

Anke Schöb

Repräsentativität – ein Begriff und seine Bedeutung

76 Prozent der Stuttgarterinnen und Stuttgarter leben gern in ihrer Stadt. Zumindest haben in der Stuttgart-Umfrage 2023 76 Prozent der Befragten so geantwortet. Weil es sich um eine repräsentative Befragung handelt, können wir annehmen, dass das Ergebnis dem Anteil in der Stuttgarter Bevölkerung entspricht. Der Begriff der Repräsentativität bezieht sich also auf die Eigenschaft einer Stichprobe (z. B. ausgewählte Personen in einer Befragung) ähnliche Merkmale aufzuzeigen wie die Grundgesamtheit (z. B. die Stuttgarter Bevölkerung). Dies stellt ein wichtiges Gütekriterium für eine Umfrage dar.

Doch was sind eigentlich die Voraussetzungen, um von einem repräsentativen Ergebnis ausgehen zu können? Um diesem Ziel der Ähnlichkeit von Stichprobe und Grundgesamtheit nahe zu kommen, müssen drei Fragen beantwortet werden:

- 1) Die Stichprobe ist ausreichend groß, um Merkmale der Gesamtheit zu repräsentieren.
(*Wann ist eine Stichprobe ausreichend?*)
- 2) Die Methode zur Auswahl der Personen aus der Gesamtheit, die teilnehmen sollen, ist zufällig.
(*Wann ist eine Auswahl zufällig?*)
- 3) Die Personen, die an der Befragung teilnehmen, unterscheiden sich nicht von Befragten, die nicht an der Befragung teilnehmen.
(*Welche Bedeutung haben Ausfälle?*)

1) Wann ist eine Stichprobe ausreichend?

Mit der Entscheidung für eine Stichprobe nimmt man eine Unsicherheit in Kauf. Das ist bei einer Vollerhebung nicht der Fall. Doch Vollerhebungen sind bei großen Gruppen (z. B. von allen 600 000 Stuttgarter Einwohnern) kosten- und zeitintensiv und bedeuten nicht zwingend ein besseres Ergebnis. Die Unsicherheit bei einer Stichprobenerhebung – so die gute Nachricht – ist allerdings berechenbar. Hierfür muss man vier Größen vorab festlegen:



- den gewünschten Grad an Sicherheit (Konfidenzniveau), mit der die Stichprobe den Wert der Grundgesamtheit beinhaltet,
- den tolerierbaren Stichprobenfehler, mit dem der Wert in der Stichprobe schwanken darf,
- wie sehr das Merkmal in der Grundgesamtheit variiert und
- wie groß die Grundgesamtheit ist, aus der die Stichprobe gezogen werden soll.

Ein Beispiel: Im Einwohnermelderegister sind 15 000 Jugendliche (Grundgesamtheit) im Alter zwischen 18 und 21 Jahren verzeichnet. Der Anteil von Frauen, so zeigt die Auswertung des Registers, liegt bei 50 Prozent. Dieser Anteil soll nun auch in einer Stichprobe bei 50 Prozent liegen, wir tolerieren jedoch einen Fehler von plus/minus zwei Prozent (d. h., der Anteil von Frauen liegt zwischen 48 und 52 %) und ein Konfidenzniveau von 95 Prozent (entspricht 1,96 Standardfehlern). Nun hat man alle vier Größen, die zur Berechnung der Stichprobengröße nötig sind (vgl. Tabelle 1).

Tabelle 1: Stichprobenumfang für endliche und unendliche Grundgesamtheiten

	Gesamtheit	95% Konfidenzniveau	Anteil	1-Anteil	Fehler	Stichprobengröße
a.	15 000	1,96	0,5	0,5	0,02	2070
	600 000	1,96	0,5	0,5	0,02	2391
b.	unendlich	1,96	0,5	0,5	0,02	2401

www.stuttgart.de/statistik

Die Größe der Stichprobe ergibt sich dann aus der Formel:

$$\text{a. } (1,96^2 \cdot 15000 \cdot (0,5 \cdot 0,5)) / (1,96^2 \cdot (0,5 \cdot 0,5) + 15000 \cdot (0,02)^2) = 2070$$

$$\text{b. } (1,96/0,02)^2 \cdot 0,5 \cdot 0,5 = 2401$$

Hieraus lässt sich folgern: Die Stichprobe müsste bei einer Gesamtheit von 15 000 Jugendlichen 2070 Personen umfassen, um mit einer Wahrscheinlichkeit von 95 Prozent einen Anteil von jungen Frauen zwischen 48 und 52 Prozent zu ermitteln. In der Tabelle ist auch ersichtlich, dass die Stichprobe bei einer Gesamtheit von 600 000 (z. B. Stuttgarter Bevölkerung) unter gleichen Bedingungen bei 2391 Personen liegt, also nicht wesentlich größer ist. Bei einer unendlich großen Gesamtheit liegt sie auch nur bei 2401. Hier lässt sich folgern: Die Repräsentativität einer Stichprobe wird nicht durch ihren relativen Anteil der Stichprobe an der Grundgesamtheit bestimmt, sondern durch die absolute Größe der Stichprobe.

In der Regel ist die Variation der Merkmale, die untersucht werden sollen, aber nicht bekannt, sodass zur Berechnung der Größe einer Stichprobe vom (in methodischem Sinne) schlechtesten Fall einer „50-prozentigen-Streuung“ ausgegangen wird (wie im Berechnungsbeispiel dargestellt). Die übliche Empfehlung bei Bevölkerungsumfragen legt eine Stichprobengröße von 1000 Befragten nahe.

2) Wann ist eine Auswahl zufällig?

Die einfachste Form von Zufallsstichproben erhalten wir, indem wir zum Beispiel für jede Person im Einwohnermelderegister zufällig eine Nummer vergeben und dann eine gewünschte Anzahl von Nummern auswählen. Diese Art der Stichprobe heißt einfache Zufallsstichprobe: Jede Person hat genau die gleiche Chance ausgewählt zu werden (für das Beispiel a aus dem vorherigen Abschnitt wäre die

Auswahlwahrscheinlichkeit für jede Person 2070/15 000 oder 2391/600 000). Für eine Zufallsauswahl ist nicht zwingend eine Liste der Gesamtheit notwendig. So ist eine Zufallsauswahl auch durch eine Zählung wie „jede*r fünfte Wähler*in eines Wahllokals“ (exit poll) oder „jede*r zehnte Besucher*in einer Webseite“ (Intercept-Befragung) möglich. Die Auswahlgesamtheit beschränkt sich dann per Definition auf die Wählenden im Wahllokal (nicht auf Briefwähler*innen) oder die Besucher*innen der Webseite (nicht auf Personen ohne Internetzugang). Ein Rückschluss auf eine weiterreichende Grundgesamtheit ist damit also nicht möglich (z. B. alle Wähler*innen). Entscheidend für eine Zufallsauswahl ist, dass die Auswahlentscheidung der Stichprobe bei der Person liegt, die die Daten erhebt, und die Auswahlwahrscheinlichkeit für jede Person berechenbar ist.

3) Welche Bedeutung haben Ausfälle?

Bei einer freiwilligen Umfrage auf der Basis von Zufallsstichproben wird es immer Nichtteilnehmer geben (z. B. aufgrund von Sprach-, Alters- oder Gesundheitsproblemen). Diese Antwortausfälle (Unit-Nonresponse) können in einer Befragung Einfluss auf die Qualität der Ergebnisse nehmen. So könnten bestimmte Haushaltstypen (ältere Menschen, junge Alleinlebende, Ausländer) unter den Ausfällen besonders häufig vorkommen, sich also erkennbare Muster zeigen. Eine gängige Praktik, um sich besser der Gesamtpopulation anzunähern, ist eine mögliche Gewichtung der erhobenen Daten. Erkennt man ein Muster, also etwa dass eine Gruppe unterrepräsentiert ist, verleiht man dieser Gruppe bei der Auswertung ein Gewicht größer 1. Dies ist eine mögliche Maßnahme, die allerdings eine Referenzstatistik (z. B. Einwohnermelderegister, Mikrozensus, Zensus) voraussetzt. Die Gewichtung ist deshalb meist nur für demografische Merkmale wie Alter, Geschlecht und Bildung möglich.

Die Gewichtungen werden als Ausfallgewichtung (Non-response-Adjustierung, Response-propensity-Technik) und Randanpassung (Kalibrierung, Poststratifizierung, Redressment) bezeichnet. Die Annahme dieser Verfahren ist, dass die verwendeten Korrekturmerkmale den Ausfall erklären beziehungsweise die Teilnehmer der Befragung stellvertretend stehen für die Nichtteilnehmer. Eine andere Möglichkeit, Ausfällen durch Nichtteilnahme zu begegnen, ist, nur Personen in die Befragung einzubinden, die sich aktiv zu einer (zukünftigen) Teilnahme an einer Befragung bereiterklärt haben. Hier gibt es nicht-zufallsgesteuerte Verfahren wie zum Beispiel im Bereich der Onlinebefragungen sogenannte Freiwilligen- oder Opt-in-Panels, bei denen sich die Teilnehmer selbst entscheiden teilzunehmen (selbstrekrutiert, das heißt aktives Eintragen in eine Liste). Ferner finden auch Verfahren Anwendung, bei denen die Teilnehmenden mit einem zufallsgesteuerten Auswahlverfahren auf Basis einer externen Liste (z. B. Telefonregister, Einwohnerregister) ausgewählt werden und sich dann in eine Teilnehmerliste eintragen.

Als ein Qualitätsindikator für eine Umfrage kann die Ausschöpfungsquote verwendet werden. Sie berechnet sich aus dem Verhältnis von teilnehmenden Personen zur Gesamtzahl eingeladenen Personen. Es trifft jedoch nicht

unbedingt zu, dass hohe Antwortquoten zu einem hohen Grad an Repräsentativität der Erhebung führen. So kann auch eine kleine Zahl an Nichtteilnehmenden systematisch von den Teilnehmenden abweichen. Umgekehrt können bei Erhebungen mit niedrigen Antwortquoten Eigenschaften der Bevölkerung korrekt wiedergegeben werden, wenn die Antwortenden im Vergleich zu den Nichtteilnehmenden nur zufällig (und nicht systematisch) variieren. Einen verbindlichen Schwellenwert für die Ausschöpfungsquote gibt es nicht; sie ist aber ein Orientierungswert und zeigt das Potenzial möglicher Verzerrungen an.

Einer möglichen Verzerrung der Ergebnisse einer Umfrage durch Ausfälle kann schon im Vorfeld begegnet werden beispielsweise durch die Gestaltung des Fragebogens beziehungsweise der Fragen, die Organisation der Feldarbeit, Erinnerungen oder nicht selektiven Anreizen zur Teilnahme an der Befragung.

Die drei genannten Aspekte – Stichprobengröße, Zufallsauswahl und Ausfälle – zusammengekommen bezeichnen den Begriff der Repräsentativität. Bei der Bewertung einer Umfrage kann man darauf achten, ob diese angesprochen werden und somit auf die Güte der Befragung als repräsentativ schließen. ●