Stichproben in der Theorie (Version 1.1)

Gabler, Siegfried; Häder, Sabine

Erstveröffentlichung / Primary Publication
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Gabler, S., & Häder, S. (2015). *Stichproben in der Theorie (Version 1.1).* (GESIS Survey Guidelines). Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.15465/gesis-sg_009

gesis Leibniz-Institut für Sozialwissenschaften

GESIS Survey Guidelines

Stichproben in der Theorie

Siegfried Gabler & Sabine Häder

Januar 2015, Version 1.1

Zusammenfassung

In diesem Kapitel werden die stichprobentheoretischen Grundlagen behandelt. Zunächst wird in die Begriffswelt der Stichproben eingeführt. Termini wie Grundgesamtheit, Auswahlgesamtheit und Auswahlrahmen werden erläutert. Dann werden die unterschiedlichen Arten von Zufallsstichproben einzeln behandelt. Dabei werden die Formeln für die uneingeschränkte, die geschichtete und die systematische Zufallsauswahl, die Klumpenauswahl, zweistufige Auswahlverfahren und Auswahlverfahren mit ungleichen Inklusionswahrscheinlichkeiten dargestellt. Schließlich folgen Ausführungen zur Bestimmung des notwendigen Stichprobenumfangs.

Zitierung

Gabler, Siegfried und Häder, Sabine (2015). Stichproben in der Theorie. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines). DOI: 10.15465/gesis-sg_009

1. Worum geht es?

Da in der Umfrageforschung die Einbeziehung aller Einheiten in der Regel viel zu kosten- und zeitaufwändig wäre, beschränkt man sich auf eine gewisse Anzahl ihrer Vertreter (Stichprobe), um Aussagen über Erscheinungen in der interessierenden Gesamtheit treffen zu können. Der erste, der eine solche „repräsentative Methode" vorstellte, war 1895 der Norweger Kiaer auf der Tagung des „International Statistical Institute". Allerdings erreichte er unter seinen Kollegen zunächst keine überwiegende Zustimmung, sondern löste einen Streit aus. Dennoch konnten sich in der Folge Stichprobenerhebungen in der Praxis der nationalen statistischen Ämter immer mehr durchsetzen. Zugleich wurde an ihrer theoretischen Begründung, der Stichprobentheorie, gearbeitet. Diese erhielt schließlich von Godambe (1955) ein einheitliches theoretisches Fundament. Inzwischen sind Stichprobenerhebungen aus dem täglichen Leben längst nicht mehr weg zu denken. Sie begegnen uns praktisch überall. Speziell im Vorfeld von Wahlen wird z.B. immer wieder darüber spekuliert, wie Prognosen von Meinungsforschungsinstituten eigentlich zustande kommen. Die theoretischen Grundlagen für die Beantwortung dieser und anderer Fragen sollen im Folgenden behandelt werden.

2. Besser Stichprobe oder Totalerhebung?

Uns interessiert ein Kennwert einer Gesamtheit, z.B. das mittlere Nettoeinkommen der Haushalte in Mannheim. Wie können wir diese Information beschaffen?

- Wir können alle Mannheimer Haushalte nach ihrem Nettoeinkommen fragen und dann den Mittelwert berechnen.
- Wir wählen einige Mannheimer Haushalte aus und bitten diese um Auskunft über ihr Nettoeinkommen. Wenn die Haushalte nach bestimmten Regeln ausgewählt werden, können wir anschließend mit einer gewissen Wahrscheinlichkeit auf das mittlere Nettoeinkommen aller Mannheimer Haushalte schließen.

Zum Ziel der Informationsbeschaffung über eine Gesamtheit stehen also zwei Wege offen: Totalerhebung (engl. census) und Stichprobe (engl. sample).

Die Totalerhebung, d.h. die Erhebung aller Elemente einer Gesamtheit, hat den Vorteil, dass die gesuchten Kennwerte genau angegeben werden können. Im oben genannten Beispiel würde unser Ergebnis z.B. lauten: Das mittlere Haushaltsnettoeinkommen in Mannheim beträgt X €.

Bei der Berechnung des Kennwerts auf der Basis einer Stichprobe würde das Ergebnis komplizierter klingen, z.B.: Mit einer Wahrscheinlichkeit von 95% beträgt das Mannheimer Haushaltsnettoeinkommen X € ± Y €. Offensichtlich ist das Ergebnis, das aufgrund der Stichprobenerhebung zustande kommt, deutlich komplexer und eben nicht „ganz sicher".

Warum wird dann nicht in jedem Fall eine Totalerhebung durchgeführt? Stichprobenerhebungen haben durchaus eine Reihe von Vorteilen:

- Sie sind billiger als Totalerhebungen.
- Die Ergebnisse stehen bei Stichprobenerhebungen schneller zur Verfügung als bei Totalerhebungen.

- Bei Stichprobenerhebungen ist eine geringere Größe des Mitarbeiterstabes notwendig als bei Totalerhebungen. Die Mitarbeiter können dafür spezieller geschult werden.
- Es ist bei Stichprobenerhebungen ein besserer Umgang mit Ausfällen, z.B. aufgrund von Nichterreichbarkeit der zu Befragenden, möglich als bei Totalerhebungen. So kann bei einer Bevölkerungsbefragung beispielsweise die Zahl der Kontaktversuche auf vier oder fünf erhöht werden. Im Rahmen einer Totalerhebung wäre dies zu kostenaufwändig. Damit hängt der zunächst paradox erscheinende Umstand zusammen, dass u.U. Stichprobenerhebungen eine höhere Messgenauigkeit haben können als Befragungen, die als Totalerhebungen geplant waren.
- Mitunter sind Stichprobenerhebungen die einzige Möglichkeit, um an Informationen über die Gesamtheit zu kommen. Dies ist z.B. dann der Fall, wenn der Untersuchungsgegenstand bei der Messung zerstört wird (z.B. Erfassung der Brenndauer von Glühlampen als Bestandteil der Gütekontrolle).
- Insgesamt ist die Last für die Befragten geringer, da weniger von ihnen um Auskunft gebeten werden.

Allerdings gibt es auch Bedingungen, die gegen den Einsatz von Stichprobenerhebungen sprechen: So ist es bei verhältnismäßig kleinen Gesamtheiten, zum Beispiel vom Umfang $N = 30$, in der Regel wenig sinnvoll, eine Stichprobe zu ziehen. So ist auch eine Totalerhebung eher angesagt, wenn es gilt, zu kleinen Subpopulationen innerhalb der Gesamtheit Aussagen zu treffen, die bei einer Stichprobenerhebung wegen zu geringer Stichprobeneinheiten sehr unpräzise sein können. Eine Totalerhebung ist auch dann empfehlenswert, wenn im Vorhinein bekannt ist, dass es sich um eine sehr heterogene Gesamtheit handelt. Ein Beispiel für eine hinsichtlich eines Merkmals extrem heterogene Gesamtheit sind die Fingerabdrücke (Muster der Papillarlinien an der Fingerkuppe). Es kann davon ausgegangen werden, dass keine zwei Fingerabdrücke auf der Welt identisch sind.

In gewissen Fällen verbietet sich eine Stichprobenerhebung, und die Totalerhebung ist die einzige Möglichkeit. Dies gilt z.B. für so genannte Rückrufaktionen bei Autos.

3. Welche Begriffe sind wichtig?

Die Menge aller Elemente, für die die aus der Stichprobe gewonnenen Aussagen gültig sein sollen, bezeichnet man als Grundgesamtheit oder Untersuchungsgesamtheit. Diese muss zu Beginn der Untersuchung sachlich, räumlich und zeitlich exakt abgegrenzt werden.

Beispiel: In einer telefonischen Erhebung soll festgestellt werden, wie sich Veränderungen im Telekommunikationsverhalten auf soziale Beziehungen auswirken. Dazu erfolgt eine Abgrenzung der Grundgesamtheit zunächst derart, dass sie alle Personen umfasst, die telefonisch erreichbar sind. Außerdem wird die Abgrenzung „deutschsprachig" getroffen, die ebenfalls auf praktische Erwägungen im Hinblick auf die geplante telefonische Befragung zurückgeht.

sachlich:	alle deutschsprachigen Personen, die über Festnetz oder Mobilfunk erreichbar sind,
räumlich:	die in der Bundesrepublik Deutschland leben und
zeitlich:	die 16 Jahre oder älter sind (bei jüngeren Personen wäre die Zustimmung der Erziehungsberechtigten nötig).

In einem nächsten Schritt muss recherchiert werden, ob es eine Liste (engl. frame) gibt, auf der die Elemente der Grundgesamtheit akzeptabel verzeichnet sind. Was bedeutet in diesem Zusammenhang akzeptabel?

Die Liste ist hinreichend aktuell. Beispiel: Einwohnermelderegister der Gemeinden haben in der Regel ein Timelag, d.h. enthalten Fehler in Bezug auf mobile Personen, Geburten und Sterbefälle. Das hat z.B. der Zensus 2011 deutlich gemacht, indem er zeigte, dass in Deutschland insgesamt ca. 1,5 Millionen Menschen weniger wohnen als aufgrund der Angaben aus den Einwohnermeldeämtern und den entsprechenden Fortschreibungen angenommen (Registerfehler). Trotzdem sind die Register eine häufig verwendete Auswahlgrundlage, da eine bessere Liste nicht verfügbar ist. Aktuell heißt:

- Jedes Mitglied der Grundgesamtheit ist einmal und nur einmal aufgeführt, d.h. die Liste weist weder Karteileichen (Elemente enthalten, die nicht zur Grundgesamtheit gehören, engl. overcoverage) noch Fehlbestände (Elemente nicht enthalten, die zur Grundgesamtheit gehören, engl. undercoverage) auf.
- Die Liste ist für die Erhebung zugänglich und ihre Nutzung ist nicht zu teuer. (Beispiel Einwohnermelderegister: Die Untersuchung muss im öffentlichen Interesse liegen. Das heißt, dass nicht für jedes beliebige Untersuchungsthema Personenstichproben ausgehändigt werden. Zudem muss das Untersuchungsinstitut eine gültige Unbedenklichkeitsbescheinigung vorweisen können (vgl. Albers 1997, S. 118f.). Die Preise für Stichproben aus Einwohnermelderegistern sind jeweils von den Bundesländern festgelegt und unterscheiden sich recht stark.)

Idealer Weise sind die auf der Liste verzeichneten Elemente (Auswahlgesamtheit, engl. frame population) und die Grundgesamtheit (engl. target population) identisch. Dies ist in der Praxis jedoch nur sehr selten der Fall. Deshalb ist es notwendig, über Größe und Art der Differenz zwischen Grund- und Auswahlgesamtheit abzuwägen. Im Allgemeinen ist das Problem geringer, wenn die Abweichungen nicht systematisch, sondern zufällig sind, d.h. keine Merkmale betreffen, die mit dem Untersuchungsziel in Verbindung stehen. So ist zum Beispiel das Telefonbuch als Liste bei bundesweiten Erhebungen aufgrund des hohen Anteils Nicht-Eingetragener nicht geeignet. Für Erhebungen in einer südwestdeutschen ländlichen Region, in der nach wie vor fast alle Haushalte im Telefonbuch registriert sind, könnte es als Auswahlrahmen dagegen durchaus brauchbar sein.

Schließlich gilt es, den Umfang der Grundgesamtheit vor Beginn der Untersuchung abzuschätzen, um beurteilen zu können, ob eine Stichprobenerhebung überhaupt sinnvoll ist. Dies ist insbesondere dann wichtig, wenn keine Liste mit den interessierenden Elementen vorliegt, sondern die Stichprobe über ein Screening rekrutiert werden soll. Beispielsweise sollte im Rahmen eines von GESIS beratenen Projektes im Jahr 2000 eine bundesweite Befragung von Eltern mit Kindern, die nicht älter als 8 Jahre sind, durchgeführt werden. Eine Erhebung über Einwohnermelderegister schied deshalb aus, weil die Stichprobe möglichst ungeklumpt sein sollte. Dies lässt sich am besten mit Hilfe eines telefonischen Screenings bewerkstelligen. Aus dem Mikrozensus 1998 wurde errechnet, dass der Anteil von Haushalten mit Kindern unter 9 Jahren in Westdeutschland 13%, in Ostdeutschland 11% betragen müsse. In einem Pretest musste allerdings festgestellt werden, dass nur 7,4% der angerufenen Haushalte in Westdeutschland und 4,6% der befragten ostdeutschen Haushalte angaben, ein Kind im Alter von unter 9 Jahren zu haben. Bei einem angezielten Netto-Stichprobenumfang von je n = 1.500 in West- und Ostdeutschland hätten damit in Westdeutschland 20.271 und in Ostdeutschland 32.609 Haushalte kontaktiert werden müssen. Dabei ist noch unterstellt, dass alle Haushalte an der Studie teilzunehmen bereit sind. Geht man aber von der realistischeren Annahme aus, dass nur etwa die Hälfte der ausgewählten Haushalte auskunftswillig ist und auch angetroffen wird, wären insgesamt mehr als 100.000 Anrufe nötig, um die angezielte Fallzahl zu rekrutieren. Es hängt nun von den finanziellen, personellen und technischen Möglichkeiten des Erhebungsinstituts ab, ob es in der Lage ist, diesen

Auftrag zu erfüllen oder ob möglicherweise ein anderer Zugang zur Bearbeitung des Forschungsthemas gewählt werden muss.

In der Markt- und Sozialforschung hat sich in den letzten Jahren der Trend durchgesetzt, vermindert persönlich-mündliche Interviews, dafür Online-Interviews durchzuführen. Dies verdeutlichen die folgenden Zahlen der Interviews, die von den Instituten des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e. V. in den letzten Jahren durchgeführt worden sind (http://www.adm-ev.de/). Die Vor- und Nachteile der einzelnen Erhebungsmodi können an dieser Stelle jedoch nicht diskutiert werden. Wir verweisen dazu auf das Kapitel „Umfragedesign".

Quantitative Interviews der Mitgliedsinstitute des ADM nach Befragungsart (in %)

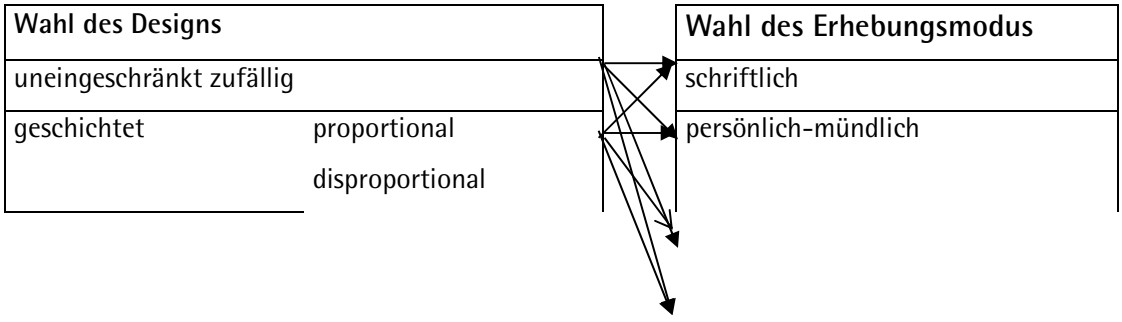
	1990	1995	2000	2005	2010	2013
Persönliche Interviews	65	60	34	24	21	22
Telefoninterviews	22	30	41	45	35	36
Schriftliche Interviews	13	10	22	9	6	6
Online-Interviews			3	22	38	36

4. Was sind Zufallsstichproben und welche Arten gibt es?

Generell gibt es unterschiedliche Möglichkeiten, Stichproben aus einer Gesamtheit zu entnehmen. Dazu gehören:

- Uneingeschränkte Zufallsauswahl
- Geschichtete Zufallsauswahl
- Systematische Zufallsauswahl
- Klumpenauswahl
- Zweistufige Auswahlverfahren
- Auswahlverfahren mit ungleichen Inklusionswahrscheinlichkeiten

Wenn z.B. eine akzeptable Liste mit den Elementen der Auswahlgesamtheit vorliegt, wird aus dieser uneingeschränkt zufällig oder – wenn Zusatzinformationen zur Verfügung stehen – geschichtet ausgewählt. Dabei spielt der Erhebungsmodus keine Rolle, d.h. diese Auswahlmechanismen sind für alle Befragungsmodi (schriftlich, persönlich-mündlich, telefonisch oder online) anwendbar.



telefonisch
per E-Mail/Internet

Liegt dagegen keine Liste vor, die für die Auswahl geeignet ist, muss eine Ersatzkonstruktion zum Einsatz kommen. Diese kann z.B. in einer mehrstufigen Flächenstichprobe mit Random-Route-Elementen bestehen.

Im Folgenden sollen nun die theoretischen Grundlagen für die verschiedenen Arten von Zufallsstichproben vorgestellt werden. Beweise zu den Aussagen findet man z.B. in den Büchern von Lohr (1999) und Särndal (1992).

Uneingeschränkte Zufallsauswahl

Bei der Auswahl von Einheiten aus einer Gesamtheit ist es von Vorteil, wenn diese Einheiten nach einem Wahrscheinlichkeitsgesetz gezogen werden, da sich dann statistisch gesicherte Aussagen über interessierende Parameter der Gesamtheit machen lassen.

Wir stellen uns zunächst vor, dass nur eine Einheit der Grundgesamtheit vom Umfang *N* auszuwählen ist und dass die Auswahl für jede Einheit gleich und demnach 1/*N* ist. Die Durchführung der Ziehung kann über ein Zufallsexperiment realisiert werden.

Wiederholt man dieses Zufallsexperiment *n* mal unabhängig und notiert die ausgewählten Einheiten in einem Vektor $(I_1,\ldots,I_n)$, wobei I_k die in der *k*-ten Wiederholung des Zufallsexperiments, d.h. im *k*-ten

Zug, ausgewählte Einheit bezeichnet, so gibt es $1/N^n$ verschiedene gleich wahrscheinliche Ergebnisse. Dieses Stichprobendesign wird auch *als uneingeschränkte Zufallsauswahl mit Zurücklegen (*engl. *srswr)* bezeichnet.

Das Stichprobenmittel $\bar{y} = \frac{1}{n}\sum_{k=1}^{n} Y_{I_k}$ ist eine Zufallsvariable mit

$$E_{srswr}(\bar{y}) = \bar{Y} \text{ mit } \bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$$

$$\text{var}_{srswr}(\bar{y}) = \frac{\sigma^2}{n} \text{ mit } \sigma^2 = \frac{1}{N}\sum_{k=1}^{N}\left(Y_k - \bar{Y}\right)^2$$

Eine erwartungstreue Schätzfunktion υ_{srswr} für die Varianz des Stichprobenmittels ist durch

$$\upsilon_{srswr} = \frac{s^2}{n} \text{ mit } s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_{I_k} - \bar{y}\right)^2 \text{ als (korrigierte) Stichprobenvarianz}$$

gegeben.
Für große Stichprobenumfänge *n* ist daher

$$\left[\bar{y} - 1{,}96\sqrt{\frac{s^2}{n}}; \bar{y} + 1{,}96\sqrt{\frac{s^2}{n}}\right]$$

ein 95%-Konfidenzintervall für den unbekannten interessierenden Mittelwert $\bar{Y}$ in der Gesamtheit.

Hat eine im *i*-ten Zug ausgewählte Einheit bei den folgenden Zügen keine positive Wahrscheinlichkeit mehr, erneut ausgewählt zu werden, und haben alle nicht bis zum *i*-ten Zug ausgewählten Einheiten dieselbe Wahrscheinlichkeit in den restlichen Zügen ausgewählt zu werden, spricht man von *uneingeschränkter Zufallsauswahl ohne Zurücklegen (*engl. *srs).*

Das Stichprobenmittel $\bar{y} = \frac{1}{n}\sum_{k=1}^{n} Y_{I_k}$ ist eine Zufallsvariable mit

$$E_{srs}(\bar{y}) = \bar{Y} \text{ mit } \bar{Y} = \frac{1}{N}\sum_{k=1}^{N} Y_k$$

$$\text{var}_{srs}(\bar{y}) = \frac{S^2}{n}\left(1-\frac{n}{N}\right) \text{ mit } S^2 = \frac{1}{N-1}\sum_{k=1}^{N}\left(Y_k - \bar{Y}\right)^2 .$$

Eine erwartungstreue Schätzfunktion υ_{srs} für die Varianz des Stichprobenmittels ist durch

$\upsilon_{srs} = \frac{s^2}{n}\left(1-\frac{n}{N}\right)$ mit $s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_{I_k} - \bar{y}\right)^2$ als (korrigierte) Stichprobenvarianz gegeben.

Für große Stichprobenumfänge *n* ist daher

$$\left[\bar{y} - 1{,}96\sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}; \bar{y} + 1{,}96\sqrt{\frac{s^2}{n}\left(1-\frac{n}{N}\right)}\right]$$

ein 95%-Konfidenzintervall für den unbekannten interessierenden Mittelwert $\bar{Y}$ in der Gesamtheit.

Ist der Auswahlsatz *n/N* klein, etwa unter 5%, so wird der Korrekturfaktor $\left(1-\frac{n}{N}\right)$ oft vernachlässigt und die Formeln des *Ziehens mit Zurücklegen* verwendet.

Ein wichtiger Spezialfall liegt vor, wenn die *y*-Werte nur die Ausprägungen 0 oder 1 annehmen können und $\bar{Y}$ als Anteilswert interpretiert werden kann. Statt $\bar{Y}$ wird dann häufig *P* geschrieben. Analog bezeichnet man das Stichprobenmittel mit *p*. Die Varianz σ^2 lässt sich schreiben als $\sigma^2 = P(1-P)$ und $s^2 = \frac{n}{n-1}p(1-p)$. Wegen der großen Bedeutung, die Anteilswerte haben, werden die Formeln explizit aufgeführt.

Bei der *uneingeschränkten Zufallsauswahl mit Zurücklegen* gilt

$$E_{srswr}(p) = P$$

$$\text{var}_{srswr}(p) = \frac{P(1-P)}{n}$$

Eine erwartungstreue Schätzfunktion υ_{srswr} für die Varianz des Stichprobenanteils *p* ist durch

$\upsilon_{srswr} = \frac{p(1-p)}{n-1}$ gegeben.

Für große Stichprobenumfänge *n* ist daher

$$\left[p - 1{,}96\sqrt{\frac{p(1-p)}{n-1}}; p + 1{,}96\sqrt{\frac{p(1-p)}{n-1}}\right]$$

ein 95%-Konfidenzintervall für den unbekannten interessierenden Mittelwert *P* in der Gesamtheit. Da *n* groß ist, kann *n*-1 auch durch *n* ersetzt werden.

Bei der *uneingeschränkten Zufallsauswahl ohne Zurücklegen* gilt

$$E_{srs}(p) = P$$

$$\mathrm{var}_{srs}(p) = \frac{P(1-P)}{n}\left(1-\frac{n-1}{N-1}\right).$$

Eine erwartungstreue Schätzfunktion υ_{srs} für die Varianz des Stichprobenanteils *p* ist durch

$$\upsilon_{srs} = \frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right) \text{ gegeben.}$$

Für große Stichprobenumfänge *n* ist daher

$$\left[p-1{,}96\sqrt{\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)};p+1{,}96\sqrt{\frac{p(1-p)}{n-1}\left(1-\frac{n}{N}\right)}\right]$$

ein 95%-Konfidenzintervall für den unbekannten interessierenden Mittelwert *P* in der Gesamtheit.
Wenn *N* und *n* groß sind, kann *N*-1 durch *N* und *n*-1 durch *n* ersetzt werden.

Geschichtete Zufallsauswahl

Häufig sind Gesamtheiten in natürlicher Weise geschichtet, und die Auswahl in einer Schicht wird unabhängig von der Auswahl in einer anderen Schicht vorgenommen. Wählt man uneingeschränkt zufällig aus und bezeichnet $N(h)$ bzw. $n(h)$ den Umfang der Gesamtheit bzw. der Stichprobe der *h*-ten Schicht (*h*=1,...,*H*), so wird das gewogene arithmetische Mittel der Stichprobenmittel $\overline{y}(h)$ zur Schätzung verwendet. Genauer gilt für

$$\overline{y}_{str} = \sum_{h=1}^{H}\frac{N(h)}{N}\overline{y}(h)$$

$$E_{str}(\overline{y}_{str}) = \overline{Y}$$

$$\mathrm{var}_{str}(\overline{y}_{str}) = \sum_{h=1}^{H}\left(\frac{N(h)}{N}\right)^2\frac{S^2(h)}{n(h)}\left(1-\frac{n(h)}{N(h)}\right).$$

Eine erwartungstreue Schätzfunktion υ_{str} für die Varianz des geschichteten Schätzers ist durch

$$\upsilon_{str} = \sum_{h=1}^{H}\left(\frac{N(h)}{N}\right)^2\frac{s^2(h)}{n(h)}\left(1-\frac{n(h)}{N(h)}\right)$$

gegeben.
Für große Stichprobenumfänge *n*(*h*) ist daher

$$\left[\overline{y}_{str}-1{,}96\sqrt{\upsilon_{str}};\overline{y}_{str}+1{,}96\sqrt{\upsilon_{str}}\right]$$

ein 95%-Konfidenzintervall für den unbekannten interessierenden Mittelwert $\overline{Y}$ in der Gesamtheit.

Bei der Aufteilung des Stichprobenumfangs *n* auf die Schichten werden üblicherweise drei Möglichkeiten angegeben:

Proportionale Aufteilung

$$n(h) = \frac{N(h)}{N}\cdot n$$

Optimale Aufteilung

$$n(h) = n \cdot \frac{N(h)\sqrt{S^2(h)}}{\sum_{g=1}^{H} N(g)\sqrt{S^2(g)}}$$

Kostenoptimale Aufteilung

$$n(h) = \frac{c}{\overline{c}(h)} \cdot \frac{N(h)\sqrt{S^2(h)\,\overline{c}(h)}}{\sum_{g=1}^{H} N(g)\sqrt{S^2(g)\,\overline{c}(g)}}$$

Dabei sind $\overline{c}(h)$ die Durchschnittskosten der Erhebung einer Einheit in der h-ten Schicht und $c = \sum_{h=1}^{H} n(h)\overline{c}(h)$ die Gesamtkosten, die die Erhebung kosten darf.

Sind die Durchschnittskosten für jede Schicht gleich, wird aus der kostenoptimalen Aufteilung die optimale Aufteilung. Sind die Varianzen der y-Werte in den Schichten gleich groß, wird aus der optimalen Aufteilung die proportionale Aufteilung. Der geschichtete Schätzer ist bei proportionaler Aufteilung mit dem Stichprobenmittel identisch.

Warum sollte man schichten?

Aus der Formel für die Varianz des geschichteten Schätzers

$$\mathrm{var}_{srs}\left(\sum_{h=1}^{H} \frac{N(h)}{N}\overline{y}(h)\right) = \sum_{h=1}^{H}\left(\frac{N(h)}{N}\right)^2 \frac{S^2(h)}{n(h)}\left(1-\frac{n(h)}{N(h)}\right)$$

ist ersichtlich, dass sie dann klein ist, wenn die y-Werte in den Schichten nur wenig streuen. In solchen Fällen erzielt man gegenüber dem Stichprobenmittel einen *Schichtungsgewinn*.

Systematische Zufallsauswahl

Bei der systematischen Zufallsauswahl werden die Einheiten der Gesamtheit vom Umfang $N = nH$ in der Regel zuerst nach einer oder mehreren Variablen geordnet. Zunächst wird eine Startzahl K mit $1 \le K \le H$ zufällig und danach die Einheiten mit den Nummern $K, K+H,..., K+(n-1)H$ ausgewählt. Dann gilt für den Schätzer

$$\overline{y}_{sys} = \frac{1}{n}\sum_{i=1}^{n} Y_{K+iH}$$

$$E_{sys}\left(\overline{y}_{sys}\right) = \overline{Y}$$

$$\mathrm{var}_{sys}\left(\overline{y}_{sys}\right) = \frac{S^2}{n}\frac{N-1}{N}\left(1+(n-1)\rho\right).$$

Dabei ist ρ der Intraklassen-Korrelations-Koeffizient, der als Korrelations-Koeffizient zwischen Paaren von Einheiten innerhalb derselben (systematischen) Stichprobe interpretiert werden kann, d.h.

$$\rho = \frac{\sum_{k=1}^{H} \sum_{\substack{i,j=0 \\ i \neq j}}^{n-1} (Y_{k+iH} - \bar{Y})(Y_{k+jH} - \bar{Y})}{(n-1)(N-1)S^2}$$

Ein erwartungstreuer Schätzer für die Varianz existiert nicht. Da $-\frac{1}{n-1} \leq \rho \leq 1$ gilt, kann $\text{var}_{sys}\left(\bar{y}_{sys}\right)$ zwischen 0 und $S^2 \frac{N-1}{N}$ liegen. Der Extremwert 0 tritt dann auf, wenn die Stichprobenmittel der systematischen Stichproben alle gleich sind. Der andere Extremfall ergibt sich, wenn die *y*-Werte in einer systematischen Stichprobe alle gleich sind. Man kann sich auch leicht überlegen, dass $\text{var}_{sys}\left(\bar{y}_{sys}\right)$ mit der Varianz des Stichprobenmittels bei uneingeschränkter Zufallsauswahl übereinstimmt, wenn $\rho = -\frac{1}{N-1}$ ist.

Die systematische Zufallsauswahl ist ein Spezialfall der Klumpenauswahl.

Klumpenauswahl

Angenommen, die Gesamtheit ist - wie bei der Schichtung - in *H* Klumpen (Cluster) zerlegt, wobei N_h der Umfang des *h*-ten Klumpens sei. *n* Klumpen werden uneingeschränkt zufällig ausgewählt. Als Schätzer für $\bar{\bar{Y}} = \frac{1}{K}\sum_{h=1}^{H} N_h \bar{Y}_h$ mit $K = \sum_{h=1}^{H} N_h$ verwendet man

$$\bar{y}_{cl} = \frac{H}{nK}\sum_{h=1}^{H} L_h N_h \bar{Y}_h$$

Dabei ist $\bar{Y}_h$ das Mittel der *y*-Werte im *h*-ten Klumpen $(h = 1,...,H)$ und

$$L_h = \begin{cases} 1 & \text{falls } h\text{-ter Klumpen ausgewählt} \\ 0 & \text{sonst} \end{cases}.$$

Dann gilt

$$E_{cl}\left(\bar{y}_{cl}\right) = \bar{\bar{Y}}$$

$$\text{var}_{cl}\left(\bar{y}_{cl}\right) = \frac{S_{cl}^2}{n}\left(1 - \frac{n}{H}\right) \text{ mit } S_{cl}^2 = \frac{1}{H-1}\sum_{h=1}^{H}\left(N_h \bar{Y}_h - \frac{K}{H}\bar{\bar{Y}}\right)^2.$$

Eine erwartungstreue Schätzfunktion υ_{cl} für $\text{var}_{cl}\left(\bar{y}_{cl}\right)$ ist im Falle *n*>1 durch

$$\text{var}_{cl}\left(\bar{y}_{cl}\right) = \frac{S_{cl}^2}{n}\left(1 - \frac{n}{H}\right) \text{ mit } s_{cl}^2 = \frac{1}{n-1}\sum_{h=1}^{H} L_h\left(N_h \bar{Y}_h - \frac{K}{H}\bar{y}_{cl}\right)^2.$$

gegeben.

Zweistufiges Auswahlverfahren

Geschichtete Auswahl und Klumpenauswahl sind zwei Spezialfälle des zweistufigen Auswahlverfahrens. Die Gesamtheit besteht aus N Primäreinheiten (primary sampling units). Die i-te Primäreinheit enthalte M_i Sekundäreinheiten. Y_{ij} bezeichne die j-Ausprägung des Untersuchungsmerkmals in der i-ten Primäreinheit. Weiter setzen wir

$$Y_i = M_i\overline{Y}_i = \sum_{j=1}^{M_i} Y_{ij}, \overline{\overline{Y}} = \frac{1}{K}\sum_{i=1}^{N} Y_i = \frac{N}{K}\overline{Y}, S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2, S_i^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(Y_{ij} - \overline{Y}_i\right)^2 \text{ mit } K = \sum_{i=1}^{N} M_i$$

Geht man davon aus, dass n Primäreinheiten uneingeschränkt zufällig ohne Zurücklegen ausgewählt und m_i Sekundäreinheiten aus der i-ten Primäreinheit ebenfalls uneingeschränkt zufällig ohne Zurücklegen ausgewählt werden, so verwendet man als Schätzer für $\overline{\overline{Y}}$

$$\hat{\overline{\overline{Y}}} = \frac{1}{n}\sum_{i=1}^{N} L_i\hat{Y}_i = \frac{1}{n}\sum_{i=1}^{N}\frac{M_i}{m_i}\sum_{j=1}^{M_i} L_{ij}Y_{ij}$$

mit

$$L_{ij} = \begin{cases} 1 & \text{falls j-te Sekundäreinheit in i-ter Primäreinheit ausgewählt wird} \\ 0 & \text{sonst} \end{cases} \qquad \left(j = 1,\ldots,M_i; i = 1,\ldots,N\right)$$

und

$$L_i = \begin{cases} 1 & \text{falls } i\text{-te Primäreinheit ausgewählt wird} \\ 0 & \text{sonst} \end{cases} \qquad \left(i = 1,\ldots,N\right).$$

Dann gilt

$$E\hat{\overline{\overline{Y}}} = \overline{\overline{Y}}$$

$$\operatorname{var}\left(\hat{\overline{\overline{Y}}}\right) = \frac{1}{K^2}\left[\frac{N^2}{n}\left(1-\frac{n}{N}\right)S_t^2 + \frac{N}{n}\sum_{i=1}^{N} M_i^2\frac{S_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right)\right]$$

$$= \frac{1}{K^2}E\left(\frac{N^2}{n}\left(1-\frac{n}{N}\right)s_t^2 + \frac{N}{n}\sum_{i=1}^{N} L_i M_i^2\frac{s_i^2}{m_i}\left(1-\frac{m_i}{M_i}\right)\right)$$

Dabei ist $\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{M_i} L_{ij}Y_{ij}$ das Stichprobenmittel in der i-ten Primäreinheit,

$s_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{M_i} L_{ij}(Y_{ij} - \overline{y}_i)^2$ die Stichprobenvarianz in der i-ten Primäreinheit und

$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{N} L_i\left(\hat{Y}_i - \overline{\hat{Y}}\right)^2$ mit $\overline{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{N} L_i\hat{Y}_i$ die Stichprobenvarianz der $\hat{Y}_i$-Werte.

Auswahlverfahren mit ungleichen Inklusionswahrscheinlichkeiten

Es bezeichne π_{ij} die Wahrscheinlichkeit dafür, das *i*-te und *j*-te Element der Grundgesamtheit auszuwählen. Statt π_{ii} schreibt man einfacher π_i. Als erwartungstreue Schätzfunktion für die Gesamtsumme $Y = \sum_{i=1}^{N} Y_i$ verwendet man den Horvitz-Thompson-Schätzer

$$\hat{Y}_{HT} = \sum_{i=1}^{N} L_i \frac{Y_i}{\pi_i}$$

mit $L_i = \begin{cases} 1 & \text{falls } i\text{-te Einheit ausgewählt wird} \\ 0 & \text{sonst} \end{cases}$ für $i = 1,\ldots,N$.

Dabei wird vorausgesetzt, dass alle π_i positiv sind. Für die Varianz des Horvitz-Thompson-Schätzers erhält man

$$\operatorname{var}\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\left(\pi_{ij} - \pi_i\pi_j\right)$$

Liegt ein Auswahlverfahren mit festem Stichprobenumfang n vor, so gilt

$$\sum_{j=1}^{N} \pi_{ij} = n\pi_i \quad \text{und} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_{ij} = n^2$$

und der sogenannte Yates-Grundy-Varianzschätzer

$$\mathrm{v}_{YG} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{L_i L_j}{\pi_{ij}}\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 \left(\pi_i\pi_j - \pi_{ij}\right)$$

schätzt $\operatorname{var}\left(\hat{Y}_{HT}\right)$ erwartungstreu, wenn alle π_{ij} positiv sind. Er ist offensichtlich nichtnegativ, wenn $\pi_i\pi_j \geq \pi_{ij}$ für alle i,j gilt.

Im Falle der uneingeschränkten Zufallsauswahl ist $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$ für $i \neq j$ und $\pi_i = \frac{n}{N}$. Der Horvitz-Thompson-Schätzer ist dann das *N*-fache des Stichprobenmittels. Bei geschichteter Zufallsauswahl ist $\pi_i = \frac{n_h}{N_h}$ für i in *h*-ter Schicht. Für $i \neq j$ beide in *h*-ter Schicht ist $\pi_{ij} = \frac{n_h(n_h-1)}{N_h(N_h-1)}$ und sonst $\pi_{ij} = \pi_i\pi_j$.

5. Wie groß muss der Stichprobenumfang sein?

Ermittlung des notwendigen Stichprobenumfangs bei uneingeschränkter Zufallsauswahl (Simple Random Sampling - SRS)

Wir befinden uns in der folgenden Situation: Der tolerierbare Stichprobenfehler und die Irrtumswahrscheinlichkeit sind vorgegeben, z.B. der Anteilswert in der Gesamtheit soll bei einer Irrtumswahrscheinlichkeit von 5% um nicht mehr als ± 3 Prozentpunkte vom Punktschätzer abweichen.

Wie groß muss dann der Stichprobenumfang sein?

Es sei

n_{srs} Stichprobenumfang unter SRS

N Umfang der Grundgesamtheit

$z_{\alpha/2}$ tabellierter Wert aus der Standardnormalverteilung, für α = 0,05 ist $z_{\alpha/2}$ = 1,96

p Anteil des interessierenden Merkmals in der Stichprobe, entweder aus Voruntersuchung bekannt oder ungünstigsten Fall p = 0,5

e zulässiger absoluter Stichprobenfehler, 2e entspricht der Länge des Konfidenzintervalls

Schätzung von Anteilswerten bei kleinem Auswahlsatz (*n/N < 0,05*)

$$n_{srs} \geq \left(\frac{z_{\alpha/2}}{e} \right)^2 \cdot p \cdot (1-p)$$

Schätzung von Anteilswerten bei großem Auswahlsatz

Berücksichtigung des Korrekturfaktors $1-\frac{n}{N}$ (Ziehen ohne Zurücklegen)

$$n_{srs} \geq \frac{N \cdot z_{\alpha/2}^2 \cdot \ p\ (1-p)}{z_{\alpha/2}^2 \cdot p\ (1-p) + N \cdot e^2}$$

Tabelle: minimaler Stichprobenumfang *n* für gegebenen absoluten Stichprobenfehler *e* bei Irrtumswahrscheinlichkeit $\alpha = 0{,}05$ für Anteile $p = 0{,}5$ und $p = 0{,}8$ (oder $p = 0{,}2$) (nach Borg 2000, S. 144)

$p = 0{,}5$			**$p = 0{,}8$ oder $p = 0{,}2$**		
N	$e = 0{,}03$	$e = 0{,}05$	*N*	$e = 0{,}03$	$e = 0{,}05$
200	168	132	200	155	110
300	234	168	300	208	135
400	291	196	400	252	152
500	340	217	500	289	165
750	440	254	750	357	185
1.000	516	278	1.000	406	197
3.000	787	341	3.000	556	227
7.500	934	365	7.500	626	238
10.000	964	370	10.000	639	240
50.000	1.045	381	50.000	674	245
100.000	1.056	383	100.000	678	245

Ermittlung des notwendigen Stichprobenumfangs bei komplexen Designs

Bei komplexen Designs gibt es in der Regel eine Varianzvergrößerung durch Clusterung und Gewichtung. Diese gilt es, bei der Bestimmung des Stichprobenumfangs zu berücksichtigen. Ein Maß dafür ist der Design-Effekt.

$$n_{kompl} = n_{srs} \cdot Deff$$

Zur Berechnung von Design-Effekten (Kish 1965, 1980, 1987)

$$Deff = v \,/\, v_0$$

mit

v Varianz des Schätzers unter komplexem Design

v_0 Varianz des Schätzers unter Simple Random Sampling
(Es hat Vorteile für v_0 die Varianz des Schätzers unter srswr zu verwenden).

Bei der Bestimmung des (Modell-)Design-Effekts für komplexe Stichprobendesigns müssen zwei Komponenten berücksichtigt werden: Der Design-Effekt aufgrund der Clusterung und der Design-Effekt aufgrund ungleicher Inklusionswahrscheinlichkeiten (Kish 1987, Beweis: Gabler/Häder/Lahiri 1999)

$$Deff = n\frac{\sum_{i=1}^{I} n_i w_i^2}{(\sum_{i=1}^{I} n_i w_i)^2}[1+(\bar{b}-1)\rho] = (1+L)[1+(\bar{b}-1)\rho] = Deff_w \cdot Deff_c$$

mit

n_i Zahl der Beobachtungen in der *i*-ten Gewichtungsklasse

w_i Gewichte in der *i*-ten Gewichtungsklasse

$n = \sum_{i=1}^{I} n_i$ Stichprobenumfang

$\bar{b}$ mittlere Clustergröße

ρ Intraklassenkorrelationskoeffizient

Über die Größenordnung von in der Umfragepraxis vorkommenden Design-Effekten gibt Kish (1987) Auskunft: „Variations of 1,0 to 3,0 of *deft* are common ...", wobei gilt $Deft = \sqrt{Deff}$. Bei Auswahlen, die gleichgewichtend auf der Haushaltsebene sind und bei denen ungleiche Wahrscheinlichkeiten lediglich auf der letzten Auswahlstufe – der Bestimmung der Zielperson im Haushalt – auftreten, ergab sich im ESS in verschiedenen Ländern in verschiedenen Jahren jeweils ein Design Effekt $Deff_w$ von ca. 1,2 (Ganninger 2010).

Literaturverzeichnis

Borg, I. (2000). *Führungsinstrument Mitarbeiterbefragung: Theorien, Tools und Praxiserfahrungen*. Göttingen: Verl. für Angewandte Psychologie.

Gabler, S., Häder, S. & Lahiri, P. (1999). A model based justification of Kish's formula for design effects for weighting and clustering. *Survey Methodology* Vol. 25, No.1.

Ganninger, M. (2010). Design Effects: Model-based versus Design-based Approach. GESIS Schriftenreihe Band 3.

Godambe, V. P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society* 17, Series B: 269-278.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kish, L. (1987). Weighting in Deft2. *The Survey Statistician*. June 1987.

Lohr, S. L., (1999). *Sampling: Design and Analysis*. Duxbury Press

Särndal, C-E, Swensson, B. & Wretman, J, (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.