

Eine Frage der Grösse? Stichproben und Repräsentativität

"Ist denn das repräsentativ?" - mit dieser Frage wird Marketagent.com regelmässig konfrontiert. Zum einen, weil Online-Marktforschung selbst 2015 noch mit mangelnder Bevölkerungsabdeckung verbunden wird, zum anderen, weil ohne statistischen Background schwer vorstellbar ist, wie mit einer Stichprobe von beispielsweise 500 Personen die Gesamtbevölkerung abgebildet werden kann. Das kann sie streng genommen auch nicht, egal ob mit 500, 1.000 oder 5.000 Interviews.

Vorweg: Bis zum Alter von 49 Jahren stellt die Erreichbarkeit mittels Online-Befragung in der Schweiz heutzutage kein Problem mehr dar. Mehr als 97% dieser Personen sind bereits online. Erst bei den 50+ knickt die Online-Nutzung nach wie vor (noch) etwas ein, liegt aber auch schon bei knapp 73% (Quelle: Net-Metrix-Base 2014 2/1). Da die Festnetz-Nutzung kontinuierlich rückläufig ist und die Nummern der Mobiltelefone nicht (oder nur unzureichend) in offiziellen Listen wie Telefonbüchern registriert sind, wird die Online-Befragung immer mehr zur Methode der Wahl, wenn es um quantitative Marktforschungsstudien geht. Dennoch stellt sich für viele Marktforschungskunden die Frage, wie man zu einer repräsentativen Stichprobe gelangt - sei es nun on- oder offline.

Die Frage der Repräsentativität

Marktforschung hat oft die Aufgabe, deskriptive Verteilungen in einer Grundgesamtheit, zum Beispiel der Bevölkerung, abzuschätzen. Wie bekannt ist mein Unternehmen? Wie viel Prozent der Population hat meine Plakatkampagne wahrgenommen? Was sind die meistgenutzten Einkaufs- oder Informationskanäle für Produkt XY und gibt es dabei Unterschiede zwischen Männern und Frauen? Solche Studienresultate, aus denen strategische (Marketing-)Entscheidungen mit weitreichenden inhaltlichen und finanziellen Folgen abgeleitet werden, sollten wenn möglich nicht verzerrt sein, sondern die tatsächlichen Verhältnisse bestmöglich wiedergeben. Um diese Übertragbarkeit der Daten auf die Gesamtbevölkerung geht es, wenn in Zusammenhang mit Stichproben über "Repräsentativität" gesprochen wird, und sie gilt allgemein als zentrales Qualitätsmerkmal.

Dabei gibt es streng genommen gar keine statistische Repräsentativität. Repräsentativität ist ein theoretisches Konstrukt, das in der Realität nicht erfüllt werden kann - weder in der Markt- noch in der Hochschulforschung. Damit eine Stichprobe die Grundgesamtheit wirklich repräsentiert, müsste sie diese hinsichtlich sämtlicher Merkmale exakt abbilden. Das ist jedoch unmöglich, weshalb man sich der Grundgesamtheit in der empirischen Praxis jeweils nur annähern kann.

Quotenstichproben

Um eine Annäherung an die Grundgesamtheit sicherzustellen, wird in der Marktforschung vielfach mit sogenannten Quotenstichproben gearbeitet. Diese Stichproben werden zum Teil auch als quotierte Zufallsstichproben bezeichnet, weil streng genommen nur bei Zufallsstichproben ein Schluss von den Daten der Stichprobe auf die Verteilung der entsprechenden Eigenschaften in der Grundgesamtheit zulässig ist (induktive Statistik). Während jedoch bei Zufallsstichproben jedes Element aus der zu befragenden Grundgesamtheit die gleiche, von Null verschiedene Wahrscheinlichkeit besitzen muss, in die Stichprobe zu gelangen, werden bei Quotenstichproben Stichproben konstruiert, die

hinsichtlich zentraler Merkmale (zum Beispiel Alter und Geschlecht) der Grundgesamtheit, beispielsweise der Gesamtbevölkerung, entsprechen. Quotenstichproben erzeugen somit ein Abbild der Grundgesamtheit, das dieser zumindest in wesentlichen Eigenschaften entspricht. Dabei geht man davon aus, dass dadurch automatisch auch andere Eigenschaften wie das Bildungs- oder Einkommensniveau annähernd bevölkerungsabbildend dargestellt werden.

Am einfachsten lassen sich Quotenstichproben erzeugen, wenn man bereits vor Durchführung der Befragung Kenntnis über die zu steuernden Merkmale bei den Befragungsteilnehmern hat. Deshalb nutzen viele Marktforschungsinstitute wie auch Marketagent.com sogenannte Panels, Pools von potentiellen Respondenten, die sich bereit erklärt haben, an Befragungen des jeweiligen Instituts teilzunehmen und so zielgerichtet entsprechend den in ihren Panelprofilen hinterlegten demographischen Daten eingeladen werden können. Um zu vermeiden, dass zwar proportional genauso viele 14-19-Jährige in der Stichprobe sind wie in der Grundgesamtheit, diese aber nur aus Männern bestehen, werden die zu steuernden Merkmale bei seriösen Marktforschungsinstituten wie Marketagent.com zusätzlich kreuzquotiert. Das heisst, man steuert nicht nur den Anteil der 14-19-Jährigen und den Anteil von Männern und Frauen in der Gesamtstichprobe, sondern auch den Anteil von Männern und Frauen innerhalb jeder einzelnen Altersgruppe.

Bei Quotenstichproben werden die Stichproben gemäss ausgewählter Eigenschaften gesteuert, die als besonders wichtig erachtet werden. Diese können je nach Studienansatz variieren. Es ist jedoch kaum bzw. nur verbunden mit erheblichem Mehraufwand, Zeit und entsprechenden Kosten möglich, mehr als drei Eigenschaften gleichzeitig zu kontrollieren. Repräsentativität in der Marktforschung ist daher immer ein Kompromiss. Es geht darum, qualitativ hochwertige Stichproben zu erzeugen, die die Grundgesamtheit beziehungsweise deren Handlungen, Meinungen und Erfahrungen bestmöglich abbilden.

Die Frage der richtigen Grösse

Grundsätzlich gilt: Je grösser, die Stichprobe, desto genauer die Ergebnisse. Mit zunehmenden Interviewzahlen nähern sich die Daten immer mehr dem tatsächlichen Ergebnis in der Grundgesamtheit an, die Stichprobengenauigkeit nimmt zu. Während man bei einer Stichprobe von $n = 100$ noch mit einer Schwankungsbreite von 3,4 bis maximal 10 Prozentpunkten rechnen muss, die Stichprobe also relativ ungenau ist, nimmt diese bis zu einem n von 500 auf 1,5 bis 4,5 Prozentpunkte ab. Bei $n = 1.000$ Netto-Interviews kann der wahre Wert nur noch um 1,1 bis maximal 3,2 Prozentpunkte vom erhobenen Wert abweichen (jeweils bei einer Irrtumswahrscheinlichkeit von 5 Prozent).

Diese Schwankungsbreiten gelten aber nur, wenn die Stichprobe nicht beispielsweise durch Über- oder Unterrepräsentation einzelner Bevölkerungsgruppen verzerrt, sondern "sauber" ist. Eine Stichprobe von $n = 5.000$ mag beeindruckend klingen, aber sie liefert unter Umständen schlechtere Daten als eine Stichprobe von $n = 500$. Wenn sie systematische Verzerrungen aufweist, weil beispielsweise die Vertreter einer bestimmten Interessensgruppe darin klar überrepräsentiert sind, wächst der Fehler mit zunehmender Stichprobengrösse sogar noch. In Zusammenhang mit Stichprobenziehung gilt also: Lieber etwas kleiner und dafür "sauber" als gross und schief.

Das heisst aber nicht, dass eine "sauber" quotierte Stichprobe von $n = 100$ Personen für alle Studienzwecke ausreichend ist. Je nachdem, welche Auswertungen und statistischen Analysen angestrebt werden, benötigt man deutlich mehr Netto-Interviews. Die Stichprobengrösse zählt also doch, aber weniger in Zusammenhang mit "Repräsentativität" als in Verbindung mit dem benötigten Mass an Stichprobengenauigkeit und der Weiterverwendbarkeit der Daten.

Nachträgliche Gewichtung

Wenn es von vornherein nicht möglich ist, eine "saubere" Stichprobe zu erzeugen oder die Stichprobe aus anderen Gründen schief ausfällt, also in ihrer Struktur hinsichtlich zentraler Merkmale wie Alter oder Geschlecht nicht der Grundgesamtheit entspricht, gibt es die Möglichkeit, die gewonnenen empirischen Daten nachträglich zu gewichten. Enthält eine Stichprobe beispielsweise nur 30% Frauen und 70% Männer, wird für die Auswertung die Meinung der Frauen hochgewichtet, die der Männer hingegen runtergewichtet, so dass das Verhältnis letztlich dem in der Gesamtbevölkerung entspricht. Statistisch ist dieses Vorgehen aber nicht unumstritten, weshalb bei einer Studie, die das Ziel verfolgt, die Verteilungen innerhalb der Gesamtbevölkerung zu schätzen, eine "saubere" Ausgangsstichprobe grundsätzlich zu bevorzugen wäre.

Marktforschung in Zusammenhang mit politischen Abstimmungen

Befragungen zu Werthaltungen, politischen Meinungen oder Abstimmungsverhalten in der Bevölkerung sind in der Politikwissenschaft und bei den Medien sehr beliebt.

Genau wie andere sozialwissenschaftliche oder Marktforschungsstudien auch, versuchen Abstimmungs- oder Wahlstudien immer, sich der Realität mittels wissenschaftlicher Methoden und nach den Gesetzen der Wahrscheinlichkeitsrechnung anzunähern. Sie können jedoch kein genaues Abbild erzeugen. Von Abstimmungs- oder Wahlstudien wird aber erwartet, dass sie eine genaue Prognose von Gewinnern und Verlierern zulassen. Ein Beispiel: Erreicht eine Vorlage eine Zustimmungsrate von 51% und die Umfrage hat ein Resultat von 55% prognostiziert, dann wird die Prognose als treffend betrachtet. Lag das Resultat der empirischen Studie jedoch bei 49%, wird sie als ungenügend betrachtet, weil sie einen falschen Ausgang vorhergesagt hat, obwohl sie statistisch gesehen näher am effektiven Resultat lag.

Aufgrund des grossen öffentlichen Interesses an Studienergebnissen aus dem Bereich der politischen Forschung sind sämtliche dem Verband Schweizer Markt- und Sozialforschung angeschlossenen Institute wie Marketagent.com dazu verpflichtet, bei der Durchführung entsprechender Studien spezielle Standesregeln einzuhalten. So dürfen beispielsweise zehn Tage vor dem offiziellen Abstimmungstag keine Ergebnisse aus bevölkerungsabbildenden Studien mehr veröffentlicht werden. Denn erst während dieser letzten „heissen Phase“ des Abstimmungskampfs fällt ein bedeutender Teil der Bevölkerung seine Abstimmungsentscheidung und die Umfrageergebnisse könnten die Meinungsbildung der Stimmbürger beeinflussen.

Ausserdem müssen die Stichproben bei Umfragen zu Abstimmung und Wahlen immer mindestens $n = 1.000$ Personen umfassen. Um auch Aussagen zu einzelnen Segmenten machen zu können, muss man sich mindestens auf eine Fallzahl von $n = 200$ beziehen können. Die vollständigen Standesregeln der vsms swiss interview institute® für Umfragen, Wahlbörsen und Medien-Analysen zu Wahlen und Abstimmungen, die zur Publikation bestimmt sind, finden Sie hier:

http://www.vsms-asms.ch/files/6314/1899/9729/Standesregeln_-_neues_CI.pdf

Die berühmte Ausnahme von der Regel

Natürlich gibt es bestimmte Fälle, in denen es nicht möglich ist, eine "saubere" Zufalls- oder Quotenstichprobe zu erzeugen. Das ist zum Beispiel der Fall, wenn man schwer erreichbare Personen - wie solche mit einer bestimmten Krankheit oder Kunden einer Privatbank - befragen möchte. In solchen Fällen ist der Wert jedes einzelnen Interviews grösser als der Wert der Repräsentativität oder die Einhaltung bestimmter Quoten innerhalb der Stichprobe. Man versucht dann, so viele Interviews wie möglich zu generieren. Dabei geht es aber auch meist nicht (oder nicht in erster Linie) darum, etwas über die tatsächliche Verteilung bestimmter Merkmale in der entsprechenden Bevölkerungsgruppe zu erfahren, sondern zum Beispiel darum, vertiefte Insights aus dem Usage&Attitude-Bereich zu erhalten. Wenn gar

nicht im Vordergrund steht, tatsächliche Verteilungen in der Gesamtbevölkerung aufzudecken, sondern stattdessen beispielsweise Hypothesen zu testen, sind repräsentative Stichproben verzichtbar.

Genutzte und weiterführende Literatur zum Thema Stichprobenziehung und Repräsentativität

- Beller, Sieghard: Empirisch forschen lernen. Konzepte, Methoden, Fallbeispiele, Tipps. 2., überarb. Aufl. Bern 2008.
- Diekmann, Andreas: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek b. Hamburg 2002. (Auch in neuerer Auflage erhältlich)