig viele Variablen zu erheben.

Man könnte zum Beispiel noch schnell einen Persönlichkeitsfragebogen dazulegen, einen IQ-Test oder, oder, oder. Aber je mehr Variablen, je mehr mögliche Effekte und Zusammenhänge analysiert werden, desto wahrscheinlicher wird einer von ihnen durch reinen Zufall signifikant. Forschende erhöhen so zwar ihre Chancen, ihre Arbeit in einem wissenschaftlichen Journal zu veröffentlichen. Falsch ist so ein Vorgehen trotzdem, denn den vermeintlichen Zusammenhang gibt es eigentlich gar nicht. Unterm Strich schadet ein solcher Zufallsbefund der Forschung, denn in Replikationsstudien wird er nicht erneut gefunden. Was dem Problem vorbeugt: nur diejenigen Variablen zu erheben, die wirklich nötig sind, um die Hypothese zu testen, oder vorab festzulegen, auf welchen Variablen der Fokus liegen soll. Im Nachhinein lässt sich eine Analyse von

vielen Variablen notfalls statistisch korrigieren.

Stehen die Variablen fest, ist die nächste Frage, wie groß ein Effekt oder Zusammenhang ausfallen muss, damit er auch praktisch bedeutsam ist. Angenommen, es geht um die Wirksamkeit eines IQ-Trainings: Um wie viele Punkte müsste der IQ steigen? Ist der Mindesteffekt bestimmt, gibt eine so genannte Poweranalyse an, wie viele Versuchspersonen nötig sind, damit sich der gesuchte Effekt höchstwahrscheinlich in den Daten zeigt – wenn es ihn denn wirklich gibt.

Wer seine Stichprobengröße auf diese Weise festlegt, gerät auch nicht in die Versuchung, zwei weitere schlechte Forschungspraktiken anzuwenden, wenn die Studie schon läuft. Erster Fehler: den Zwischenstand immer mal wieder anzuschauen und die Studie zu stoppen, sobald sich in den Daten der erhoffte Effekt abzeichnet – auch wenn die angedachte Versuchspersonenzahl noch gar nicht erreicht ist. Zweiter Fehler: so lange immer mehr Teilnehmer zu testen, bis schließlich ein signifikantes Ergebnis vorliegt, auch wenn die geplante Stichprobengröße überschritten ist.