

Peter von der Lippe, Andreas Kladroba  
**Repräsentativität von Stichproben**

Prof. Dr. P. von der Lippe  
Universität Essen  
Fachbereich Wirtschaftswissenschaften  
Universitätsstr. 12  
45117 Essen  
Tel. 0201/183-3639  
[plippe@vwl.uni-essen.de](mailto:plippe@vwl.uni-essen.de)

Dr. Andreas Kladroba  
Universität Essen  
Fachbereich Wirtschaftswissenschaften  
Universitätsstr. 12  
45117 Essen  
Tel. 0201/183-2302  
[kladroba@vwl.uni-essen.de](mailto:kladroba@vwl.uni-essen.de)

**Schlüsselworte:** Statistik, Repräsentativität, Stichprobentheorie

**keywords:** statistics, representativity, sampling theory

# Repräsentativität von Stichproben

**Zusammenfassung:** Repräsentativität gilt als eines der wichtigsten Gütekriterien für Stichproben in empirischen Untersuchungen. Die vorliegende Arbeit widerspricht der allgemeinen Auffassung über die Notwendigkeit der Repräsentativitätseigenschaft und versucht ihre Schwächen und Widersprüchlichkeit vor allem im Verhältnis zu anderen Konzepten der Stichprobentheorie aufzuzeigen.

**Abstract:** Representativity is one of the most popular qualities which in common opinion samples have to get. This paper discusses the principle of representativity compared to other concepts of the sampling theory.

## 1. Einleitung

Wenn es ein Konzept in der Statistik gibt, von dem im Alltag jeder glaubt, dass es ihm geläufig ist und dass es unverzichtbar ist, ist es die „Repräsentativität“. Während ein großzügiger Umgang mit diesem Schlagwort in den Nachrichten oder Tageszeitungen aus einer gewissen fachlichen Unwissenheit von Journalisten und Redakteuren notfalls verständlich ist, erstaunt es um so mehr, dass der Begriff der Repräsentativität in betriebswirtschaftlichen empirischen Untersuchungen mindestens genau so überstrapaziert wird. Dabei wird auch hier häufig auf eine nähere Erläuterung, was darunter zu verstehen sei, verzichtet, wohl in der Annahme, dass das Konzept der Repräsentativität allgemein bekannt ist.<sup>1</sup>

Es ist sicherlich Tiede/Voß zuzustimmen, wenn sie leicht ironisch anmerken: „Einem Ergebnis, das nicht repräsentativ ist, haftet ein Makel an. Mit ihm lässt sich nicht viel Staat machen. [...] Repräsentativität ist als Gütesiegel für ein Forschungsergebnis begehrt, weil das repräsentative Ergebnis den Eindruck der Verlässlichkeit, Glaubwürdigkeit und Verallgemeinerungsfähigkeit erweckt.“ (Tiede/Voß 2000, S. 84)

Betrachtet man die gängige Lehrbuchliteratur zur Statistischen Methodenlehre, wird man das Stichwort „Repräsentativität“ in der Regel vergeblich suchen. Dies hat der Statistik als Wissenschaft oftmals den Vorwurf eingebracht, sie würde nur ein einem elfenbeinturmartigen Denken entsprungenes mathematisches l'art pour l'art betreiben und sich nicht um praxisrelevante Probleme der Anwender kümmern.<sup>2</sup> Es stellt sich die Frage, ob die eher anwendungsorientierte Literatur, also vor allem die zur Betriebswirtschaftslehre und zur Marktforschung, hier hilfreicher ist oder ob man mit Tiede/Voß sagen muss: „Repräsentativität ist ein schillernder Begriff, der häufig verwendet wird, obwohl ziemlich unklar ist, was er zum Inhalt hat.“

Zunächst ist aber mit einem weit verbreiteten Vorurteil aufzuräumen (vgl. z.B. Herrmann/Homburg 1999, S. 63 und Pepels 1998): Der „Idealfall“ einer empirischen Untersuchung besteht auf keinen Fall darin, möglichst eine Totalerhebung vorzunehmen und auf Stichproben nur im Notfall zurückzugreifen („Hilfskonstruktion“). Davon abgesehen, dass Totalerhebungen vor allem bei großen Grundgesamtheiten mit einem gewaltigen Aufwand und damit Kosten verbunden sind, ist vor allem anzumerken, dass die oft angenommene Fehlerfreiheit von Totalerhebungen ganz sicher

<sup>1</sup> Für die unterschiedlichen Bereiche der betriebswirtschaftlichen Empirie vgl. z.B. Pekruhl/Nordhause-Janzen 2000, Werani 2000, Steffens-Duch 2000, Karsch 2000 oder Keller/Lerch/Matzke 2000. Natürlich gibt es auch in der Betriebswirtschaftslehre kritische Stimmen bezüglich des Konzepts der Repräsentativität (z.B. Bausch 1995). Allerdings stellen diese eine Minderheit dar.

<sup>2</sup> Dieser Vorwurf mag in anderen Bereichen durchaus berechtigt sein. Hier ist er es allerdings eindeutig nicht.

nicht gegeben ist, wie die Auszählung der jüngsten Wahlen zum Präsidenten der USA eindrucksvoll bewiesen hat. Stichprobenverfahren sind ein probates Mittel zur empirischen Analyse und sicher kein Instrument zweiter Klasse.

Die gängige Vorstellung von Repräsentativität lässt sich wohl am besten folgendermaßen beschreiben: Die Auswahl einer Teilgesamtheit ist so vorzunehmen, dass „aus dem Ergebnis der Teilerhebung möglichst exakt und sicher auf die Verhältnisse der Gesamtmasse geschlossen werden kann.“ Dies ist dann der Fall, „wenn sie [die Teilerhebung] in der Verteilung aller interessierenden Merkmale der Gesamtmasse entspricht, d.h. ein zwar verkleinertes, aber sonst wirklichkeitsgetreues Abbild der Gesamtheit darstellt.“ (Berekoven u.a. (1999), S. 50)<sup>3</sup> Das Problem besteht allerdings darin, dass gerade die Verteilung der „interessierenden Merkmale“ in der Grundgesamtheit unbekannt ist. Ansonsten würde sich die angestrebte Untersuchung ja auch erübrigen. Daher wird in der Regel vorgeschlagen auf Merkmale auszuweichen, die mit den zu untersuchenden hoch korrelieren. Dass dieses Bestreben zum Teil recht absonderliche Blüten treibt, zeigt folgendes Beispiel: „Das Ziehen einer Stichprobe aus den Studenten einer Hochschule für die Durchführung eines Produkttests einer Joghurtmarke ist ein relativ schnelles Stichprobenverfahren. *Das Geschmacksempfinden* [...] ist aber wegen des geringen Einkommens [Unterstreichungen von den Verfassern] von Studenten keineswegs repräsentativ für die mögliche Gesamtkäuferschaft.“ (Rogge 1981, S. 168)

Zusammenfassend lässt sich sagen, dass im allgemeinen Sprachgebrauch Repräsentativität<sup>4</sup> einer Teilgesamtheit dann vorliegt, wenn sie in bestimmten Merkmalen eine ähnliche Struktur aufweist wie die Grundgesamtheit. Daraus wird gefolgert, dass man dann - und für viele Autoren (z.B. Zentes 1996, S. 383) nur dann - von der Teil- auf die Grundgesamtheit schließen kann.

Die vorliegende Arbeit vertritt demgegenüber die bei Statistikern im allgemeinen vorherrschende Meinung,

1. dass „Repräsentativität“ ein aus einer im Falle von Stichprobenuntersuchungen falschen Vorstellung heraus entstandenes Konzept ist, und dass
2. Repräsentativität kein Qualitätsmerkmal für eine Untersuchung ist sondern in einer widersprüchlichen Beziehung zu den theoretisch fundierten statistischen Qualitätskriterien steht.

Daraus lässt sich dann ersehen, dass die Tatsache, dass man die Behandlung von Repräsentativität in der Literatur zur Induktiven Statistik oder zur Stichprobentheorie in der Regel vergeblich sucht, keine Ignoranz der Statistiker gegenüber den Problemen der Anwender ist, sondern nur Ausdruck dafür, dass von Seiten der Anwender ein Problem konstruiert wurde, das in Wirklichkeit keines ist.

## 2. Repräsentativität vs. Stichprobentheorie

An dieser Stelle scheint es sinnvoll zu sein, zunächst einige Begriffe zu definieren:

1. Unter einer *Grundgesamtheit* versteht man eine sachlich, zeitlich und örtlich abgegrenzte Menge von Merkmalsträgern, wobei die Abgrenzung in der Praxis oftmals relativ schwer fällt. Beispielsweise könnte man bei einer Umfrage nach dem

<sup>3</sup> Ähnlich drückt sich auch Koch 1997, S. 30 aus.

<sup>4</sup> In diesem Zusammenhang sei ausdrücklich erwähnt, dass wir Repräsentativität nicht für einen statistischen Fachausdruck halten, so dass nicht einfach von einer sprachlichen Diskrepanz zwischen Fach- und Alltagssprache ausgegangen werden kann.

Internetverhalten als Grundgesamtheit sämtliche Personen mit Internetanschluss, alle Computerbesitzer oder sogar die gesamte Bevölkerung wählen. In den letzten beiden Fällen wären die Antworten „Ich besitze kein Internet“ oder „Ich besitze keinen Computer“ auch eine Form von Internetverhalten.<sup>5</sup>

2. Eine *Teilgesamtheit* ist eine Teilmenge der Grundgesamtheit, wobei man nach der Art des Auswahlverfahrens unterscheidet:
  - a. Willkürliche Auswahl
  - b. Quotenauswahl (entspricht der „repräsentativen“ Teilgesamtheit)<sup>6</sup>
  - c. bewusste Auswahl
  - d. Zufallsauswahl (= Stichprobe)
  - e. sonstige

Das entscheidende ist nun, dass eine Stichprobe im statistisch-mathematischen Sinne *immer* eine Zufallsauswahl ist. Eine Quotenauswahl, also eine Teilgesamtheit, die das Prinzip der „Repräsentativität“ umsetzt, ist somit *keine* Stichprobe.<sup>7</sup> Eine Stichprobe ist durch die Art des Zustandekommens der Auswahl gekennzeichnet und auf sie ist die Wahrscheinlichkeitsrechnung anwendbar. Eine Quotenauswahl ist durch das Ergebnis der Auswahl (ähnliche Struktur) gekennzeichnet. Auf sie ist die Wahrscheinlichkeitsrechnung nicht anwendbar.

Was macht es für einen Unterschied, ob eine Auswahl zufällig oder nicht-zufällig gezogen wurde?

Auf rein deskriptiver Basis besteht natürlich kein Unterschied. Auf beide Gesamtheiten sind die Instrumente der Deskriptiven Statistik gleichermaßen anwendbar. Der Unterschied besteht in der Verallgemeinerung, also in den Rückschlüssen, die aus der Stichprobe auf die Grundgesamtheit gezogen werden können. Das besondere an einer Zufallsauswahl ist, dass die Stichprobenvariable eine Zufallsvariable ist. Damit besitzt sie eine Wahrscheinlichkeitsverteilung und das gesamte Instrumentarium der inferentiellen Statistik (Verwendung von Punkt- und Intervallschätzern unter Angabe eines Stichprobenfehlers sowie von Tests aller Art) ist anwendbar. Bei einer nicht-zufälligen Auswahl kann eine Verallgemeinerung auf die Grundgesamtheit nur verbal erfolgen.

Dabei ist zu bemerken, dass eine rein deskriptive Datenanalyse nichts Ehrenrühriges ist. Es spricht außerdem nichts dagegen, dass auch rein verbal sehr vernünftige und wohlbegründete Schlussfolgerungen gezogen werden können. Allerdings ist immer wieder zu beobachten, dass es für viele Autoren offensichtlich wichtig ist, dass ihrer Untersuchung nicht nur das vermeintliche „Gütesiegel“ repräsentativ anhaftet, sondern dass darüber hinaus auch die Aussage „Verwendung der Inferenzstatistik“ ein Qualitätsmerkmal ist. Dabei wird meist nicht gesehen, dass beides auf sehr unterschiedlichen Ebenen steht.

Es muss noch einmal deutlich betont werden, dass eine Stichprobe, also eine auf Zufall basierende Auswahl, nicht repräsentativ im Sinne einer der Grundgesamtheit

---

<sup>5</sup> Vgl. zu dieser Problematik auch *Lander 1998, Will 1998 und Bogner/Mayer 2000*.

<sup>6</sup> Interessanterweise bezeichnet Malhotra (2000) gerade die Quotenauswahl als nicht repräsentativ. Auf eine Definition, was Repräsentativität dann bedeutet, wird allerdings verzichtet.

<sup>7</sup> Dabei ist sehr genau zu prüfen, ob eine Auswahl zufällig ist oder nicht. Z.B. ist eine Teilgesamtheit, die durch einen hohen Prozentsatz an Antwortausfällen zustande kommt, im Gegensatz zur Meinung vieler Autoren keine Zufallsauswahl, sondern hängt von verschiedenen, in der Regel aber unbekanntem Einflussfaktoren ab.

gleichen Struktur sein muss. Sie *kann* es sein, ist es dann aber eben nur zufällig. Man kann aber auch einfach das „Pech“ haben, eine äußerst „ungünstige“ Stichprobe zu ziehen, die zu Schätzungen von Grundgesamtheitseigenschaften führt, die von den „wahren“ Eigenschaften weit entfernt liegen. Der Vorteil einer echten Zufallsstichprobe besteht nun darin, dass diese „Gefahr“ mittels des Stichprobenfehlers zum einen quantifizierbar und zum anderen – was vielleicht noch wichtiger ist – steuerbar ist. Die Steuerung erfolgt über den Stichprobenumfang. Wir werden darauf noch näher eingehen.

Darüber hinaus versuchen einige Autoren die „Repräsentativität“ ihrer „Stichprobe“ mit Hilfe spezieller Tests zu überprüfen. Dabei handelt es sich in der Regel um den Chi-Quadrat-Test zur Überprüfung der Anpassung einer empirischen an eine theoretische Verteilung.<sup>8</sup> Selbst wenn wir annehmen, dass die vorliegende Teilgesamtheit tatsächlich eine Stichprobe ist und damit die Grundvoraussetzung für die Verwendung eines Tests überhaupt geschaffen ist, sind diese Tests von höchst zweifelhaftem Wert. Der Hauptkritikpunkt liegt darin, dass sie zu einer Umkehrung der Testlogik in ihr genaues Gegenteil führen. Statt wie üblich eine Hypothese über die unbekannte Grundgesamtheit an einer bekannten Stichprobe zu testen, wird hier eine Stichprobe an einer Grundgesamtheit getestet. Die Entscheidung lautet somit nicht auf Ablehnung oder Annahme der Hypothese sondern auf Ablehnung oder Annahme der Stichprobe.

### 3. Kritik am Konzept Repräsentativität

Es ist nicht unser Ziel, die Vor- und Nachteile von zwei Auswahlverfahren gegenüberzustellen, nämlich einerseits der echten Zufallsstichprobe und andererseits des Quotenverfahrens, das auf dem beschriebenen Konzept der "Repräsentativität" beruht und von der folgenden Erwartung ausgeht: „In dem verkleinerten Modell [...] sind auch die untersuchenden Ansichten und Antwort-Kategorien in modellgerechtem Anteil vertreten" (Noelle 1963, S. 110)<sup>9</sup>. Unser Ziel ist nur zu zeigen, dass Repräsentativität zwar intuitiv überzeugend und daher als Konzept wohl unverwundlich ist, es sich aber auf den zweiten Blick herausstellt, dass

1. neben Repräsentativität auch andere nicht weniger plausible Kriterien an die Güte einer Auswahl herangetragen werden können, und diese es nicht selten verlangen, Repräsentativität bewusst nicht zu realisieren,
2. sich das Konzept Repräsentativität, anders als das des Stichprobenfehlers, nicht aus einer Theorie über ein Auswahlverfahren entwickeln lässt (wie z.B. der Stichprobentheorie als eine Theorie der Zufallsauswahl) und sich aus Repräsentativität auch keine mathematisch formulierten Folgerungen ergeben,
3. Repräsentativität in einem ungeklärten, wenn nicht widersprüchlichen Zusammenhang zu anderen Variablen steht, die ohne Zweifel entscheidend sind für die Güte einer Auswahl (z.B. der Umfang  $n$  der Teilgesamtheit oder Stichprobe und die Varianz des Untersuchungsmerkmals in der Grundgesamtheit (GG) als Ausdruck der Homogenität der GG).

---

<sup>8</sup> Eine Alternative zu dieser üblichen Vorgehensweise in Form eines verteilungsfreien Tests bietet Polasek 1997 an.

<sup>9</sup> Eine der kritischen Fragen an dieses Verfahren ist: Woher nimmt man die Gewissheit, dass man alle relevanten Quotenmerkmale berücksichtigt hat, und dass dieser Schluss gerechtfertigt ist, dass also die Gesamtheit nach Belieben maßstäblich verkleinert (wie weit verkleinert?) werden kann?

In den folgenden Abschnitten 4 und 5 stellen wir die ersten beiden Einwände gegen die Repräsentativität dar und in den Abschnitten 6 und 7 den dritten Einwand. Dabei ergeben sich Fragen, deren Beantwortung wir zumindest für problematisch halten.

Wir wollen dazu annehmen, eine Grundgesamtheit (GG) bestehe zu je 50% aus Männern und Frauen. Das Merkmal Geschlecht (das nicht unser Untersuchungsmerkmal ist) soll als Quotenmerkmal dienen, d.h. anhand dieses Merkmals machen wir den „Grad der Repräsentativität“ fest. Aus dieser Grundgesamtheit seien Teilgesamtheiten (TG) gezogen worden. Außerdem nehmen wir an, es existiere eine Maßzahl, anhand derer wir die Repräsentativität der  $i$ -ten TG (Maßzahl  $R_i$ ) messen können.

#### 4. Gewollte Nichtrepräsentativität

Wir haben bisher die These vertreten, dass Repräsentativität kein Gütekriterium für eine Untersuchung ist. Man sollte sich an dieser Stelle bewusst machen, dass es sogar genügend Situationen gibt, in denen eine bewusst nichtrepräsentative Auswahl vorzuziehen ist. Wir führen hierfür zwei Beispiele an:

1. die nichtproportionale Aufteilung im Rahmen einer geschichteten Stichprobe. So kann es z.B. sinnvoll sein, nach Bundesländern zu schichten und eine Überrepräsentation kleinerer Länder anzustreben<sup>10</sup>,
2. das Konzentrationsprinzip als ein Beispiel für eine *nichtzufällige* Auswahl, die bewusst nicht repräsentativ (zumindest hinsichtlich der Größe der Einheiten) ist.

Es sind daher folgende Überlegungen anzustellen:

1. Sieht man davon ab, dass innerhalb der *Schichten zufällig* ausgewählt wird, so besteht eine gewisse Verwandtschaft zwischen der geschichteten Stichprobe und der Quotenauswahl, die ja die Logik der "Repräsentativität" verfolgt. Dass man – ganz im Sinne der "Repräsentativität" - die gleichen Proportionen  $n_1 : n_2 : \dots : n_K$  in der Stichprobe (oder TG) hat wie in der GG (mit  $N_1 : N_2 : \dots : N_K$ ) ist nur eine von unendlich vielen Möglichkeiten der geschichteten Stichprobe. Wenn die Varianzen innerhalb der Schichten ( $\sigma_k^2$ ,  $k = 1, 2, \dots, K$ ) unterschiedlich sind, dann ist es sinnvoll die (geschichtete) Stichprobe *nicht*-proportional aufzuteilen. Angenommen, die GG bestehe aus  $N_1 = 30$  Mill. Einpersonenhaushalten und  $N_2 = 50$  Mill. Mehrpersonenhaushalten. Wenn man  $n = 80$  Haushalte befragen wollte, dann erhält man eine „repräsentative“ Auswahl, wenn es  $n_1 = 30$  Einpersonens- und  $n_2 = 50$  Mehrpersonenhaushalte sind (sog. "proportionale Aufteilung"). Aber aus der Theorie der "optimalen" (nichtproportionalen) Aufteilung<sup>11</sup> einer geschichteten Stichprobe ist bekannt, dass es sehr viel besser sein kann, etwa nur 20 Einpersonenhaushalte und dafür aber 60 Mehrpersonenhaushalte zu befragen (also eine bewusst *nicht* "repräsentative" Auswahl vorzunehmen), weil z.B. die Standardabweichung  $\sigma_M$  innerhalb der Mehrpersonenhaushalte viel größer ist als die Standardabweichung innerhalb der Einpersonenhaushalte  $\sigma_E$ .
2. Ein extremer Fall von Nicht-Proportionalität und bewusster Verletzung der "Repräsentativität" ist das in der amtlichen Statistik und auch in der Marktforschung

<sup>10</sup> Das geschieht sehr oft in der amtlichen Statistik, um zu einer möglichst gleichen Genauigkeit und Sicherheit der Aussagen für alle Bundesländer zu gelangen.

<sup>11</sup> Hier geht es um die Minimierung der Gesamtvarianz, bei der oben beschriebenen Praxis der amtlichen Statistik dagegen um vergleichbar große Fehlervarianzen für alle Schichten (z.B. Bundesländer). In beiden Fällen wird von einer proportionalen („repräsentativen“) Aufteilung abgewichen.

sehr beliebte "cut-off-principle" oder "Konzentrationsprinzip" (Abschneideverfahren): Man befragt nur die Unternehmen mit 20 und mehr Beschäftigten, womit dann aber z.B. 85% des Umsatzes einer Branche abgedeckt wird. Die große Zahl der kleineren Unternehmen, die aber nur 15% des Umsatzes auf sich vereinigen, werden nicht befragt. Die Frage ist dann natürlich: Sollte man solche Erhebungen ablehnen, weil sie ja dem Prinzip der "Repräsentativität" (zumindest was das Merkmal Unternehmensgröße betrifft) ganz bewusst widersprechen?

## 5. In welchem theoretischen Zusammenhang steht der Begriff "Repräsentativität"?

Es ist zuzugeben, dass die Wahrscheinlichkeitsrechnung nicht leicht zugänglich ist. Nichtstatistiker haben oft Schwierigkeiten, "zufällige" und "willkürliche" Auswahl zu unterscheiden, sie haben auch – wie beschrieben – wenig Skrupel, dort statistische Tests anzuwenden, wo überhaupt keine Zufallsauswahl vorliegt und neigen meist sehr zum Konzept der Repräsentativität. Worauf sind die Beliebtheit der Repräsentativität und der Quotenauswahl einerseits und die Verständnisprobleme bei der Stichprobentheorie andererseits zurückzuführen? Wir glauben mit *Wendt* 1960, S. 35: "Das muss doch – zumindest auch – daran liegen, dass die theoretische Seite dieser Angelegenheit nicht überzeugend im Gedächtnis haften bleibt". Es bleibt eben nicht so leicht haften, dass unterschiedliche Stichproben theoretisch alle gleich gut sind.

Was die "theoretische Seite" betrifft, so ist der Begriff Repräsentativität zwar leichter zugänglich als das Konzept Stichprobenfehler<sup>12</sup> etwa des Mittelwerts  $\bar{x}$  oder des Anteilswerts  $p$ , also  $\sigma_{\bar{x}}$  bzw.  $\sigma_p$ , aber was leichter eingeht, muss nicht auch mehr Einsichten vermitteln. Im Gegenteil folgt aus „Repräsentativität“ wenig oder nichts,

wenn man darunter mathematisch exakte Aussagen versteht. Aus  $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

folgt, dass der Fehler geringer wird, wenn der Stichprobenumfang  $n$  größer wird. Mehr Menschen zu befragen bedeutet bei einer echten Stichprobe ein kleinerer Fehler  $\sigma_p$  und damit ein kleineres Konfidenzintervall. Der Umfang  $n$  erscheint also in einer Formel. Aber in welcher Formel erscheint die Repräsentativität und was bedeutet mehr Repräsentativität für die Genauigkeit und Sicherheit der Aussage?

Was kann man überhaupt über die Verallgemeinerungsfähigkeit  $V$  auf der Basis einer wie auch immer definierten Repräsentativitätsmaßzahl  $R$  aussagen? Es mag sicher *plausibel* (aber wohl auch nicht mehr als das) sein, anzunehmen, dass man z.B. das Durchschnittseinkommen der GG "besser" (was immer das heißen mag) schätzen kann, wenn man eine TG von 50 Frauen und 50 Männern hat, als im Falle von 80 Frauen und 20 Männern. Beweisen kann man das aber wohl im üblichen Fall begrenzter Kenntnisse über die GG nicht. Kann man ein Maß dafür angeben, *um wie viel* größer  $V$  ist (z. B. nach Art der geringeren Breite eines Konfidenzintervalls) im Fall von  $R_1$  (50:50) im Unterschied zu  $R_2$  (80:20), und wenn ja aufgrund welchen mathematischen Modells? Um *wie viel* "besser" ist denn der Fall  $R_1$  als der Fall  $R_2$  wirklich? Wie groß ist  $V_i$  bzw.  $V_j$  wenn  $R_i$  z.B. 7 und  $R_j = 5$  ist! Welche Formel gilt hier? Aus welcher Theorie über  $R$  wird sie abgeleitet?

Zu der fehlenden Theorie gehört auch, dass das Konzept Repräsentativität völlig unabhängig davon ist, wie die betrachtete Gesamtheit zustande kommt, während die

<sup>12</sup> Standardabweichung der Stichprobenverteilung.

Stichprobentheorie doch gerade eine Theorie über das Zustandekommen einer Untersuchungsgesamtheit darstellt. Es gibt das Beispiel, dass eine willkürliche Auswahl aus einer Kartei eines psychologischen Instituts oder einer Klinik als "repräsentativ" betrachtet wurde, nur weil die entsprechenden Proportionen mit denen der Grundgesamtheit übereinstimmen, obgleich doch völlig klar ist, dass es schon etwas zu bedeuten hat, dass man überhaupt in diese Kartei gelangt. Sind 50 Frauen und 50 Männer aus einer Patientenkartei genau so "repräsentativ" wie 50 Frauen und 50 Männer aus dem Telefonbuch oder aus dem Mitgliederverzeichnis eines Vereins usw.?

## 6. Repräsentativität und Anzahl der Befragten

Wie bereits erwähnt soll hier das Verhältnis des Begriffs Repräsentativität zu anderen Begriffen wie dem Stichprobenumfang  $n$  in Form von Fragen (insgesamt sechs Fragen) problematisiert werden.

Wie wir gerade festgestellt haben, gilt in der Stichprobentheorie die einfache Regel: Je größer  $n$ , desto kleiner der Stichprobenfehler. Welche Rolle spielt  $n$  aber bei der Repräsentativitätsbetrachtung? Oder bezüglich unseres Beispiels (50% Männer, 50% Frauen) gefragt:

1. Kann man sagen, dass es um ... % besser ist 50 Männer und 50 Frauen zu befragen als 485 Männer und 502 Frauen? Mit welcher Formel kommt man hier zu einem konkreten Prozentsatz der Verbesserung (von z. B. 12,3%)?
2. Die Repräsentativität (sie sei  $R_1$ ) müsste bei 50 Männern und 50 Frauen und bei 500 Männern und 500 Frauen ( $R_2 = R_1$ ) gleich sein. Trotzdem ist zu vermuten, dass wir im zweiten Fall meist die "besseren" Daten haben dürften (obgleich  $R_1 = R_2$ ). Also: Wie differenziert man, wenn  $R_1 = R_2$  ist, es aber hinsichtlich des Umfangs  $n$  der zu vergleichenden Teilgesamtheiten erhebliche Unterschiede gibt? Wären die TGen echte (Zufalls-) Stichproben, dann wäre die Standardabweichung der Schätzung des Anteils  $\pi$  der Männer durch den Stichprobenanteil  $p$ , also  $\sqrt{\frac{p(1-p)}{n-1}}$  im ersten Fall (bei  $p = \frac{1}{2}$ ,  $n = 100$ ) 0,05025, im zweiten Fall ( $p = \frac{1}{2}$ ,  $n = 1000$ ) aber nur 0,01582 (also 1,582% statt 5,025%). Folglich würde ein 95% Konfidenzintervall ( $z = 1,96$ ) im ersten Fall von 0,4015 bis 0,5985 gehen und im zweiten Fall sehr viel enger sein, nämlich von 0,468996 bis 0,531006. Bei gleicher Sicherheit erhält man eine wesentlich genauere Aussage. Das bedeutet doch, dass wir im zweiten Fall "bessere" Daten haben. Das Ergebnis ist auch intuitiv verständlich, zumal der Stichprobenumfang zehnmal so groß ist. Gemessen an der "Repräsentativität" hätten wir aber durch Verzehnfachung unserer Teilgesamtheit überhaupt nichts gewonnen, denn schließlich ist  $R_1 = R_2$ , egal ob die Relation 50:50 oder 500:500 ist. Ist dieses Ergebnis sinnvoll?
3. Aber auch bei einer Verkleinerung des Umfangs unserer Beobachtungen könnte  $R$  gleich bleiben. Die Frage ist nur "wie lange"? Wie wäre es bei  $n = 3$ ? Hier wäre *keine einzige TG* "repräsentativ", denn schließlich kann der Anteil der Männer (und entsprechend auch der Frauen) nur 0,  $\frac{1}{3}$ ,  $\frac{2}{3}$  oder 1 sein, aber nicht  $\frac{1}{2}$ . Wie wäre es mit  $n = 1$ ?<sup>13</sup> Die Logik des "getreuen Abbilds" kann doch nicht bei

<sup>13</sup> Es stellt sich sowieso die Frage, wieso in der Marktforschung immer nach einer „Struktur“ gefragt wird und nicht nach dem „typischen Element“ (wie es in anderen Wissenschaften, wie z.B. der Biologie durchaus üblich ist). Wieso sucht man nicht einfach nach dem „Durchschnittsbürger“ mit dem „Durchschnittsgeschmack“ und befragt nur den?

einer nicht näher bestimmten Zahl  $n$  halt machen (etwa dergestalt, dass sie noch bei  $n = 12$  gilt, aber nicht mehr bei  $n = 11$ )<sup>14</sup>.

4. Interessant ist in diesem Zusammenhang auch die Frage der Verbesserung oder Verschlechterung einer TG durch Hinzu- oder Wegnahme eines oder mehrerer Elemente. Angenommen, man habe aus Versehen 51 Frauen und 50 Männer befragt, also eine Frau zu viel. Nach der allgemeinen Vorstellung von Repräsentativität müsste sich nun der Datensatz dadurch verbessern lassen, dass man die überzählige Frau herausnimmt, wobei es sogar ganz egal ist, welche der 51 Frauen dies ist. Tritt in *jedem* Fall eine "Verbesserung" durch Entfernen einer Einheit (egal welcher Frau) ein? Das müsste eigentlich so sein, denn schließlich vergrößert sich ja  $R$  dadurch. Die nächste Frage wäre natürlich: Wie groß ist die Verbesserung?

## 7. Repräsentativität und Homogenität der Grundgesamtheit sowie Korrelation zwischen Quoten- und Untersuchungsmerkmal

Völlig außer acht gelassen wird beim Konzept „Repräsentativität“ auch die Homogenität (gemessen an der Varianz  $\sigma^2$ ) der GG und die Forderung an das Quotenmerkmal.

5. Angenommen, es gilt die Körpergröße  $Y$  zu schätzen und in der GG sind alle gleich groß ( $\sigma_Y^2 = 0$ ). Es genügt dann, wie unmittelbar einleuchtet, eine einzige Person zu befragen, und zwar egal welche. Für eine repräsentative Stichprobe kann man aber nur mindestens  $n = 2$  Personen befragen und zwar einen Mann und eine Frau (auch nicht etwa zwei Frauen). Soll man auch in diesem Falle ( $\sigma_Y^2 = 0$ ) lieber eine Person zu viel befragen oder eine TG von zwei Männern oder zwei Frauen statt einem Mann und einer Frau ablehnen?

Es mag sein, dass der Fall  $\sigma^2 = 0$  unrealistisch ist. Es geht uns nur darum zu zeigen, dass die Größe von  $\sigma^2$  eine Rolle spielt,  $\sigma^2$  (anders als  $R$ ) deshalb mit Recht in  $\sigma_x^2 = \sigma^2/n$  vorkommt und dass es auch völlig dem gesunden Menschenverstand entspricht, dass die TG bei größerem  $\sigma^2$  mehr Einheiten und bei kleinerem  $\sigma^2$  weniger Einheiten umfassen sollte (im Extremfall  $\sigma^2 = 0$  ist es  $n = 1$  Einheit)<sup>15</sup>. Bei der Betrachtung einer „repräsentativen“ TG ist es aber offenbar ganz egal, wie groß  $\sigma^2$  (und auch  $n$ ) ist, Hauptsache man hat ein bestimmtes  $R$ . Bei großem  $\sigma^2$  des Untersuchungsmerkmals kann die TG mit 485 Männern und 502 Frauen *erheblich* besser sein als die TG mit 50 Männer und 50 Frauen. Unter Repräsentanzgesichtspunkten ist aber die Situation wegen der geringeren  $R$  im ersten Fall *stets* schlechter als im zweiten, egal wie groß  $\sigma^2$  ist. Das ist doch wohl schwer zu rechtfertigen. Offensichtlich spielt  $\sigma^2$  eine Rolle für die "Verallgemeinerungsfähigkeit" ( $V$ ) einer TG oder einer Stichprobe. Gibt es aber eine

<sup>14</sup> Tiede 1996 versucht den Gedanken der "Repräsentativität" anhand des Grenzfalles von  $n = 1$  ad absurdum zu führen.

<sup>15</sup> Dieser Zusammenhang wird z.B. völlig übersehen, wenn manche Autoren argumentieren, der Stichprobenumfang müsse eine „allgemein akzeptierte“ Größe haben und die „heute übliche Norm“ seien „1000 oder 2000 Fälle“ (Wyss 1991, S. 498/499). Die Zahlen 1000 und 2000 sind offensichtlich ganz gefühlsmäßig aus der Luft gegriffen. Bei geringer Varianz und geringerer Genauigkeitsanforderung kann 1000 zu viel und entsprechend kann bei größerer Varianz und größeren Anforderungen an die Genauigkeit 2000 zu wenig sein.

Formel, mit der man von R zusammen mit einem bestimmten  $\sigma^2$  und n auf V schließen kann?

6. Wenn es vor allem auf die Struktur der TG und GG ankommt, fragt man sich natürlich auf *welches* Merkmal (oder auf welche Merkmale) sich die "Struktur" beziehen soll (das ist die Frage nach dem bzw. den Quotenmerkmal(en)) und wie detailliert sie gemessen werden soll? Ein Quotenmerkmal X sollte, so heißt es, mit dem Untersuchungsmerkmal Y korrelieren. Wie sieht es aber aus, wenn die TG sehr "repräsentativ" ist hinsichtlich des Merkmals X aber weniger beim Merkmal Z, aber andererseits die Korrelation  $r_{zy}$  wesentlich größer ist als die Korrelation  $r_{xy}$ ? Treten diese Korrelationen überhaupt in einer *zahlenmäßigen* Betrachtungen auf oder werden sie allenfalls verbal bedacht? Man kann ja wohl schlecht R bezüglich *irgendeines* Merkmals fordern. Es gibt offenbar unterschiedlich geeignete Merkmale (wohl zu messen an der Korrelation r), wenn es gilt, etwas über V aussagen zu können. Wo ist die Formel für V, in der neben R auch r und (siehe oben)  $\sigma^2$  und n auftreten?

## 8. Fazit

Es ist zu bezweifeln, dass das Konzept "Repräsentativität" im Unterschied zum Auswahlfehler überhaupt ein sinnvolles Kriterium für die Güte einer Auswahl ist. Es gibt *erstens* zu viele Unklarheiten, wenn nicht gar Widersprüche, wenn man – wie hier versucht – einmal einige Aussagen über R und deren Konsequenzen mit den auf dem ersten Blick offenbar schwer zugänglichen, auf dem zweiten Blick aber durchaus plausiblen Erkenntnissen aus der Stichprobentheorie vergleicht und *zweitens* vermisst man bei der "Repräsentativität" eine Theorie oder ein mathematisches Modell, mit dem man zeigen kann, was aus einem größeren oder kleineren R folgt. Eine solche Theorie ist uns nicht bekannt und deshalb halten wir „Repräsentativität“ nicht für ein brauchbares Konzept.

Was ist nun für empirische Untersuchungen aus diesem Fazit zu folgern? Zunächst ist streng zwischen zufälligen und nicht-zufälligen Teilgesamtheiten und den damit verbundenen Auswertungsmöglichkeiten zu trennen. Wir sehen dabei sehr wohl das Problem, dass in vielen Fällen eine „echte“ Zufallsstichprobe nicht zu erreichen ist. Dann muss auf die Verwendung wahrscheinlichkeitstheoretisch fundierter Methoden wie der Schätz- und Testtheorie verzichtet werden. Dies ist nicht unbedingt als Nachteil anzusehen, da selbstverständlich auch die Methoden der Deskriptiven Statistik zu einer Vielzahl von relevanten Aussagen führen können. Im Falle einer „echten“ Zufallsauswahl ist Repräsentativität schon von sich aus kein Gütekriterium, weil das zufällige Ziehen einer repräsentativen Auswahl in dem Sinne, dass Strukturgleichheit zwischen der Grund- und der Teilgesamtheit besteht, als ziemlich unwahrscheinlich anzusehen ist und daher fast alle Zufallsstichproben als „nicht repräsentativ“ zu bezeichnen wären. Hier sollte zur Beurteilung der „Qualität“ einer Schätzung auf das Konzept des Stichprobenfehlers zurückgegriffen werden bzw. sollten Formen der Stichprobenziehung verwendet werden, die eine Verkleinerung dieses Fehlers bewirken (z.B. geschichtete Stichproben).

## Literatur

Bausch, T. (1995), Auswahlverfahren in der Marktforschung, in: Tietz, B/R. Köhler/J. Zentes (Hrsg.), Handwörterbuch des Marketing, Stuttgart, Sp. 155 - 166

- Berekoven, L./W. Eckert/P. Ellenrieder* (1999), Marktforschung: Methodische Grundlagen und praktische Anwendung, 8. Auflage, Wiesbaden
- Bogner, W./M. Mayer* (2000), Die Validität von Onlinebefragungen II – Land in Sicht?, in: Planung und Analyse 1/2000, S. 50 -55
- Herrmann, A./C. Homburg* (1999), Marktforschung: Methoden, Anwendungen, Praxisbeispiele, Wiesbaden
- Karsch, W.* (2000), Aktienkultur: Volk der Aktionäre?, in: Die Bank 10/2000, S. 710 – 711
- Keller, B./S. Lerch/S. Matzke* (2000), Umfrage: Kundenbindung und Wechselbereitschaft, in: Die Bank 6/2000, S. 376 - 381
- Koch, J.* (1997), Marktforschung: Begriffe und Methoden, 2. Auflage, München/Wien
- Lander, B.* (1998), Güte von Internet-Umfragen, in: Planung und Analyse 5/98, S. 63 – 66
- Malhotra, N.* (2000), Marketing Research, Harlow u.a.
- Noelle, E.* (1963), Umfragen in der Massengesellschaft: Einführung in die Methoden der Demoskopie, Reinbek bei Hamburg
- Pekruhl, U./J. Nordhause-Janzen* (2000), Gruppenarbeit: Konzept und Realität: Ergebnisse einer repräsentativen Beschäftigtenbefragung, in: Personal – Zeitschrift für Human Resource Management 7/2000, S. 326 - 331
- Pepels, W.* (1998), Auswahlverfahren in der quantitativen Marktforschung, in: Planung und Analyse 1/98, S. 47 - 51
- Polasek, W.* (1997), Schließende Statistik, Springer, Berlin/Heidelberg/New York
- Rogge, H.-J.* (1981), Marktforschung: Elemente und Methoden betrieblicher Informationsgewinnung, Hanser, München/Wien
- Steffens-Duch, S.* (2000), Commitment – die Bank im Urteil der Mitarbeiter, in: Die Bank 3/2000, S. 183 - 185
- Tiede, M.* (1996), Repräsentativität, Diskussionspapier aus der Fakultät für Sozialwissenschaft der Ruhr-Universität Bochum 96 – 12, Bochum
- Tiede, M./W. Voß* (2000), Schließen mit Statistik – Verstehen, München/Wien
- Wendt, F.* (1960) Wann wird das Quotenverfahren begraben?, in: Allgemeines Statistisches Archiv 44, S. 35 – 40
- Werani, T.* (2000), Der Wert von kooperativen Geschäftsbeziehungen in industriellen Märkten, in: Die Unternehmung 2/2000, S. 123 – 143
- Will, C.* (1998), Augen zu und durch: Überlegungen zu den Fußangeln bei Online-Befragungen, in: Planung und Analyse 1/1998, S. 20 - 24
- Wyss, W.* (1991), Marktforschung von A – Z, Demoscope, Adligenwil
- Zentes, J.* (1996), Grundbegriffe des Marketing, 4. Auflage, Stuttgart