

Kurzdarstellung des Rösch-Telefonstichproben-Systems (RTS)

1. Methodische Vorbemerkungen

Das generelle Problem bei der Bildung von Stichproben für telefonische Umfragen in der Bundesrepublik Deutschland besteht darin, dass heute - nach Ansicht aller Experten - weit mehr als 20% der Telefonhaushalte nicht in den einschlägigen Verzeichnissen eingetragen sind. Es handelt sich bei diesen Haushalten in erster Linie um solche mit Neuanschlüssen aus den verschiedensten Gründen (Umzug, Wechsel auf ISDN, gerade in den Neuen Bundesländern in großem Umfang auch viele Erstanschlüsse nach der Wiedervereinigung, etc.)

Siegfried Gabler und Sabine Häder (ZUMA Mannheim) haben als Erste die herkömmlichen Verfahren für Telefonstichproben in Deutschland wissenschaftlich analysiert und nachgewiesen, dass bei diesen Stichproben die Auswahlwahrscheinlichkeit der Telefonhaushalte

- nicht gleich,
- nicht bekannt und
- auch nicht berechenbar sind.

Mit anderen Worten: diese Stichproben - Random Digit Dialing (RDD), Random Last Digit (RLD), etc. - sind statistisch verzerrt und diese Verzerrungen sind weder vernachlässigbar noch korrigierbar. (Näheres hierzu in den Anmerkungen am Ende dieses Textes.)

Sie haben daraufhin (1997) ein zu seiner Zeit bahnbrechendes Modell entwickelt, das auf sogenannten Hunderterblocks basiert. Ein valider Hunderterblock ist definiert durch eine eingetragene Telefonnummer eines Privatanschlusses und Variation der letzten beiden Ziffern (etwa der Privatanschluss 08141-94965 erzeugt den Hunderterblock 08141-949xx, d.h. alle 100 Nummern von 08141-94900 bis 08141-94999). Wenn man nun alle Hunderterblocks mit ihrer jeweiligen Eintragsdichte (Anzahl der eingetragenen Anschlüsse pro Hunderterblock) und die Blockdichte (Anzahl valider Nummern pro Hunderterblock) ermittelt oder hinreichend genau (erwartungstreu) schätzt, so kann man daraus – entsprechendes Fachwissen vorausgesetzt – Stichproben erstellen, bei denen die Auswahlwahrscheinlichkeiten der Stichprobeneinheiten gleich oder zumindest berechenbar sind.

Leider hat dieses Verfahren in der ursprünglichen Form eine Reihe von spezifischen Nachteilen:

Die betreffenden Stichproben sind nach den Telefonanschlüssen 'repräsentativ', verteilt - diese sind aber von untergeordnetem Interesse, denn in der Regel bilden '**Privathaushalte**' bei den Bevölkerungsumfragen die Grund- bzw. Auswahlgesamtheit der sekundären Auswahlstufe.

- Mehrfacheinträge von Privathaushalten werden dabei nicht berücksichtigt,
- Firmen- und Behördeneinträge werden nur sehr unvollkommen erkannt, etwa über sog. 'Bindestrich-Nummern', also Durchwahlnummern von Telefonanlagen.

- Die naturwüchsigen Aggregate der Einheiten bzw. Hunderterblocks sind Vorwahlnummern, über die keine hinreichenden sekundärstatistischen Daten oder Strukturen verfügbar sind. Die 'normalen' Gebietseinheiten, wie Gemeinden, Kreise, Regierungsbezirke oder Bundesländer sind für dieses System inkompatibel. Die kleinräumige Variation dieser Stichproben ist zudem sehr hoch.
- Der Ausgleich von Non-Response-Fehlern oder differenzierte Redressementgewichtungen sind somit nur eingeschränkt möglich.

Damit ist das "reine" System für Umfragen in der Markt-, Medien- und Sozialforschung in der Regel nicht optimal geeignet.

2. Das Rösch-Telefonstichproben-System (RTS)

2.1 Auswahlgrundlage

Für das Rösch-Telefonstichproben-System wurde der grundlegende Ansatz von Gabler und Häder konsequent umgesetzt und um einige wesentliche Aspekte erweitert.

Die Basis des RTS bildet eine Datenbank von allen eingetragenen Telefonanschlüssen, die aus den verschiedensten verfügbaren und kritisch überprüften Quellen¹ ständig aktualisiert wird.

Die zugehörige Adressenbasis wird in einem komplexen Verfahren zunächst auf nicht-private Anschlüsse geprüft (Abgleich mit Branchenverzeichnissen, spezielle und selbstlernende Erkennungssoftware, etc.) und die betreffenden Anschlüsse werden - genauso wie z.B. erkannte Telefaxanschlüsse - entsprechend gekennzeichnet.

Danach werden Mehrfacheinträge einer Nummer (unter verschiedenen Namen) eliminiert und in mehreren unterschiedlichen Schritten die Zahl der Mehrfacheinträge eines Haushalts ermittelt und festgehalten.

Die enthaltenen Anschlüsse haben ja naturgemäß immer einen Namen und eine (vollständige) Telefonnummer, ansonsten aber durchaus unterschiedliche Informationsniveaus. Diese Informationsniveaus werden optimal ausgeschöpft, um die Anschlüsse mit verschiedenen regionalen Merkmalen wie z.B. Gemeindeschlüssel und Stadtteil (in größeren Städten) anzureichern.

¹ Der Aspekt der kritischen Überprüfung der Quellen wird in der Fachliteratur entweder übersehen oder völlig unterschätzt (was sich im Übrigen auch bei den Diskussionen in Workshops o.ä. immer wieder bestätigt). Es ist leider eine schmerzliche Tatsache, dass **alle** (!) Quellen mehr oder weniger – punktuell zum Teil sogar extrem - fehlerhaft sind.

Somit erhalten wir:

- voll qualifizierte Nummern eingetragener Anschlüsse (Gemeindeschlüssel, ggfs. Stadtteil, Postleitzahl, Ortsname, Adresse)
- teilweise qualifizierte Nummern eingetragener Nummern mit unterschiedlichem Niveau:
 - o Gemeindeschlüssel ohne Stadtteile
 - o Ortsnetz mit Wahrscheinlichkeiten für mehrere Gemeinden
 - o Ortsnetz ohne weitere Qualifizierung
- voll qualifizierte Nummern nichteingetragener Anschlüsse (s.o., ohne Adresse)
- teilweise qualifizierte nichteingetragener Anschlüsse (s.o.)

Die nichteingetragenen Nummern in den Hunderterblocks enthalten dabei naturgemäß einen gewissen Anteil nicht existierender Anschlüsse, die (bei korrekter Behandlung) als stichprobenneutrale Ausfälle übergangen werden können.

Mit dieser Auswahlgrundlage, die im Übrigen ständig aktualisiert und an neuere Sachverhalte, wie z.B. Gebiets- und Bevölkerungsstandsänderungen, angepasst wird, können in speziellen Stichprobenverfahren hinreichend erwartungstreue und kleinräumig hoch auflösende Stichproben für die Grundgesamtheit "Privathaushalte" (und nicht: "Telefonanschlüsse" oder "Telefonhaushalte") gezogen werden.

2.2 Stichproben

Das konkrete Stichprobenverfahren selbst hängt natürlich auch von den Zielsetzungen der jeweiligen Forschungsaufgabe ab und ist auf Grund der Gegebenheiten äußerst komplex und entsprechend kompliziert ("erweiterte Mikrostratifikation"). Es stellt aber sicher, dass jede Stichprobe grundsätzlich überschneidungsfrei zu allen vorangegangenen Stichproben des Systems ist.

Die Konzeption des vorgestellten Systems wurde 1998 entwickelt, der erste Einsatz erfolgte im April 1999. Seitdem wurde es ständig weiterentwickelt und optimiert (Update-Routinen, Konsistenzchecks, Laufzeitoptimierung, Erkennungssoftware, Gebietstandsaktualisierungen, Stichprobenverfahren, etc.) und hat sich in einer großen Zahl der unterschiedlichsten nationalen und regionalen Stichproben mit vielen Millionen Auswahleinheiten bestens bewährt.

2.3 Gewichtung

Telefonstichproben in der Bevölkerung sind häufig zunächst Haushaltsstichproben (meist in der zweiten Auswahlstufe), bei denen jeder Privathaushalt im Prinzip die gleiche Auswahlchance hat. Wenn in den Haushalten dann eine Zielperson zufällig ausgewählt wird, so sind die Auswahlchancen abhängig von der jeweiligen auf die Zielgruppe eingeschränkten Haushaltsgröße. Daher sind die re-

sultierenden Stichproben auf Personenebene statistisch verzerrt und müssen deshalb gewichtet werden, um diese design-gegebene Verzerrung auszugleichen - es sei denn, man beschränkt sich nur auf die Auswertung von haushaltsbezogenen Merkmalen. Den Ausgleich unterschiedlicher Auswahlwahrscheinlichkeiten nennt man "Design-Gewichtung", hier, in diesem speziellen Fall, handelt es sich um eine "Transformation" der Haushalts- in eine Personenstichprobe. Zu dieser Klasse gehört auch die vom Design abhängige Gewichtung der nichteingetragenen Haushalte in der Stichprobe.

Daneben ergeben sich - abhängig vom äußeren Rahmen der Untersuchung (wie z.B. Thema, Interviewlänge, Feldzeit, etc.), durch Nonresponse und andere Ausfälle - Abweichungen von der ursprünglichen modellgetreuen Stichprobe, die erheblich sein können und die in der Regel auch nicht zu vernachlässigen sind (z.B. werden kleinere Haushalte oft schwerer angetroffen, jüngere Personen haben u.U. weniger Zeit, andere weniger Interesse am Thema, usw.)

Deshalb ist (zumindest bei den üblichen Ausschöpfungsraten) eine a-posteriori-Anpassung ("Redressement") der Stichprobe an vorliegende soziodemografische Strukturen (z.B. der amtlichen Statistik oder aus anderen Quellen) dringend erforderlich, um die zwangsläufigen und unvermeidbaren Defizite der realisierten Stichprobe auszugleichen oder wenigstens zu verringern.

Die entsprechenden Merkmale und ihre Ausprägungen hängen insbesondere von der Fallzahl und anderer Gegebenheiten der jeweiligen Umfrage, der Verfügbarkeit aktueller und valider Strukturen sowie der Feinheit der geplanten Auswertungen ab.

Designgewichtung und Redressement sind in der Regel unerlässlich für qualitativ anspruchsvolle Stichproben in der Umfrageforschung. Sonst ist die Bezeichnung "repräsentativ" jedenfalls nicht gerechtfertigt.

Das eingesetzte Gewichtungsverfahren ist nach dem aus der Informationstheorie stammenden Prinzip "des minimalen Informationsverlusts" entwickelt und stellt sicher, dass die vorgegebenen komplexen Strukturen exakt und mit minimaler Veränderung der Stichprobe statistisch vollkommen transparent reproduziert werden.

Anmerkungen:

1. RDD

Das reine "Random Digit Dialing"-Verfahren (RDD, alternative, ebenso korrekte und verbreitete Schreibweise: "Random Digit Dialling"), bei dem Telefonnummern mit Zufallszahlengeneratoren erzeugt werden, ist generell weniger optimal und besonders nicht in Ländern angezeigt, in denen die Telefonnummern (wenigstens innerhalb von Ortsnetzen) keine feste Länge haben, was außer in Deutschland z.B. auch in Österreich und Großbritannien der Fall ist. Abgesehen von einer Reihe statistischer Nachteile, ist es wegen der geringen Trefferquote zudem höchst ineffizient. Seine Anwendung ist generell nur dann gerechtfertigt, wenn keinerlei forschungsökonomisch akzeptablen Alternativen zu Verfügung stehen.

2. RLD

Beim sogenannten "Random-Last-Digit"-Verfahren (RLD) werden zunächst eingetragene Nummern ausgewählt (wir wollen hier einmal unterstellen, dass diese Auswahl fachgerecht erfolgt) und es wird beabsichtigt, nichteingetragene Haushalte durch die Variation der letzten Ziffer oder der beiden letzten Ziffern dieser Telefonnummern zu erhalten. Fatalerweise bleibt man dadurch vor allem in den herkömmlichen "alten" Hunderterblocks mit relativ hoher Eintragsdichte jedoch relativ geringer Blockdichte, also mit wenigen validen nichteingetragenen Nummern, wobei man dann bei Treffern valider Nummern von privaten Haushalten überwiegend solche vorfindet, die ihrerseits auch eingetragen sind. Mehrfach eingetragene Haushalte haben dabei natürlich auch eine entsprechend höhere Auswahlchance.

Weil darüber hinaus wesentliche Kenntnisse über den Nummernraum fehlen und diese auch nicht mit statistischen Mitteln *sui generis*, also aus der Stichprobe selbst heraus, geschätzt werden können, sind die Treffer bei so erzeugten Nummern völlig willkürlich. Somit können die Verzerrungen auch nachträglich nicht durch Gewichtung geheilt werden, denn sie entziehen sich jeder formalen statistischen Einschätzung.

Und weil sich die Verzerrungseffekte zudem gegenseitig verstärken, ist RLD tatsächlich noch schlechter zu bewerten, als wenn sich die Auswahl nur auf eingetragene Telefonnummern alleine beschränkt. Leider findet dieses relativ einfache, aber gleichwohl völlig ungeeignete Verfahren aus Unkenntnis oder Ignoranz heraus immer wieder Anwendung.

Günther Rösch
Maisach-Frauenberg, im Juni 2004

Büro für Erhebungsdesign & Datenanalyse
Tel: 08141-39121
Fax: 08141-39123
guenther.roesch@roesch-research.de

Anhang:

Alters- und Geschlechtsverteilung einer telefonischen Umfrage der deutschen Bevölkerung ab 18 Jahren in Privathaushalten in Bayern im Dezember 2007 (n=1000)

Die Zahlen unterstreichen die soziodemographischen Unterschiede zwischen eingetragenen und nicht-eingetragenen Privathaushalten. Die Vergleichswerte der entsprechenden Wohnbevölkerung dienen nur der Veranschaulichung und dürfen hier schon wegen der speziellen Design-Gegebenheiten (z.B. unterschiedliche Auswahlwahrscheinlichkeiten für Personen in unterschiedlich großen Haushalten) keineswegs als Sollwerte - quasi als Qualitätsmaßstab für die ungewichtete Stichprobe - fehlinterpretiert werden!

	Eingetragene ungewichtet	Nicht-Eingetr. ungewichtet	Wohnbevölkerung in Privat-HH
männlich	%	%	%
18-29	5,1	10,1	8,3
30-39	5,7	11,6	8,8
40-49	9,0	9,5	10,4
50-64	9,5	6,5	11,1
65+	8,9	3,0	9,6
weiblich			
18-29	5,7	15,1	8,2
30-39	11,5	12,6	8,5
40-49	16,2	17,1	10,1
50-64	15,7	13,1	11,2
65+	12,6	1,5	13,8
	100,0	100,0	100,0