

Repräsentativität

Der Begriff Repräsentativität ist in der Marktforschung viel diskutiert. Es heißt, eine Stichprobe, die nicht repräsentativ ist, würde keine sinnvollen Ergebnisse und Schlussfolgerungen erlauben. Allerdings: nicht jede Darstellung, die man zu diesem Thema findet, wird dessen Bedeutung auch gerecht. Eine Recherche im Internet fördert zahlreiche Beiträge zu dem Thema hervor, die in ihrer Kürze einen falschen Eindruck vermitteln (ein paar Hinweise finden sich im Anhang).

Eine Stichprobe ist genau dann repräsentativ, wenn man davon ausgehend auf die Gegebenheiten in der Grundgesamtheit schließen kann. Entsprechend wird eine Schlussfolgerung von der Stichprobe auf die Grundgesamtheit auch

Repräsentativitätsschluss genannt. Um eine repräsentative Stichprobe zu erhalten, fordert die Statistik lediglich, dass jede Person der Grundgesamtheit (oder welche Einheit auch immer einen Fall ausmacht) die gleiche Wahrscheinlichkeit besitzt, um in die Stichprobe aufgenommen zu werden. Was aber in der Theorie so einfach klingt, begründet in der Praxis eines der fundamentalen Probleme der Empirie.

Gedankenexperiment 1:

Wir wollten herausfinden, wie groß die Wahrscheinlichkeit sei, auf der Straße jemandem mit langen Haaren zu begegnen. Aus organisatorischen Gründen sind wir dazu gezwungen, die Erhebung an einem Nachmittag zwischen 14 und 16 Uhr in der Fußgängerzone durchzuführen. Dadurch werden die Ergebnisse verfälscht. In unserer westlichen Welt ist es nach wie vor recht üblich, dass die Männer einer Familie arbeiten, während die Frauen sich um die Familien kümmern und diese Zeit nutzen, um Besorgungen zu machen. Da Frauen wiederum sehr viel häufiger lange Haare tragen als Männer, werden wir die gesuchte Wahrscheinlichkeit überschätzen.

Aus Gedankenexperiment 1 ließe sich Folgendes ableiten. Das Merkmal "Geschlecht" ist in der Stichprobe ganz anders aufgeteilt als in der Grundgesamtheit. Die Stichprobe "repräsentiert" die Grundgesamtheit nicht hinreichend, um eine zuverlässige Aussage zu treffen. In diesem Zusammenhang wird häufig der Kern gesehen, warum eine Stichprobe repräsentativ sein müsse.

Repräsentativität der interessierenden Merkmale

Repräsentativität wird dementsprechend auch häufig definiert als dann gegeben, wenn die Verteilung aller interessierenden Merkmale in der Stichprobe auch der der Grundgesamtheit entspricht. Um eine repräsentative Stichprobe zu erhalten identifiziert der Forscher all diejenigen Variablen, auf die es bei seiner Fragestellung ankommt. Anschließend rekrutiert er eine Stichprobe, die hinsichtlich dieser Variablen eine

verkleinerte Kopie der Grundgesamtheit darstellt. Eine solche Stichprobe wird als **geschichtete Stichprobe** bezeichnet.

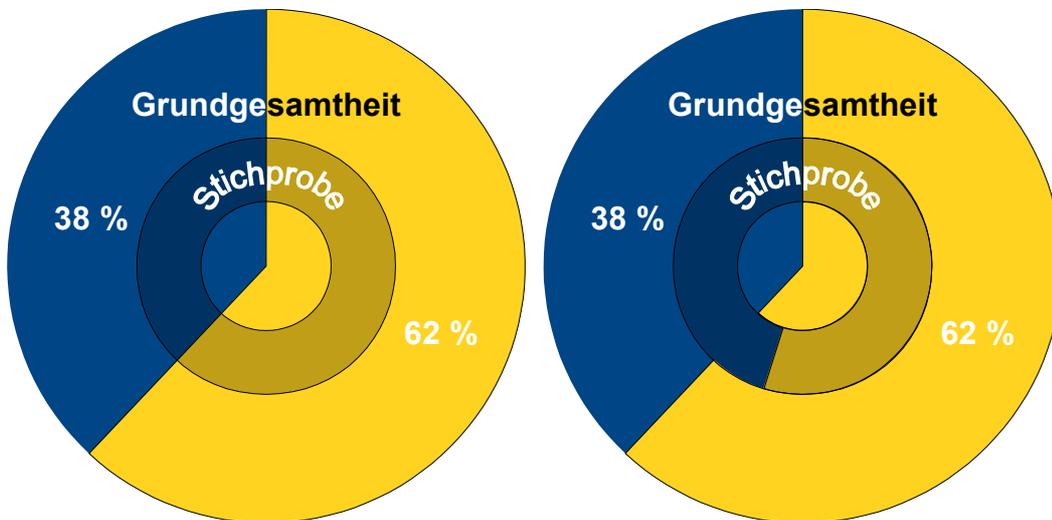


Abbildung 1: Links – die Stichprobe ist ein Abbild der Grundgesamtheit; rechts – blau ist in der Stichprobe stärker vertreten als in der Grundgesamtheit.

Diese Betrachtungsweise hat aber einen entscheidenden Haken. Was sind die *interessierenden Merkmale*? Es mag Beispiele geben, in denen dies sehr gut bekannt ist. Aber für welche Fragestellung existiert eine Forschungstradition, die lang genug ist, dass der Forscher diese Merkmale als bekannt ansehen kann? Die Wahlforschung mag ein Beispiel sein. In der Regel können wir über diese Merkmale sicherlich einige Spekulationen anstellen. Manche Spekulationen wird sich sogar solide begründen lassen. Aber – in der Regel – führen wir eine empirische Studie gerade deswegen durch, weil wir die interessierenden Merkmale nicht kennen und diese kennen lernen wollen! Ein zweiter Haken an diesem Verständnis besteht in der praktischen Umsetzung dieser Forderung. Wenn ich für jeden potentiellen Probanden möglichst unkompliziert entscheiden möchte, ob ich ihn/sie in die Stichprobe aufnehme, muss ich mich an Merkmale halten, die leicht zu bestimmen sind. Ein Screening nach komplexen Merkmalen wie Einstellungen oder Motiv-Konstellationen scheidet somit praktisch aus und es bleiben lediglich Oberflächenmerkmale wie Alter, Geschlecht oder sozialer Status. Diese sind aber nicht immer die wirklich interessanten Merkmale, sondern haben sehr häufig nur einen kleinen und indirekten Einfluss.

Der dritte Haken ergibt sich durch die statistischen Rahmenbedingungen. Wenn man es sehr genau nimmt, dürften die üblichen inferenzstatistische Verfahren bei einer geschichteten Stichprobe gar nicht verwendet werden (siehe z.B. Brogan, 1998). Und

Gedankenexperiment 2 zeigt, dass diese auch nicht nötig wäre, vorausgesetzt, die Stichprobe wäre tatsächlich perfekt geschichtet.

Gedankenexperiment 2:

Angenommen, wir wüssten über alle Variablen, die Haarlänge beeinflussen perfekt Bescheid und würden eine Stichprobe ziehen, die tatsächlich exakt repräsentativ wäre. Die Verteilungen, Mittelwerte, Varianzen der Stichproben ließen sich ebenso exakt auf die Grundgesamtheit übertragen.

Wozu würden wir also noch Statistik benötigen? Die Eigenschaften der Grundgesamtheit schlagen sich in diesem Fall exakt in der Stichprobe nieder und somit ist jede statistische Aussage unnötig.

In den meisten Fällen ist diese Form der Repräsentativität eine Krücke. Meistens ist sie zu kurz oder sie ist zu lang. Selten wird sie unserer Größe entsprechen. Und selbst wenn, haben wir in den wenigsten Fällen die Möglichkeit dies zu überprüfen und die Qualität unserer Daten einzuschätzen. Es wirkt, als sei sie der Versuch, die Wahrscheinlichkeitslehre aus zu tricksen. Da dies nun aber reichlich unrealistisch ist, dient sie allenfalls als Hilfskonstruktion und als Plausibilitätskriterium. Trotzdem gibt es Marktforscher, die diese Form der Repräsentativität über alles stellen.

Gedankenexperiment 3:

Überlegen Sie mal, welche Eigenschaften es sein könnten, die Haarlänge beeinflussen. Dann vergleichen Sie meine Liste am Ende des Dokuments mit Ihrer. Was habe ich übersehen? Woran haben Sie nicht gedacht?

Repräsentativität der Stichprobengewinnung

Auch wenn wir nicht durch die praktischen Unzulänglichkeiten der Wirklichkeit eingeschränkt wären und wir mit perfekten (sic!) Mitteln eine Stichprobe ziehen könnten, eine repräsentative Stichprobe ist nicht gewährleistet. Denn wer eine empirische Stichprobe zieht, geht zwangsläufig das Risiko ein, dass die Stichprobe die Gegebenheiten in der Grundgesamtheit eben nicht exakt widerspiegelt. Das liegt im Bereich des Zufalls. Genau deshalb gibt es die Statistik. Sie bestimmt die das Fehlerrisiko, wenn ich eine Stichprobe nehme und die Gegebenheiten auf die Grundgesamtheit übertrage.

Dies ist also kein Problem: das ist so. *Um eine gute Stichprobe zu erhalten, muss jede Person die gleiche Wahrscheinlichkeit besitzen, um in die Stichprobe aufgenommen zu werden.* Ein Verfahren zur Stichprobenziehung soll eben dies sicherstellen, denn in der Praxis ist es nicht so einfach, eine echte Zufallsauswahl zu gewährleisten. Für Deutschland gäbe es in den Einwohnermeldeämtern eine Liste aller Bundesbürger aus der man zufällig Personen heraus greifen könnte. Aber diese Daten sind (glücklicher-

weise) nur den Behörden zugänglich. Und deshalb ist es in der Praxis deutlich komplizierter, eine Stichprobe zu gewinnen, der kein Stichprobenfehler anhaftet. Es gibt verschiedene Ansätze. Sie nennen sich Geburtstagsmethode, "der Schweden-Schlüssel", das Random-Route-Verfahren, ADM-Verfahren, und so weiter (Schnell, Hill und Esser, 2005; siehe auch Tabelle 1). Einige wählen Haushalte aus einer größeren Einheit, andere wählen Personen aus Haushalten oder einer anderen Basis aus (siehe Tabelle 1). Ein einfaches Beispiel ist die Geburtstagsmethode. Sie dient dazu, aus einem Haushalt zufällig mit mehreren Personen zufällig eine auszuwählen. Man wählt genau diejenige, deren Geburtstag am weitesten zurückliegt. Die Geburtstage der untersuchten Personen verteilen sich nicht gleichmäßig über das Jahr. Trotzdem wird diese Methode v.a. in Telefoninterviews gerne eingesetzt, weil sie einfach durchzuführen ist.

Die korrekte Stichprobengewinnung ist der einzige Weg zu einer Stichprobe ohne Stichprobenfehler. Das gilt auch obwohl auch diese Ansätze nicht perfekt sind. Ein Beispiel: jedes Verfahren, dass zuerst Haushalte auswählt, um auf dieser Basis Personen zu ermitteln, gibt Personen aus kleineren Haushalten eine höhere Auswahlwahrscheinlichkeit. Das betrifft unter anderem das ADM-Verfahren, dass von allen größeren Marktforschungsinstituten eingesetzt wird. Die Geburtstagsmethode wiederum wählt in größeren Haushalten bevorzugt Personen aus, deren Geburtstag länger zurückliegt. Die bessere Alternative hier ist der Schweden-Schlüssel.

Tabelle 1: Beispiele für Stichprobenziehungsverfahren

	"Land" → Bezirk	Bezirk → Haushalt	Haushalt → Person
ADM	Zufallsauswahl von Bezirken	Random-Route	Schweden-Schlüssel
		RLD-Verfahren	Geburtstagsmethode
		Random Digit Dialing	
Mikrozensus	Zufallsauswahl von Bezirken		(Vollerhebung)
Allbus	1. Stufe des Allbus	2. Stufe: Zufallsauswahl nach Einwohnermelderegister ¹	

¹ Das Einwohnermelderegister ist wirtschaftlich und wissenschaftlich tätigen Forschern leider nicht zugänglich.

Schlussfolgerung

Repräsentative Stichproben gibt es in zwei "Ausbaustufen":

1. Repräsentativität der interessierenden Merkmale
2. Repräsentativität der Stichprobengewinnung

Prinzipiell ist Stufe 2 besser als Stufe 1. Dennoch sind die Zusammenhänge komplexer, als sie auf den ersten Blick scheinen. Empirische Studien werden nicht im Elfenbeinturm durchgeführt, sondern in der "freien Weltnis", und es ist schwierig, immer Stufe 2 gerecht zu werden. Sie ist aufwändig und teuer. Aber vielmehr: sie ist nicht für jede Fragestellung erforderlich. Egal ob die Studie einen wissenschaftlichen oder ökonomischen Hintergrund hat, das Kosten-Nutzen-Verhältnis muss berücksichtigt werden.

Stufe 1 ist sinnvoll:

- Zur rudimentären Qualitätssicherung in Panels (d.h. Datenbanken mit Personen, die sich zur Teilnahme an Studien bereit erklärt haben). Allerdings ist auch hier das Auswahlverfahren das deutlich bessere Qualitätskriterium.
- Als Ergänzung zu Stufe 2, um die statistische Genauigkeit zu erhöhen. Auf diese Weise lässt sich mit einer verkleinerten Stichproben eine gleichwertige Aussage treffen.
- Sie kann auch eingesetzt werden bei großangelegten Studien. In diesem Fall werden für eine bestimmte Region riesige Stichproben erhoben. Aufgrund der Größe der Stichprobe und ihrer geplanten Zusammensetzung geht man davon aus, dass die Daten tatsächlich direkt auf die Grundgesamtheit übertragbar sind. Hier wird keine Inferenzstatistik mehr durchgeführt. Die Verteilungen der erhobenen Merkmale werden rein beschreibend ausgewertet (da sich, wie oben beschrieben, eine statistische Analyse erübrigt, wenn die Stichprobe die Grundgesamtheit exakt widerspiegelt).

Beispiele für solche Studien sind die regelmäßigen Umfragen GfK-Konsumklima MAXX (2000 Probanden ab 14 Jahren) oder Allensbacher Computer- und Technik-Analyse (10.000 Probanden zwischen 14 und 64 Jahren). Für eine repräsentative Umfrage für **ganz** Deutschland müssen Stichproben ca. 1.000-3.000 Probanden kalkuliert werden inkl. aufwändiger Stichprobenziehungsverfahren (Schnell, Hill und Esser, 2005).

Unter keinen Umständen ist Stufe 1 ausreichend um eine "vertrauenswürdige" Repräsentativität zu erreichen. Sie kann lediglich als grobe Kontrolle dienen, wenn

vermutet wird, dass bei der Zufallsauswahl Verzerrungen aufgetreten sind (Schnell, Hill und Esser, 2005). Stufe 2 ist an sich immer angebracht, aber vor allem:

- Inferenzstatistische Auswertung (t-Tests, Varianzanalysen, ...)
- Data Mining im weitesten Sinne – damit sind statistische Auswertungen gemeint, die in einem Datensatz nach Strukturen suchen. Beispiele sind Faktoren²- oder Clusteranalyse, etc.

Repräsentativität in Teilnehmerdatenbanken

Eine Datenbank mit potentiellen Studienteilnehmern ist ein beliebtes Werkzeug. Wer über so etwas verfügt, hat eine mächtige Infrastruktur zur Hand. Aus Sicht der Repräsentativität ist sie aber bei weitem nicht so mächtig wie man glaubt. Eine Stichprobe aus einem Panel ist in erster Linie repräsentativ für das Panel selbst. Und sonst für gar nichts. Der Wärter des Panels ist lediglich in der Lage, seine Teilnehmer zu optimieren, damit sie hinsichtlich gewisser Oberflächenmerkmale (Alter, Geschlecht, Status, ...). Eine wirkliche Qualitätssicherung, die sich näher an den "interessierenden Variablen" orientiert ist nur dann möglich, wenn das Panel einem festen Zweck dient. Das GfK-Fernsehpanel zum Beispiel dient der Vorhersage von Fernsehverhalten. Die Wärter können deshalb regelmäßig verschiedene Methoden einsetzen und kreuzvalidieren die Struktur des Fernsehpanels. Bei Bedarf wird sofort korrigiert. Wenn das Panel aber keine derart klare Zielsetzung hat, dann sind Methoden zur Kreuzvalidierung nur unter großen Einschränkungen einsetzbar. Wenn heute Waschmittel untersucht werden, morgen berührungslose Interaktionstechniken am Computer, übermorgen die Vorlieben beim Geruch von Reinigungsmitteln, und am Tag danach die Preisgestaltung bei Floristik-Lieferunternehmen, und, und, und ... nach welchen Kriterien soll das Panel dann sinnvollerweise geschichtet werden?

Stichprobengröße und Repräsentativität

Die Stichprobengröße hat keinen Einfluss auf den Stichprobenfehler. Wenn die Datenerhebung fehlerbehaftet ist, kann auch Größe nicht helfen. Die Stichprobengröße entscheidet lediglich, ...

1. welche statistischen eingesetzt werden können
2. welche Zusammenhangs-/Unterschiedshypothesen getestet werden können.

Wenn kein Stichprobenfehler vorliegt, dann erhöht eine größere Stichprobe auf jeden Fall die Präzision. Wenn wir einen Stichprobenfehler nicht ausschließen können, dann

2 Wobei sich eine Faktorenanalyse auch konfirmativ rechnen lässt und sie dann eher als inferenzstatistisches Verfahren zu werten ist.

helfen mehr Probanden auch nicht weiter. Wenn das Auswahlverfahren allerdings ordentlich ist, dann kann auch eine Stichprobe von 2 Probanden repräsentativ sein. Diese Problematik fällt insbesondere bei Online Panels ins Gewicht. Stichproben aus einem Online Panel sind prinzipiell erst einmal repräsentativ für das Panel selbst. Ob sie auch auf andere Grundgesamtheiten übertragbar sind, hängt von der Qualität des Panels ab und vom Zufall.

Inhaltliche Repräsentativität

Der Begriff inhaltliche Repräsentativität (manchmal auch psychologische Repräsentativität) muss von dem oben beschriebenen Begriff klar abgegrenzt werden. Alles was ich bisher geschrieben habe bezieht sich auf die "statistische" oder "quantitative Repräsentativität". Die inhaltliche Repräsentativität ist in der Auseinandersetzung mit qualitativen Verfahren entstanden. Von Seiten quantitativer Forscher wird der qualitativen Forschung häufig vorgeworfen, sie sei weniger Aussagekräftig. Manche betrachten qualitative Forschung gar als unseriös (vor allem, wenn sie von einem wissenschaftlichen Standpunkt aus argumentieren).

Aus der Usability Forschung kennen wir das Phänomen, dass wir bei einer homogenen Stichprobe ab dem fünften bis achten Probanden immer wieder dieselben Schwachstellen beobachten (Nielsen, 2000). Bei weniger homogenen Stichproben, bei komplexen Produkten und/oder Aufgaben sind ein paar mehr Probanden nötig (Caulton, 2001; Woolrych & Cockton, 2001), aber trotzdem kommt man mit erstaunlich wenigen Probanden aus. Dabei sprechen wir hier nicht vom Bauchgefühl eines naiven Empirikers: die Gründe dafür wurden *statistisch* erörtert und belegt (Nielsen & Landauer, 1993). Inferenzstatistische Verfahren könnten bei derart kleinen Stichproben erstaunlich wenig ausrichten. Die Konfidenzintervalle sind derart breit, dass ein Repräsentationsschluss hinsichtlich der Häufigkeit eines Phänomens kaum Sinn macht.

Bei qualitativen Untersuchungen kann man mit wenigen Probanden oft alle wirksamen Faktoren aufdecken und verstehend beschreiben. Dass dies möglich ist, wird mit dem Begriff "inhaltliche Repräsentativität" auf den Punkt gebracht. Denn bei qualitativen Untersuchungen hängt die Qualität der Forschung nicht von der Anzahl der untersuchten Fälle ab, sondern mit der Qualität des Vorgehens. Es geht nicht darum, **wie viel** es von einer Sache gibt, sondern, **wie** diese Sache überhaupt beschaffen ist. In welchem quantitativen Ausmaß die Faktoren oder Mechanismen von Bedeutung sind, kann und will eine qualitative Studie nicht beantworten. Wenn dies für eine Fragestellung wichtig ist, dann können Folgestudien durchgeführt werden, die durch die qualitativen Ergebnisse auf ein solides Fundament aufbauen können.

Um von einer inhaltlich repräsentativen Stichprobe zu sprechen, wird bei der Rekrutierung auf unterschiedliche spezielle Testfälle/Probanden geachtet. Üblicherweise

verwendet man extreme Fälle, typische Fälle, und kritische Fälle. Konsequenterweise werden Probanden auch nicht zufällig ausgewählt, sondern gezielt aus der Grundgesamtheit herausgezogen.

Quellen

- Brogan, D. (1998) "[Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data](#)", Nachdruck mit Genehmigung der Encyclopedia of Biostatistics, Wiley, 1998.
- Caulton, D. A. (2001) "Relaxing the homogeneity assumption in usability testing", Behaviour & Information Technology, 2001, 20(1), 1-7
- Nielsen, J., and Landauer, Thomas K. (1993) "A mathematical model of the finding of usability problems," Proceedings of ACM INTERCHI'93 Conference, Amsterdam, S. 206-213.
- Nielsen, J. (2000) "Why you only need to test with 5 users". Alertbox, March 19, 2000. <http://www.useit.com/alertbox/20000319.html>
- Schnell, R., Hill, P., Esser, E. (2005) "Methoden der empirischen Sozialforschung", 7. Auflage, München: Oldenbourg
- Woolrych, A., and Cockton, G. (2001). "Why and when five test users aren't enough". In J. Vanderdonckt, A. Blandford, & A. Derycke (Eds.), Proceedings of IHM-HCI 2001 Conference: Vol. 2 (pp. 105- 108). Toulouse, France: Cépadèus.

Literatur – Gute, weil differenzierte Erläuterungen

[Marktforschungswiki](#)

[Arbeitsblätter Stangl](#)

v.d.Lippe / Kladroba, Repräsentativität, Beitrag zu "Marketing, 24 (2002), S. 227 -238

[Wikipedia](#)

[Marktforschung.de](#)

Literatur – Leider weniger gute Erläuterungen

Trotz ihrer Schwächen waren diese Darstellungen sehr hilfreich, da sie mir geholfen haben, mein Verständnis für das Thema zu schärfen.

[Beschreibung und Repräsentativität der Stichprobe](#)

[nhi² AG](#)

[Statistikberatung.eu](#)

Mögliche Einflussfaktoren von Haarlänge

Hier einige Vorschläge für die Einflussfaktoren von Haarlänge. Einige sind sicherlich miteinander korreliert. Welche nun die besten Prädiktoren darstellen ist unbekannt. Welche der Variablen als Prädiktoren überflüssig sind ist unbekannt. Welche der Variablen ursächlich sind ist unbekannt.

1. Kulturelle Herkunft
2. Alter
3. Sozialer Status
4. Beruf
5. Geschlecht
6. Haardichte
7. Neigung des Haars zur Lockenbildung
8. Musikgeschmack (Stichwort: Heavy Metal)
9. Sportliche Aktivitäten
10. Allgemeine Einstellung zur Körperpflege

Über dieses Dokument

Titel: Repräsentativität

Autor: Jan Seifert

Sprache: Deutsch

Version vom 04. Dezember 2012

Schlagworte: Repräsentativität, Repräsentanz; Empirie, Empirik, User Research, Marktforschung; Begriffsklärung; schließende Statistik, Inferenzstatistik; Repräsentativitätsschluss

Weitergabe: <http://creativecommons.org/licenses/by-nd/3.0/de/>

Die Inhalte in diesem Dokument spiegeln ausschließlich die persönlichen Ansichten des Autors wider.

