

Zeitschrift: Horizonte : Schweizer Forschungsmagazin
Herausgeber: Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung
Band: 34 (2021)
Heft: 131: Publizieren geht über Studieren

Artikel: Gleiche Information, andere Schlussfolgerung
Autor: Cartlidge, Edwin
DOI: <https://doi.org/10.5169/seals-1089076>

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: 12.01.2026

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

Gleiche Information, andere Schlussfolgerung

Haben Status, Geschlecht, Hautfarbe einen Einfluss auf Redezeit, Teilnahme, Bestrafung?
Studien dazu widersprechen sich manchmal sogar dann, wenn sie genau
dieselben Daten analysiert haben. Woran das liegt und was man dagegen tun könnte.

Text Edwin Cartlidge

Nehmen wir an, dass mehrere Forschungsteams denselben Datensatz mit dem Auftrag erhalten, bestimmte Hypothesen zu untersuchen. Wenn sie verantwortungsbewusst und sorgfältig vorgehen, sollten alle Gruppen ähnliche Ergebnisse erhalten. Zumindest denken wir normalerweise so über etablierte wissenschaftliche Methoden.

Doch dem ist nicht so. Zu diesem Schluss kommt eine Studie von Forschenden aus Sozialwissenschaften, Computerwissenschaften und Statistik, die im Juni 2021 in der Fachzeitschrift *Organizational Behavior and Human Decision Processes* veröffentlicht wurde. Unabhängige Forschende wurden beauftragt, rund vier Millionen Wörter aus einem akademischen Online-Forum zu analysieren und zwei vermeintlich recht einfache Hypothesen zu prüfen: nämlich ob Geschlecht und akademischer Status einen Einfluss auf die Forumbeiträge haben. Die Ergebnisse fielen verblüffend unterschiedlich aus, mit wilden Variationen sowohl bei den Analysemethoden als auch bei den Schlussfolgerungen – bis hin zu diametral entgegengesetzten Antworten.

Aus heiterem Himmel kommt diese Erkenntnis nicht. Schon in den letzten Jahren haben Studien gezeigt, dass viele Forschungsergebnisse in Disziplinen von Soziologie bis hin zur Medizin nicht reproduzierbar sind. Doch die Diskussion konzentrierte sich auf die Problematik, dass in verrauchten Daten krampfhaft nach statistischer Signifikanz gesucht wird oder die Ergebnisse für die Publikation verzerrt werden. Gemäss der neuen Studie sind jedoch auch unterschiedliche Forschungsmethoden ein grosses Problem.

Abraham Bernstein, Computerwissenschaftler an der Universität Zürich, hat am Projekt mitgewirkt und ist der Ansicht, dass nicht nur die Daten genau beschrieben werden sollten, sondern auch die Analyseschritte. Die Aussage «Wir haben den Test X gemacht» lasse zu viel Spielraum. «All diese Entscheidungen müssen klar dargelegt werden.»

Warum nicht die Offsideregeln anstatt Roter Karten?

Die neue Studie ist ein Beispiel für das, was als Crowdsourcing bezeichnet wird – die Rekrutierung unabhängiger Forschungsgruppen, meistens online, zum Analysieren eines Datensatzes. Bei einem Experiment von 2018 hatten die Forschungsteams die Aufgabe, Daten aus Fussball-Ligen auszuwerten und zu prüfen, ob ein Zusammenhang zwischen der Hautfarbe eines Spielers und der Anzahl erhaltener Roter Karten besteht. Die Schlussfolgerungen fielen unterschiedlich aus: Die meisten Forschenden fanden einen statistisch signifikanten, aber nicht besonders ausgeprägten Effekt, andere sahen keine Korrelation.

Während hier jedoch eine bestimmte Korrelation zu analysieren war – Rote Karten und nicht zum Beispiel Offsideregeln –, war es in der neueren Arbeit den Analysierenden überlassen, relevante Variablen zu definieren. Der Auftrag bestand darin, zwei Hypothesen zu kommentieren zu überprüfen, die 700 Personen, davon 128 Frauen

auf der Website *edge.org* über fast zwei Jahrzehnte abgegeben hatten. Die erste Hypothese lautete: «Die Bereitschaft, dass sich eine Frau aktiv an einer Konversation beteiligt, korreliert positiv mit der Anzahl an der Diskussion beteiligter Frauen.» Die zweite: «Teilnehmende mit höherem Status verfassen längere Beiträge als solche mit niedrigerem Status.»

Durchgeführt wurde die Studie im Rahmen eines internationalen Projekts unter der Leitung von Martin Schweinsberg, Psychologe an der European School of Management and Technology in Berlin. Aus einem ursprünglichen Pool von 49 beteiligten sich 19 Forschende daran. Mit der speziell konzipierten Website *Dataexplained* erfassten und erklärten sie ihre Analyseschritte – sowohl für die gewählten wie auch für die verworfenen Methoden.

Das Experiment zeigt, wie verschieden Analysen sein können. Verwendet wurde ein breites Spektrum statistischer Techniken und ein noch breiteres Spektrum von Variablen – zur Codierung des «Status» unter anderem die akademische Funktion, ein Dokortitel, die Zahl der Zitierungen und der h-Index für häufig zitierte Forschende.

Diese Methodenvielfalt widerspiegelte sich in uneinheitlichen Ergebnissen: Rund zwei Drittel der Forschenden kamen zum Schluss, dass Frauen häufiger mitdiskutieren, wenn andere Frauen beteiligt sind, ein Fünftel zog ein gegenteiliges Fazit. Bei der Frage zum Status variierten die Ergebnisse noch stärker: 27% bekräftigten die Hypothese, 20% widerlegten sie, und bei den übrigen resultierten statistisch nicht signifikante Werte.

Mit Voranmeldung gegen Vorurteile

Gewisse Stimmen ausserhalb des Projekts mahnen, dass diese Ergebnisse mit Vorsicht zu interpretieren seien. Leonhard Held von der Universität Zürich betont, dass die Forschenden bei diesem Projekt zwar die Variablen und statistischen Techniken frei wählen konnten, nicht aber das analysierte Forum. Dies habe die Ergebnisse möglicherweise beeinflusst, da die Stichprobengrösse die statistische Signifikanz schmälern kann. Er stellt auch die Frage, wie realistisch diese Analyse zur Verallgemeinerung der Gruppendynamik ist, da nur ein einziges Forum analysiert wurde.

Trotz dieser Einschränkungen begrüsst Held jedoch die Studie und ist der Ansicht, dass sie «klar zeigt, dass eine ausgeprägte Forschungsfreiheit bei explorativen Analysen gewisse Probleme mit sich bringt». Anna Dreber Almenberg von der Stockholm School of Economics in Schweden ist begeistert und beschreibt die Forschungsarbeit als «super wichtig» für das Anliegen, die Reproduzierbarkeit zu verbessern. Für sie kommt darin eine ungenügende «Pre-Registration» zum Ausdruck, bei der Forschende ihre Methoden und statistischen Tests festlegen, bevor sie Daten sammeln und analysieren. Eine Voranmeldung



Wird Denis Zakaria hier wegen seiner Hautfarbe diskriminiert? Daten liegen beim Fussball in rauen Mengen vor, doch welche genau analysiert werden sollen, ist unklar. Übrigens: Hatten Sie Vorurteile? Die Rote Karte beim Spiel Schweiz – Spanien galt Remo Freuler, der im Bild nicht zu sehen ist. Foto: Anton Vaganov/Pool/AFP/Keystone

kann die Verlässlichkeit der Ergebnisse verbessern, legt allerdings nicht endgültig fest, welche spezifischen Analysen durchgeführt werden.

Die Forschungsgruppe um Schweinsberg argumentiert denn auch, dass die Analysevielfalt «eine fundamentalere Herausforderung für die Forschung» darstelle als das p-Hacking oder das Peeking beim Sichten der Daten vor dem Test. Lösen lässt sich dieses Problem nach Ansicht der Gruppe durch eine Voranmeldung oder die Verwendung einer verblindeten Analyse, bei der nicht bewusst oder unbewusst eine Analysemethode gewählt werden kann, die das gewünschte Signal hervorbringt.

Dagegen führe die Bandbreite an Wissen, Überzeugungen und Interpretationen der Analysierenden zu unterschiedlichen Ergebnissen, selbst wenn diese Personen transparent und nach bestem Wissen und Gewissen handeln. «Subjektive Entscheidungen und deren Auswirkungen beruhen oft auf früheren theoretischen Annahmen und können den wissenschaftlichen Prozess in zentralen Aspekten prägen.»

Alle Entscheidungen explizit darlegen

Ein Projekt, das solche Variationen explizit machen will, ist Many Paths. Das von einer fünfköpfigen Gruppe aus Deutschland, den Niederlanden und der Schweiz geleitete Projekt will «the messy middle» im Forschungsprozess aufdecken. Dazu arbeiten praktisch tätige Wissenschaftlerinnen und Wissenschaftler aus verschiedenen Disziplinen an gemeinsamen Vorhaben. Der Stein wurde ins Rollen gebracht, indem Forschende aus Politikwissenschaften, Philosophie, Psychologie und weiteren Disziplinen aufgefordert wurden, eine uralte Frage zu diskutieren: Macht Macht korrupt? Rund 40 Expertinnen und Experten wirken bereits mit. Many Paths verwendet das Tool Hypergraph, mit

dem die Forschungsschritte modular dokumentiert werden. Wie Projektmitglied Martin Götz von der Universität Zürich erklärt, besteht die Idee darin, «das klassische wissenschaftliche Paper in Einzelteile zu zerlegen», sodass die Beteiligten das bearbeiten können, was sie am meisten interessiert, wie Theorie, Datenerhebung oder Metaanalysen. Das Fernziel besteht darin, «das heutige Publikationssystem zu ersetzen», das seines Erachtens mehr Anreize zum Generieren spektakulärer Schlagzeilen als für solide Forschung bereithält.

Grundsätzlich herrscht in der Forschungsgemeinde Einigkeit darüber, dass die Reproduzierbarkeit verbessert werden muss. Es wird aber auch vor allzu viel Pessimismus gewarnt. Eine dieser Stimmen ist John Ioannidis, Medizinerwissenschaftler und Epidemiologe an der Stanford University in den USA. Er argumentiert, dass es zwar Forschungsfragen gebe, bei denen die Ergebnisse sehr beliebig ausfallen können, dass jedoch in den meisten Fällen Folgeanalysen aufzeigen, dass gewisse Ergebnisse plausibler sind als andere. «Es gibt keinen Grund für radikalen Nihilismus», ist er überzeugt.

Auch Bernstein bleibt zuversichtlich und glaubt an einen soliden wissenschaftlichen Prozess, wenn die Forschenden klar offenlegen, welche Entscheidungen sie im Laufe ihrer Arbeit treffen. Wichtig sei, «dass sie diese Entscheidungen explizit darlegen und diese somit durch andere überprüfbar sind».

Edwin Cartlidge ist Wissenschaftsjournalist in Rom.