

Jacobsen, Jannes; Richter, David

Article — Accepted Manuscript (Postprint)

Gibt es repräsentative Umfragen?

Psychotherapie, Psychosomatik, Medizinische Psychologie

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Jacobsen, Jannes; Richter, David (2019) : Gibt es repräsentative Umfragen?, Psychotherapie, Psychosomatik, Medizinische Psychologie, Thieme, Stuttgart, Vol. 69, Iss. 5, pp. 203-204, <https://doi.org/10.1055/a-0859-8631> , <https://www.thieme-connect.de/products/ejournals/abstract/10.1055/a-0859-8631>

This Version is available at:

<https://hdl.handle.net/10419/222530>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

This is the accepted version of an article published in *Psychotherapie, Psychosomatik, Medizinische Psychologie* 69 (2019), 5, S. 203-204. The final authenticated version is available online at: <https://dx.doi.org/10.1055/a-0859-8631>

Gibt es repräsentative Umfragen?

Jannes Jacobsen & David Richter, DIW Berlin

Was wird erklärt?

Sozialwissenschaftliche Untersuchungen basieren häufig auf Umfragen. Gelten diese als repräsentativ so ist dies ein Gütesiegel für die Möglichkeit, Aussagen über eine Grundgesamtheit zu treffen. Was genau bedeutet „repräsentativ“ und welche Umfragen können als „repräsentativ“ bezeichnet werden?

Repräsentativität gibt es nicht

Wissenschaftliche Forschungsarbeiten auf Basis von Umfragen wollen in der Regel „repräsentativ“ sein. In der öffentlichen, nicht akademischen, Debatte ist dieser Begriff ein Ritterschlag – gleichbedeutend mit einer hohen Validität genutzter Daten. Aus der Sicht statistischer Verfahren und der Forschung auf der Basis von Befragungen ist dieser Begriff jedoch irreführend, da er impliziert, dass eine Stichprobe inhaltlich gleichbedeutend mit der Grundgesamtheit sei.

Mit repräsentativen Umfragen sind jedoch in der Regel einfache Zufallsstichproben gemeint. Eine einfache Zufallsstichprobe zeichnet sich dadurch aus, dass jede Person einer definierten Grundgesamtheit die gleiche Wahrscheinlichkeit hat, in eine Stichprobe aufgenommen zu werden (z.B. Diekmann, 2010; Schnell, Hill & Esser, 2013). Ist dies nicht der Fall (bei disproportionaler Ziehung bestimmter Fälle - wenn also bestimmte Personen eine höhere Wahrscheinlichkeit haben als andere, in die Befragung aufgenommen zu werden), so muss die unterschiedliche Ziehungswahrscheinlichkeit zumindest abgebildet werden können. Über die Nutzung von Gewichten ist es möglich, ungleiche Ziehungswahrscheinlichkeiten statistisch auszugleichen. Nur so kann die Aussagekraft von Zufallsstichproben erhalten werden, wenn durch die Ziehung der Stichprobe nicht für die gesamte Population die gleiche Ziehungswahrscheinlichkeit vergeben wird. Neben einer zufälligen Auswahl ist es daher ebenso wichtig, die Ziehungswahrscheinlichkeit aller Personen in der Population zu kennen.

Sind die Stichprobenmitglieder zufällig gezogen, kommt der Zentrale Grenzwertsatz zum tragen. Dies bedeutet, dass Schätzer verschiedener aber unter gleichen Bedingungen gezogener Zufallsstichproben (bspw. durchschnittliches Alter) normalverteilt um den wahren (aber unbekannt) Wert der Grundgesamtheit streuen. Auf Basis des zentralen Grenzwertsatzes kann der Stichprobenfehler bestimmt werden. Dieser gibt an, wie unsicher bestimmte Schätzer einer Stichprobe sind – also um den wahren Wert streuen. Grundsätzlich gilt: je größer die Stichprobe, desto geringer ist der Stichprobenfehler.

Repräsentativität (also die Gleichsetzung von Stichprobe und Grundgesamtheit) ist somit kein Begriff der Statistik, sondern ein Begriff des Alltags. Wissenschaftlich betrachtet gibt es nur den Unterschied zwischen Zufallsstichproben und nicht zufälligen Stichproben. Sind Stichproben nicht zufällig gezogen, kann der Stichprobenfehler nicht sinnvoll im Hinblick auf die Grundgesamtheit interpretiert werden.

Aussagen über die Grundgesamtheit

In der Natur von Zufallsstichproben liegt es, dass aus ihnen nur probabilistische Aussagen abgeleitet werden können. Somit ist davon auszugehen, dass Punktschätzer (bspw. Mittelwert, Median oder Effektgrößen in Regressionen) zwischen Stichprobe und Grundgesamtheit immer abweichen. Die Angabe von Punktschätzern auf Basis von Zufallsstichproben täuscht implizit vor, eine exakte Aussage über die Grundgesamtheit treffen zu können. Da bei Zufallsstichproben die Bestimmung des Stichprobenfehlers möglich ist, können aber zusätzlich sogenannte Konfidenzintervalle (oder Standardfehler) angegeben werden. Diese geben Aussage darüber, wie zuversichtlich Forscher auf Basis einer Stichprobe sind, dass der wahre Punktschätzer in einem bestimmten Wertebereich liegt. Ein 95% Konfidenzintervall bedeutet, dass durchschnittlich bei 100 Ziehungen unter gleichen Bedingungen 95% der Konfidenzintervalle den wahren Wert einschließen. Dies bedeutet auch, dass jeder Wert im Konfidenzintervall gleichwertig ist.

Auch Zufallsstichproben können selektiv sein

Der Charakter von Zufallsstichproben kann während der Befragungsphase stark eingeschränkt werden. Das größte Problem stellen dabei Personen dar, die an einer Befragung nicht teilnehmen wollen. Unterscheiden sich Befragte und nicht-Befragte systematisch, so geht der Charakter der Zufallsstichprobe verloren. Gewichte sollen dem entgegenwirken, indem sie die geringere Teilnahmebereitschaft bestimmter Personengruppen dadurch kompensieren, dass diese Personen in den Analysen stärker berücksichtigt werden, also ein größeres Gewicht bekommen. Im gleichen Zug werden dabei systematisch überrepräsentierte Personengruppen herunter gewichtet; also in den Analysen weniger stark berücksichtigt.

Die Gewichtungsverfahren stoßen dort an Grenzen, wenn die Teilnahmebereitschaft soweit sinkt, dass bestimmte Personengruppen kaum oder gar nicht für die Befragung gewonnen werden können. Wenn beispielsweise Personen mit einer Angststörung grundsätzlich nicht an einer Befragung teilnehmen wollen und sich diese Personen dann gar nicht oder nur in sehr geringer Zahl in der Stichprobe wiederfinden, dann kann diese systematische Verzerrung während der Feldphase auch nicht durch Gewichte hinreichend ausgeglichen werden. Dies ist insbesondere dann problematisch, wenn Eigenschaften der weniger befragungsbereiten Personen systematisch mit dem Untersuchungsgegenstand der Befragung zusammenhängen. Um beim Beispiel der Angststörung zu bleiben: Wenn Personen mit einer Angststörung besonders anfällig für Ein- und Durchschlafstörungen sind, besonders selten an Befragungen teilnehmen und eine Untersuchung die Wirksamkeit eines neuen Schlafmittels testen möchte, dann wird die Wirksamkeit dieses Mittels möglicherweise überschätzt. Daher werden auch immer wieder sogenannte Quotenstichproben herangezogen, um Aussagen über eine Grundgesamtheit zu treffen.

Quotenstichproben

Oft werden auch Quotenstichproben zu den sogenannten repräsentativen Stichproben gezählt. Diese befragen Personen entlang eines sogenannten Quotenplans, der bestimmt, wie viele Personen mit bestimmten Merkmalen an einer Umfrage teilnehmen sollen, um die Grundgesamtheit möglichst genau abzubilden. Quotenstichproben erfüllen dabei jedoch nicht die Kriterien von Zufallsstichproben und ihre Qualität ist in besonderem Maße von der Qualität des Quotenplans abhängig. Wenn bestimmte Merkmale der Population im Quotenplan nicht berücksichtigt werden, dann kann die Quotenstichprobe bezüglich dieser Merkmale verzerrt sein. Das ist wieder dann problematisch, wenn diese Merkmale systematisch mit dem Untersuchungsgegenstand der Befragung zusammenhängen. Um wieder auf das Beispiel der Angststörung zurück zu kommen: wenn 10% der Population eine Angststörung haben, Menschen mit Angststörung besonders häufig Ein- und Durchschlafstörungen haben und eine Untersuchung die Wirksamkeit eines neuen Schlafmittels testen möchte, dann sollte dieses Merkmal im Quotenplan auch in dieser Häufigkeit berücksichtigt werden. Da aber auch ein „perfekter“ Quotenplan nicht zu einer Zufallsstichprobe führt, können Stichprobenfehler und Konfidenzintervalle nicht im Hinblick auf die Grundgesamtheit interpretiert werden. Es sollte Aussagen von Quotenstichproben über eine definierte Grundgesamtheit somit nur Vertrauen geschenkt werden, wenn dem Quotenplan vertraut wird, der die Befragung bestimmt.

Bequemlichkeitsstichproben

Bei Bequemlichkeitsstichproben, den sogenannte Convenience Samples, findet die Stichprobenauswahl an Hand von einfacher Verfügbarkeit (z.B. Studenten oder Freunde) statt (z.B. Stangl, 2019). Diese Art der Stichprobengewinnung ist vor allem in der Psychologie üblich, wo häufig mit Studierenden gearbeitet wird. Da nicht abgeschätzt werden kann, ob und in welchem Ausmaß sich die bereitwilligen Studienteilnehmer von der Population, über die Aussagen getroffen werden, unterscheiden, ist die Verlässlichkeit solcher Stichproben unklar. Convenience Samples ermöglichen somit nur Aussagen über einen kleinen Teil der Grundgesamtheit (bspw. Studenten an einer bestimmten Universität) und sind damit im Grunde für viele wissenschaftliche Untersuchungen wenig geeignet.

Auch offene Befragungen ohne Stichprobenziehung, bei denen die Teilnahme nicht auf bestimmte Personen begrenzt ist, können verzerrt sein. Häufig erfolgt eine Selbstselektion in die Stichprobe durch Teilnehmer, die sich besonders für Befragungen im Allgemeinen oder für die konkrete Befragung im speziellen interessieren. Die Selbstselektion in die Stichprobe kann die Ergebnisse verzerren und sollte bei der Präsentation der Ergebnisse diskutiert werden.

Fazit für die Praxis

Wenn möglich sollten Zufallsstichproben genutzt und beim Berichten von Ergebnissen Konfidenzintervalle angegeben werden. Ist eine Stichprobe nicht zufällig, so sollte dies klar gekennzeichnet werden, damit die Aussagekräftigkeit der Analysen bestimmt werden kann.

Literatur

Diekmann, A. (2007). Empirische Sozialforschung. Grundlagen. Methoden. Anwendungen. 4. Überarbeitete Auflage (2010). Reinbek bei Hamburg: Rowohlt.

Schnell R., Hill, Paul B., Esser, E. (2013). Methoden der empirischen Sozialforschung. München: Oldenbourg

Stangl, W. (2019). Stichwort: 'Convenience Sample | Online Lexikon für Psychologie und Pädagogik'. Online Lexikon für Psychologie und Pädagogik.