

Stichprobenziehung

Nicola Döring, Jürgen Bortz

- 9.1 Population und Stichprobe – 292
- 9.2 Stichproben für qualitative Studien – 302
- 9.3 Stichproben für quantitative Studien – 305
- Literatur – 318

Lernziele

- Wissen, wie Population und Stichprobe definiert sind.
- Das Konzept der Repräsentativität erläutern und problematisieren können.
- Die wichtigsten qualitativen Stichprobenarten kennen.
- Die wichtigsten quantitativen Stichprobenarten kennen.
- Die Vor- und Nachteile probabilistischer und nicht-probabilistischer Stichproben abwägen können.

Die meisten qualitativen und quantitativen sozialwissenschaftlichen Studien untersuchen weder sämtliche interessierenden Personen, Objekten oder Verhaltensweisen noch einen singulären Einzelfall, sondern arbeiten mit Stichproben („samples“), d.h. mit einer Auswahl von Fällen. In der qualitativen Forschung spricht man im Zusammenhang mit der Bildung von Stichproben meist von **Fallauswahl**. In der quantitativen Forschung hat sich die Bezeichnung **Stichprobenziehung** eingebürgert. Im Folgenden wird zunächst das vor allem für quantitative Stichproben wichtige Verhältnis zwischen Population und Stichprobe behandelt, mit dem das Qualitätskriterium der Repräsentativität verbunden ist (► Abschn. 9.1). Anschließend stellen wir die wichtigsten Verfahren der Stichprobenbildung für qualitative Untersuchungen (► Abschn. 9.2) sowie für quantitative Studien (► Abschn. 9.3) vor. Während die „Theoretische Stichprobe“ den Prototyp einer qualitativen Fallauswahl und die „probabilistische Stichprobe“ den Prototyp eines quantitativen Samples darstellt, existieren eine Reihe von Verfahren der Fallauswahl bzw. Stichprobenziehung („sampling“), die sowohl in qualitativen als auch in quantitativen Studien zum Einsatz kommen (z. B. Gelegenheitsstichprobe, Schneeball-Stichprobe).

9.1 Population und Stichprobe

Die Gesamtheit aller Fälle, über die in einer Studie wissenschaftlich etwas ausgesagt werden soll, heißt in der quantitativen Forschung **Population** (auch: Grundgesamtheit, Universum; „population“). In der qualitativen Forschung wird eher vom (Untersuchungs-) **Feld** und den dort vorfindbaren **Fällen** gesprochen, um die Verortung der Fälle in ihrem lebensweltlichen Kontext zu betonen (zu wissenschaftstheoretischen Grundlagen der qualitativen Forschung ► Abschn. 2.3). In der empirischen Sozialforschung haben wir prinzipiell die Möglichkeit, alle interessierenden Fälle zu untersuchen (Vollerhebung) oder eine Stichprobe von Fällen auszuwählen (Teilerhebung). Stichprobenuntersuchungen sind der Regelfall,

wobei die Qualität von Stichproben hinsichtlich ihrer Aussagekraft über die Gesamtheit der interessierenden Fälle sehr unterschiedlich zu bewerten ist. Methodologische Überlegungen zum Verhältnis zwischen Population und Stichprobe und die Differenzierung verschiedener Stichprobentypen spielen in der quantitativen Forschung eine wichtige Rolle, während methodischen Fragen der Fallauswahl in der qualitativen Forschung deutlich weniger Aufmerksamkeit geschenkt wird (Glaser & Strauss, 1999, S. 238).

9.1.1 Vollerhebung der Population

Die **Zielpopulation** (auch: angestrebte Grundgesamtheit; „target population“) ergibt sich aus dem jeweiligen **Forschungsinteresse einer Studie**. Zielpopulationen können unterschiedlich definiert werden: die erwachsene Bevölkerung eines Landes, die Gesamtheit aller Besucher einer Veranstaltung, die Fünftklässler in Spanien, 70- bis 80-jährige Diabetespatienten in Deutschland etc. Neben einzelnen Personen können auch soziale Gruppen, Maßnahmen, Objekte oder Verhaltensweisen die Untersuchungseinheiten sein, so dass Stichproben dann aus entsprechenden Populationen zu ziehen sind (z. B. Population aller mittelständischen Unternehmen in Rumänien; Population aller Großfamilien einer Stadt; Population aller Zwischenrufe in Bundestagsdebatten im letzten Jahr; Population aller bisherigen Anti-Rauch-Kampagnen der Bundeszentrale für gesundheitliche Aufklärung [BZgA] etc.).

Population bzw. Zielpopulation – Unter Population bzw. Zielpopulation („population“; „target population“) versteht man die Gesamtmenge aller Untersuchungseinheiten (z. B. Personen, Verhaltensweisen, Objekte), über die im Rahmen einer quantitativen empirischen Studie wissenschaftliche Aussagen getroffen werden sollen.

Neben dem Erkenntnisinteresse müssen in der empirischen Forschung auch **pragmatische Überlegungen** hinsichtlich der Durchführbarkeit einer Untersuchung bei der Festlegung der Zielpopulation beachtet werden. Will man Belästigungen durch Fluglärm untersuchen, ist hierfür die Grundgesamtheit der Personen, die höchstens einen Kilometer vom Flughafen entfernt wohnen (dieses populationsdefinierende Kriterium ist problemlos messbar), einer Populationsdefinition vorzuziehen, die sich auf alle Personen bezieht, die sich durch Fluglärm beeinträchtigt fühlen. Bei der **Definition einer Zielpopulation** sind einfache manifeste Merkmale zu verwenden, die leicht operationalisierbar sind. Auch wird man Zielpopulationen meist zeitlich, örtlich und kulturell einschränken und nicht z. B. Aussagen über alle Menschen aller Altersgruppen zu

9.1 · Population und Stichprobe

allen historischen Zeiten in allen Ländern oder Kulturen anstreben.

Vollerhebung – Unter Vollerhebung (Totalerhebung, „total population survey“) versteht man die Untersuchung aller Objekte einer definierten Population.

Vollerhebungen sind **möglich**, wenn es sich um überschaubare und gut erreichbare Populationen handelt. Sie sind **notwendig**, wenn eine Population sehr klein und heterogen ist, so dass Stichproben das Gesamtbild verfehlen. Interessiert man sich z. B. für Arbeitsmotivation, Zufriedenheit und Verbesserungsvorschläge aller Mitarbeiterinnen und Mitarbeiter eines Unternehmens ($N = 22$, Population), so sollte eine Vollerhebung durchgeführt werden. Vollerhebungen sind nicht nur bei Populationen von Personen, sondern auch bei anderen Untersuchungsobjekten möglich und sinnvoll: So können in einer kommunikationswissenschaftlichen Analyse z. B. alle Folgen einer TV-Serie ausgewertet werden oder in der Verkehrsforschung alle Unfälle betrachtet werden, die sich während eines Jahres an einer bestimmten Straßenkreuzung ereignet haben.

Ein Sonderfall der Vollerhebung ist die **Volkszählung** (Zensus, „census“), bei der mit großem Aufwand für die **amtliche Statistik** Daten ganzer nationaler Bevölkerungen erhoben werden. Die Teilnahme an Volkszählungen ist in der Regel gesetzlich verpflichtend (s. z. B. für Deutschland das Zensusgesetz 2011; ■ Abb. 9.1).

Oftmals ist es jedoch **nicht** möglich, eine Vollerhebung durchzuführen. Dies ist immer dann der Fall, wenn

- die Population nicht endlich (finit), sondern **unendlich** (infinif) groß ist (Beispiel: Verbreitung nationaler Stereotype in allen Ausgaben der in Deutschland – täglich neu – erscheinenden Tageszeitungen),
- die Population **nur teilweise bekannt** ist (Beispiel: Erfassung des Gesundheitszustandes aller medikamentenabhängigen Männer in der Schweiz),
- die Art der Untersuchung die Population zu stark **beeinträchtigt oder gar zerstört** (Beispiel: „Crash-tests“ zur Qualitätskontrolle der gesamten Jahresproduktion eines Automobilherstellers) oder
- die Untersuchung der gesamten Population **zu aufwändig** wäre (Beispiel: Untersuchung zum Musikgeschmack bei allen europäischen Jugendlichen zwischen 14 und 16 Jahren).

Insbesondere in Beobachtungsstudien, in denen es um die direkte Erfassung von Verhaltensweisen geht (z. B. Flirtverhalten, Streitverhalten), sind Vollerhebungen des gesamten Verhaltensstroms zu allen Zeitpunkten und an allen Orten kaum möglich. Stattdessen müssen **Zeit- oder**

FIRST AND LAST DAY AS CENSUS TAKER



■ **Abb. 9.1** Die Teilnahme an Volkszählungen ist verpflichtend, die Interviewenden müssen dementsprechend sorgfältig ausgewählt und geschult werden. (© Graham Harrop/Search ID: ghrn508, Rights Available from CartoonStock.com)

Ereignisstichproben des interessierenden Verhaltens gezogen werden (► Abschn. 10.1.1). Relativ verbreitet sind Vollerhebungen z. B. in der Soziologie und Politikwissenschaft, wenn mit **Makrodaten** gearbeitet wird (z. B. Wahlbeteiligung in allen EU-Ländern; Bruttoinlandsprodukt in allen OECD-Ländern; Telefonanschlüsse in allen Ländern weltweit etc.).

Findet im Rahmen einer quantitativen Studie eine Vollerhebung statt, so sind bei der statistischen Auswertung **deskriptiv-statistische Angaben** meist ausreichend: Wenn vollständige Daten über die Population vorliegen, ist **kein** Inferenzschluss mithilfe eines statistischen Signifikanztests (► Abschn. 12.5.1) notwendig, um zu prüfen, ob die Stichprobendaten auf die Population zu verallgemeinern sind, denn es liegen ja keine Stichproben-, sondern direkt die Populationsdaten vor. Es wird jedoch diskutiert, ob unter bestimmten theoretischen Annahmen nicht doch die **Anwendung von statistischen Signifikanztests auf Vollerhebungen** sinnvoll ist (Behnke, 2005; Broscheid & Geschwend, 2005). Ein anderer kontroverser Punkt im Zusammenhang mit Populationen ist die Abgrenzung zwischen finiten (abzählbaren, endlichen) und **infiniten Populationen** und den daraus resultierenden Konsequenzen für Vollerhebungen, Teilerhebungen und statistische Analysen (Kozak, 2005).

9.1.2 Teilerhebung der Population: Stichprobenverfahren

Teilerhebungen bzw. Stichprobenuntersuchungen sind weniger aufwändig als Vollerhebungen. Sie lassen sich schneller durchführen und auswerten und sind deshalb aus forschungsökonomischen Gründen meist der Vollerhebung vorzuziehen. Für Stichprobenuntersuchungen spricht zudem die Möglichkeit, wegen der im Vergleich zur Vollerhebung deutlich reduzierten Anzahl von Untersuchungsfällen eine größere Anzahl von Merkmalen sorgfältiger und kontrollierter erfassen bzw. individuelle Sichtweisen differenzierter rekonstruieren zu können. Dies ist ein Umstand, der gelegentlich zu der Behauptung veranlasst hat, Vollerhebungen seien weniger aussagekräftig als gut geplante und sorgfältig durchgeführte Stichprobenuntersuchungen (Scheuch, 1974, S. 5).

Stichprobenerhebung – Eine Stichprobenerhebung (Teilerhebung, „sample“) liegt vor, wenn nur eine Auswahl von Fällen aus der Zielpopulation untersucht wird.

Jede Stichprobe zeichnet sich durch zwei zentrale Merkmale aus:

1. **Stichprobenart** („sampling method“): Nach welchem Auswahlverfahren Elemente aus der Population in die Stichprobe gelangen, hat wesentlichen Einfluss auf die Aussagekraft einer Studie. Grundsätzlich sind zufallsgesteuerte (**probabilistische**) Auswahlmethoden, bei denen die Stichprobenauswahl nach einem statistischen Zufallsverfahren erfolgt, von nicht-zufälligen (**nicht-probabilistischen**) Methoden zu unterscheiden. Probabilistische Auswahlverfahren sind nur bei großen Stichproben im Rahmen der quantitativen Forschung sinnvoll einsetzbar. Nicht-probabilistische Stichproben sind für die qualitative Forschung einschlägig, sie werden jedoch oft auch in der quantitativen Forschung eingesetzt. Bei nicht-zufälligen Stichproben differenziert man zwei Vorgehensweisen: Entweder es werden ganz gezielt bestimmte Elemente aus der Population herausgegriffen (**bewusste bzw. absichtsvolle Auswahl**; z. B. Quotenstichprobe: Befragung einer definierten Anzahl von Personen aus der Population, die gemäß Quotenplan bestimmte Merkmale oder Merkmalskombinationen aufweisen sollen) oder es werden relativ ungezielt vor allem gut erreichbare Elemente ausgewählt (**willkürliche Auswahl**; z. B. Selbstselektionsstichprobe: Befragung aller Personen, die sich auf eine veröffentlichte Umfrageeinladung melden).
2. **Stichprobenumfang** („sample size“): Der **Stichprobenumfang** ist definiert als Anzahl aller Objekte

oder Personen in der Stichprobe. In der quantitativen Forschung symbolisiert n oder N den Stichprobenumfang, N die Populationsgröße. Die Fallzahl eines qualitativen Samples wird ohne „ n “ berichtet. Der **Populationsumfang** wird mit N symbolisiert. Mit wachsendem Stichprobenumfang steigt die Aussagekraft einer Studie in der qualitativen wie in der quantitativen Forschung. Bei der Theoretischen Stichprobenbildung werden in der qualitativen Forschung im Zuge der Datenerhebung und Datenauswertung schrittweise zusätzliche Fälle hinzugezogen, bis **theoretische Sättigung** eintritt, also weitere Fälle keine neuen Erkenntnisse liefern (► Abschn. 9.2.1). Bei quantitativen Studien wird idealerweise im Vorfeld festgelegt, welche Effektgrößen im statistischen Signifikanztest noch nachweisbar sein sollen (hypothesenprüfende Studien) bzw. wie präzise Parameterschätzungen ausfallen sollen (populationsbeschreibende Studien) und auf dieser Basis ein **optimaler Stichprobenumfang** (► Abschn. 14.3) bestimmt. Zu kleine Stichproben würden in quantitativen Studien aufgrund mangelnder statistischer Teststärke zu uneindeutigen Resultaten führen und zu große Stichproben würden unnötig Ressourcen verbrauchen. Um die – mit dem Stichprobenumfang variierende – Genauigkeit oder Ungenauigkeit von statistischen Parameterschätzungen auszudrücken, spricht man in der Umfrageforschung vom **Stichprobenfehler** („sampling error“), der als statistisches Maß über den Standardfehler („margin of error“) quantifiziert wird (zu verschiedenen Arten von **Umfragefehlern gemäß dem Total Survey Error-Modell** ► Abschn. 10.2.3).

Betrachtet man das Verhältnis von quantitativer Stichprobe und Grundgesamtheit, so sind **drei verschiedene Populationskonzepte** zu differenzieren: Ziel-, Auswahl- und Inferenzpopulation. Idealerweise liegt eine vollständige Liste aller Elemente der (als finit definierten) **Zielpopulation** vor (**Auswahlgrundlage**, „sample frame“, „sampling frame“), aus der eine Auswahl von Elementen für die Stichprobe gezogen werden kann. In diesem Fall haben prinzipiell alle Elemente der Zielpopulation eine positive Auswahlwahrscheinlichkeit, somit sind Zielpopulation und **Auswahlpopulation** („frame population“) identisch. Beispiel: Ein Unternehmen möchte die Zufriedenheit seiner Kundschaft erfassen und nutzt zur Auswahl der Stichprobe das vollständige Kundenverzeichnis. Da jedoch nicht alle angeschriebenen Kunden an der Umfrage teilnehmen, können die Umfrageergebnisse nicht auf den gesamten Kundenstamm (Zielpopulation), sondern nur auf die auskunftswilligen und erreichbaren Kunden bezogen werden (**Inferenzpopulation**).

Auswahl- und Inferenzpopulation – Im Zusammenhang mit Stichprobenerhebungen müssen **drei Populationskonzepte** differenziert werden: Während die **Zielpopulation** („target population“) die Gesamtheit aller Objekte meint, über die Aussagen getroffen werden sollen, umfasst die **Auswahlpopulation** („frame population“) die Gesamtheit aller Objekte, die prinzipiell eine Chance haben, in die Stichprobe zu gelangen. Angesichts von Stichprobenausfällen erlaubt eine konkrete Stichprobe meist keinen direkten Rückschluss auf die Auswahlpopulation, sondern nur auf eine sog. **Inferenzpopulation** („inference population“), also die Gesamtheit aller Objekte, aus der die Stichprobe tatsächlich stammt.

Eine vollständige Zielpopulationsliste liegt leider nur in seltenen Fällen vor und kann bei infiniten Populationen definitionsgemäß gar nicht erstellt werden. Selbst wenn eine Liste vorliegt, ist sie z. B. aus Datenschutzgründen oft nicht für Forschungszwecke verfügbar. Meist muss deswegen mit **unvollständigen** bzw. **verzerrten Auswahlgrundlagen** gearbeitet werden. In einem solchen Fall weicht die Auswahlpopulation als Gesamtheit aller Objekte, die prinzipiell eine Chance haben, in die Stichprobe zu gelangen, von der Zielpopulation ab. Die Abweichung zwischen Zielpopulation und Auswahlpopulation bezeichnet man als **Abdeckungsfehler** („coverage error“). Man spricht von **Überabdeckung** („overcoverage“), wenn Elemente, die nicht zur Population gehören, im Auswahlrahmen erscheinen oder Populationselemente mehrfach im Auswahlrahmen auftauchen. Umgekehrt liegt ein Abdeckungsfehler im Sinne von **Unterabdeckung** („undercoverage“) vor, wenn Elemente im Auswahlrahmen fehlen, die zur Population gehören (► Studienbeispiel „Eine sehr große und sehr verzerrte Stichprobe“). Definiert man z. B. die Bevölkerung einer Stadt als Zielpopulation einer Befragungsstudie und verwendet für die Stichprobenziehung als Auswahlgrundlage das örtliche **Telefonbuch**, so entsteht ein beträchtlicher Unterabdeckungsfehler dadurch, dass alle Personen bzw. Haushalte, deren Nummern nicht im Telefonbuch veröffentlicht sind, überhaupt nicht in die Stichprobe gelangen können. Die entsprechende Studie kann somit keine Aussagen über die **Zielpopulation** (lokale Einwohner), sondern nur über die **Auswahlpopulation** (lokale Einwohner, deren Telefonnummern im Telefonbuch veröffentlicht sind) treffen (zu speziellen **Stichprobenverfahren für Telefonumfragen**, bei denen z. B. Telefonnummern nach dem Zufallsprinzip erzeugt werden ► Abschn. 10.2.3 „Vollstrukturiertes Telefoninterview“). Das Beispiel der Telefonumfrage illustriert, dass die Methode der Stichprobenziehung (z. B. mittels Telefonbuch oder künstlich generierten Telefonnummern) von der Methode der Datenerhebung (hier: Telefoninterview) abhängt. Bei der Datenerhebung mittels persönlichem Interview können zur Ziehung einer Stichprobe der lokalen Bevölkerung wiederum andere Stichprobentechniken zum Einsatz kommen: Aus dem **Melderegister des Einwohnermel-**

Studienbeispiel

Eine sehr große und sehr verzerrte Stichprobe

Legendär ist eine misslungene Stichprobenuntersuchung der **Zeitschrift *Literary Digest***, anhand derer das Wahlverhalten der amerikanischen Bevölkerung im Jahr 1936 prognostiziert werden sollte: 10 Mio. Wahlberechtigte in den USA, deren Adressen man über Telefonbücher, Mitgliedskarteien von Clubs und Vereinen etc. ermittelt hatte (willkürliche Auswahl), erhielten den Untersuchungsfragebogen postalisch zugeschickt. Eine **Stichprobe von rund $n = 2.40$ Mio. Wahlberechtigten** sendete ausgefüllte Fragebögen zurück. Diese Stichprobe sollte Auskunft geben über das politische Stimmungsbild in der **Population aller Wahlberechtigten in den USA**, also etwa $N = 100$ Mio. Wahlberechtigte. Die erhobenen Fragebogendaten legten den Schluss nahe, dass die Demokraten mit ihrem Spitzenkandidaten Franklin Roosevelt dem republikanischen Kandidaten Alfred Landon unterliegen würden und nur 43 % der Stimmen auf sich vereinigen könnten. Tatsächlich erreichte Roosevelt jedoch eine Stimmenmehrheit von 62 %. Eine der größten Stichproben in der Geschichte der Meinungsforschung führte so zu einer Fehlschätzung von knapp 20 % (Freedman, Pisani, & Purves, 1978, S. 302ff.).

Zwei Fehlerquellen waren im Spiel: Zunächst wurden durch die willkürliche (nicht-probabilistische) Anwerbung über Mitgliedskarteien und Telefonbücher (im Jahr 1936!) Angehörige der Mittel- und Oberschicht unverhältnismäßig häufig angesprochen im Sinne eines Überabdeckungsfehlers („oversampling“, „overcoverage“), während Angehörige der unteren Schichten eine sehr viel geringere Auswahlwahrscheinlichkeit hatten im Sinne eines Unterabdeckungsfehlers („undersampling“, „undercoverage“). Zum **Abdeckungsfehler** kam noch die für postalische Umfragen charakteristische **hohe Ausfallrate** („non-response“) hinzu, die wiederum die Angehörigen der unteren Schichten benachteiligte (sie antworteten seltener). Da aber gerade die unterprivilegierten Schichten in der Bevölkerung Roosevelts Politik befürworteten, konnte die stark in Richtung Oberschicht verzerrte Stichprobe das Wahlergebnis schlecht vorhersagen. Daran hätte sich auch nichts geändert, wenn man über dieselben Rekrutierungswege zusätzliche Personen befragt und die Stichprobengröße z. B. verdoppelt hätte. Merke: Bei einer systematisch verzerrten Auswahl hilft auch ein großer Stichprobenumfang nicht, den Fehler zu beheben. Er wiederholt sich nur in großem Stil.

Mithilfe einer sehr viel kleineren **Stichprobe von $n = 2000$ Wahlberechtigten**, die nach wichtigen soziodemografischen Merkmalen quotiert zusammengesetzt war (Quotenstichprobe) und somit höhere Repräsentativität erreichte, konnte das **Gallup-Institut** den Wahlerfolg von Roosevelt dagegen recht präzise vorhersagen (Diekmann, 2005, S. 325).

deamtes als Auswahlrahmen lässt sich eine Stichprobe von Wohnadressen für persönliche Interviews ziehen. Alternativ lässt sich eine Zufallsauswahl von Wohnadressen ziehen, indem man durch die Stadt geht und per vordefiniertem blindem Zufallsprinzip jeweils an verschiedenen Türen klingelt (sog. **Random-Route-Verfahren** der Stichprobenziehung), um die Interviews durchzuführen (zu **Stichprobentechniken für mündliche Interviews**

► Abschn. 10.2.3 „Vollstrukturiertes persönliches Interview“).

Werden aus einem Auswahlrahmen zufällig Elemente in die Stichprobe gezogen, so können nur dann Aussagen über die Auswahlpopulation getroffen werden, wenn tatsächlich alle für die Stichprobe vorgesehenen Personen oder Objekte tatsächlich untersucht werden. Dies ist jedoch oft nicht der Fall: Manche Personen sind z. B. für Befragungen nicht erreichbar oder verweigern die Teilnahme. Auch kann es durch technische Probleme oder Fehler bei der Datenerhebung oder Datenauswertung zu Informationsverlusten kommen. Die tatsächlich realisierte Stichprobe erlaubt also streng genommen keinen direkten Rückschluss auf die Auswahlpopulation, sondern nur auf eine sog. **Inferenzpopulation** als Gesamtheit aller Objekte, aus der die Stichprobe tatsächlich stammt (also **Auswahlpopulation minus Ausfälle**). Die Inferenzpopulation ist in der Regel nicht exakt bekannt. Sie kann nur erschlossen werden auf der Basis von Informationen oder Vermutungen über Art und Umfang der Stichprobenausfälle durch Nichterreichbarkeit, Teilnahmeverweigerung etc.

In der Umfrageforschung spricht man vom **Non-Response-Fehler** („non-response error“), um die Abweichung zwischen der laut Auswahlrahmen geplanten Stichprobe und der tatsächlich realisierten Stichprobe zu kennzeichnen (zu Umfragefehlern ► Abschn. 10.2.3). Wurde als Auswahlrahmen das Telefonbuch verwendet und daraus eine Zufallsauswahl von Anschlüssen ausgewählt, so wird man im Zuge der Datenerhebung feststellen, dass eine Reihe der Nummern nicht erreichbar sind oder dass die angerufenen Personen die Teilnahme an der Umfrage ablehnen. Die entsprechende Stichprobe trifft am Ende nur Aussagen über die Population der lokalen Einwohner, deren Telefonnummern im Telefonbuch veröffentlicht sind und die im Untersuchungszeitraum unter ihrer Telefonnummer erreichbar und zur Teilnahme an der Umfrage bereit waren. Inhaltlich ist dies so zu interpretieren, dass in der Inferenzpopulation z. B. alle Einwohner fehlen, die temporär nicht telefonisch erreichbar sind (z. B. wegen Krankenhaus-, Gefängnis- oder Auslandsaufenthalt) oder die aus verschiedenen Gründen keine Teilnahmebereitschaft haben (z. B. wenig Zeit, mangelndes Vertrauen in die Forschung, kein Interesse am Befragungsthema etc.; zu weiteren Details von Telefonumfragen ► Abschn. 10.2.3). Bei der Rekrutierung von Stichproben ist es ein wichtiges Anliegen, den Non-Response-Fehler zu verringern, etwa dadurch, dass man vorgesehene Stichprobenmitglieder **mehrfach kontaktiert**, wenn sie beim ersten Versuch nicht erreichbar sind. Zudem können die **Formulierung der Anfrage** und das **Auftreten der Forschenden beim Erstkontakt** darüber entscheiden, ob ein vorgesehene Stichprobenmitglied sich letztlich für oder gegen die Teilnahme an der Studie entscheidet und somit

in der geplanten Stichprobe verbleibt oder aus der Stichprobe herausfällt.

Das Ausmaß der Ausfälle wird über die **Ausschöpfungsrates** oder **Rücklaufquote** quantifiziert. Generell ist eine hohe Ausschöpfungsrates (bzw. ein geringer Non-Response) wünschenswert. Abgesehen vom quantitativen Ausmaß der Ausfälle ist es für die Repräsentativität der Stichprobe aber noch wichtiger, ob es sich um unsystematische Ausfälle handelt oder um **systematische Ausfälle**, die zu Stichprobenverzerrungen führen. Eine systematische Verzerrung würde z. B. vorliegen, wenn der Non-Response bei Berufstätigen in einer Telefonumfrage deutlich höher ist als z. B. bei Rentnern, weil die Anrufzeiten innerhalb üblicher Arbeitszeiten liegen.

Unterschiedliche **Befragungsmodi** (z. B. mündliches Interview, telefonisches Interview, Online-Befragung, mobile Befragung etc.) gehen bei der Stichprobenziehung mit unterschiedlichen Verzerrungen einher (spezifische Über- oder Unterabdeckungsfehler, unterschiedliche Anteile und Merkmale der Antwortverweigerer etc.). Eine Möglichkeit, die Verzerrungen eines bestimmten Befragungsmodus zumindest teilweise auszugleichen ist die **Multi-Mode-Befragung** („multi mode survey“), bei der mindestens zwei Befragungsmodi kombiniert werden, um eine weniger einseitig verzerrte Stichprobe zu erlangen (z. B. bei einer kombinierten Online- und Telefonbefragung gelangen sowohl Personen in die Stichprobe, die Online gut erreichbar sind, als auch diejenigen, die telefonisch gut erreichbar sind; zu unterschiedlichen Befragungsmodi in der quantitativen Forschung ► Abschn. 10.2.3 und ► Abschn. 10.3.3).

Bei der Festlegung von Auswahlmethode und Umfang der Stichprobe geht es einerseits um die Aussagekraft der geplanten Untersuchung, aber auch um forschungsökonomische Aspekte, etwa den Zeit- und Kostenaufwand. Im Zusammenhang mit Stichprobenziehung („sampling“) sind drei Phasen zu unterscheiden:

1. **Stichprobenkonstruktion:** Entscheidung für ein bestimmtes Verfahren der Stichprobenauswahl, Festlegung des Stichprobenumfangs und des Stichprobenplans (z. B. Entscheidung für eine bewusste Auswahl nach dem Quotenverfahren und Festlegung des Quotenplans).
2. **Rekrutierung:** Kontaktierung und Anwerbung von Untersuchungspersonen entsprechend der vorangegangenen Stichprobenkonstruktion (z. B. Versendung von personalisierten Einladungs-E-Mails; Kontaktaufnahme durch das Forschungsteam; Aufsuchen der Befragungspersonen an ihren Wohnadressen).
3. **Stichprobenerhebung:** Tatsächliche Datenerhebung mit allen rekrutierten Personen bzw. Objekten (z. B. Austeilen und Einsammeln der Fragebögen unter den ausgewählten Befragungspersonen). Die Anzahl der vollständig und plausibel ausgefüllten Fragebögen

(zur Datenaufbereitung ► Kap.11) entspricht dann der endgültigen Stichprobe.

Während bei der Stichprobenkonstruktion methodische Fragen hinsichtlich der Aussagekraft der zu ziehenden Stichprobe im Mittelpunkt stehen, sind im Zusammenhang mit der **Rekrutierung** vor allem **praktische** und **organisatorische Fragen** zu lösen, allen voran die Frage: Wie können möglichst schnell und kostengünstig genügend Untersuchungsteilnehmende aus der Zielpopulation gewonnen werden? Das Spektrum der Vorgehensweisen reicht von einer Anwerbung über den **eigenen Bekann-tenkreis** bis zum Rückgriff auf große Datenbanken von Personen, die sich freiwillig zur regelmäßigen Teilnahme an Studien bereit erklärt haben: Derartige **Access-Panels** werden vor allem von Forschungsinstituten aufgebaut und verwaltet. Die Panel-Mitglieder werden durch Incentives (z. B. Gutscheine, Verlosungen) zur langfristigen Teilnahme motiviert. Für alle Mitglieder eines Access-Panels liegen Kontaktdaten und soziodemografische Basisdaten vor. Bei Bedarf können aus diesem Freiwilligen-Pool jederzeit unterschiedliche Stichproben gezogen und z. B. für Fragebogenerhebungen genutzt werden. Bei Stichproben aus Access-Panels handelt es sich meist um nicht-probabilistische Quotenstichproben. Für probabilistische Stichproben greift man auf entsprechende **Auswahlrah-men** zurück, die möglichst viele Populationsmitglieder enthalten (z. B. Einwohnermeldeamtsregister, Kunden-kartei, Mitgliederverzeichnis).

Manche Populationen sind sehr schwer erreichbar, etwa weil sie klein, geografisch verstreut oder ihre Mitglieder nicht ohne weiteres erkennbar sind bzw. sich nicht öffentlich zu erkennen geben wollen. Typische Beispiele für solche **speziellen Populationen** sind Drogenkonsumenten, Prostituierte, Obdachlose, aber auch Anhänger eines seltenen Hobbys oder Experten mit seltenen Fachgebieten. Vollständige Populationslisten als Auswahlgrundlagen für probabilistische Stichproben wird man hier kaum erhalten. Statistische Angaben zur Zusammensetzung der jeweiligen Populationen fehlen meist ebenfalls, so dass Quotenstichproben nicht möglich sind. Es bleiben zunächst nur **Gelegenheitsstichproben** in der Variante **Selbstselektionsstichprobe**, etwa indem man öffentliche Teilnahme-Aufrufe startet (► Abschn.9.3.1). So lassen sich viele seltene Populationen heute über entsprechend spezialisierte Online-Foren ansprechen, in denen man Studieneinladungen veröffentlicht. Eine andere Möglichkeit besteht im **Screening**: Hierbei wird eine sehr große und heterogene Stichprobe gezogen und zunächst mit einer Screeningfrage bzw. einem Screening-Kurzfragebogen erfasst, wer zur gesuchten speziellen Population gehört, die untersucht werden soll. Es kann in diesem Fall auch auf spezielle **Stichprobenverfahren für schwer er-**

reichbare Populationen zurückgegriffen werden, etwa die **Schneeball-Stichprobe** (► Abschn.9.3.1).

9.1.3 Repräsentativität von Stichproben

Welche Stichprobenmerkmale anzustreben sind, hängt von der Zielsetzung einer Studie ab: Für qualitative wie quantitative Erkundungsstudien (**explorative Studien**) sind kleine, nicht-zufällige Stichproben ausreichend. Auch für theorieprüfende oder theoriebildende Untersuchungen (explanative Studien) greift man in der quantitativen sowie qualitativen Forschung in der Regel auf kleine bis mittlere, nicht-zufällige Samples zurück. Große probabilistische Stichproben sind für **populationsbeschreibende quantitative Studien** (deskriptive Studien), insbesondere für die quantitative **Umfrageforschung** („survey research“) nötig: Will man politische Einstellungen oder Merkmale des Gesundheitszustandes der breiten Bevölkerung möglichst genau erfassen, so sollte die Stichprobe möglichst unverzerrt die Populationsverhältnisse widerspiegeln.

Populationsbeschreibende quantitative Studien zielen darauf ab, die Ausprägung bestimmter Merkmale in der Population (**Populationsparameter**) möglichst präzise anhand der entsprechenden Merkmalsausprägungen in der Stichprobe (**Stichprobenkennwerte**) zu schätzen, entweder durch Berechnung eines einzelnen Schätzwertes (**Punktschätzung**) oder – das ist wesentlich aussagekräftiger – durch Angabe eines Schätzbereiches, in dem der wahre Populationsparameter mit hoher Wahrscheinlichkeit liegt (**Intervallschätzung**). Die Intervallschätzung gibt durch die Breite des **Konfidenzintervalls** für den geschätzten Populationsparameter an, wie genau oder ungenau die Schätzung ausfällt (zur Datenanalyse mittels Parameterschätzung im Rahmen deskriptiver Studien ► Abschn.12.4). Grundsätzlich können alle eine Population beschreibenden uni-, bi- oder multivariaten Parameter mittels Stichprobendaten geschätzt werden (z. B. arithmetischer Mittelwert, Medianwert, Modalwert, absolute Häufigkeit, relative Häufigkeit, Standardabweichung, Spannweite, Schiefe, Exzess, Regressions- und Korrelationskoeffizient, Kovarianz etc.; zur Bedeutung dieser Maße s. Bortz & Schuster, 2010; ■ Tab.9.1). Die konkrete Berechnung dieser statistischen Schätzungen sowie deren Genauigkeit hängen davon ab, welche Variante einer probabilistischen Stichprobe eingesetzt wird. So kann z. B. ein Populationsmittelwert in der Regel genauer geschätzt werden, wenn die Daten aus einer geschichteten Zufallsstichprobe anstelle einer einfachen Zufallsstichprobe (► Abschn.9.3.2) stammen. Denn die geschichtete Zufallsstichprobe bildet die Population hinsichtlich aus-

■ **Tabelle 9.1** Populationsbeschreibende quantitative Stichproben-Studien: Unbekannte Populationsparameter werden durch die empirischen Stichprobenkennwerte geschätzt (drei Beispiele)

	Populationsparameter (griechische Buchstaben)	Stichprobenkennwerte (lateinische Buchstaben)
Arithmetischer Mittelwert	μ (my) z. B. durchschnittliche TV-Nutzungszeit in der Population	$M = \bar{x}$ z. B. durchschnittliche TV-Nutzungszeit in der Stichprobe
Relative Häufigkeit	π (pi) z. B. Anteil der Personen mit Depression in der Population	p z. B. Anteil der Personen mit Depression in der Stichprobe
Bivariates lineares Zusammenhangsmaß	ρ (rho) z. B. linearer Zusammenhang (bivariater Korrelationskoeffizient) zwischen TV-Nutzungszeit und Internet-Nutzungszeit in der Population	r z. B. linearer Zusammenhang (bivariater Korrelationskoeffizient) zwischen TV-Nutzungszeit und Internet-Nutzungszeit in der Stichprobe

9

gewählter Schichtungsmerkmale (z. B. Einkommen) noch präziser ab als eine einfache Zufallsstichprobe.

Hat man sich anstelle einer Vollerhebung für eine **Teilerhebung** bzw. **Stichprobenerhebung** entschieden, so stellt sich – insbesondere bei populationsbeschreibenden Studien – die Frage, wie die Stichprobe zusammengesetzt werden soll, damit diese möglichst aussagekräftige Informationen über die Population liefert. Das Verhältnis zwischen Stichprobe und Population wird mit dem Qualitätsmerkmal der **Repräsentativität** („sample representativeness“) beschrieben, wobei es sich jedoch streng genommen **nicht** um einen statistischen Fachbegriff handelt (Schnell, Hill, & Esser, 2008, S. 305f.). Idealerweise stellt eine repräsentative Stichprobe ein exaktes **Miniaturationabbild der Population** dar: Die Merkmale und Effekte in der Population sollten sich im Kleinen 1:1 in der Stichprobe wiederfinden.

Repräsentativität einer Stichprobe – Die Repräsentativität einer Stichprobe („sample representativeness“) gibt an, wie gut bzw. unverzerrt die Merkmalszusammensetzung in der Stichprobe die Merkmalszusammensetzung in der Population widerspiegelt. Zu unterscheiden sind dabei merkmalspezifisch-repräsentative sowie global-repräsentative Stichproben.

Merkmalspezifisch-repräsentative Stichprobe

Angenommen, eine Hochschule möchte ihre Lehrangebote im fächerübergreifenden **Studium Generale** stärker den aktuellen Studierendeninteressen anpassen und zu diesem Zweck eine schriftliche Befragung durchführen, die mit möglichst geringem Zeit- und Kostenaufwand umgesetzt werden soll. Zielpopulation ist hierbei die **Gesamtheit aller Studierenden**, die zum aktuellen Zeitpunkt an der betreffenden Hochschule eingeschrieben sind (z. B. $N = 6\,294$ Studierende an der TU Ilmenau im WS 2010/2011). Die Hochschulleitung geht in diesem fiktiven Beispiel davon aus, dass die Themeninteressen

der Studierenden sich stark in Abhängigkeit vom Studienfach unterscheiden und strebt somit eine Stichprobe an, die hinsichtlich der Verteilung der Studienfächer der Population entspricht (merkmalspezifisch-repräsentative Stichprobe im Hinblick auf das **Merkmal Studienfach**). Zunächst wird also der Hochschulstatistik die Fächerverteilung in der Population entnommen (z. B. von den $N = 6\,294$ Studierenden in der Population sind 62 % in den Ingenieurwissenschaften, 13 % in Mathematik und Naturwissenschaften, 13 % in Rechts-, Wirtschafts- und Sozialwissenschaften und 12 % in Sprach- und Kulturwissenschaften eingeschrieben). Dieselbe Verteilung wird nun im Sinne eines Quotenverfahrens für die Stichprobe zugrunde gelegt, wobei der Stichprobenumfang aus forschungsökonomischen Gründen auf $n = 400$ Studierende begrenzt wird.

Merkmalspezifisch-repräsentative Stichprobe – Eine merkmalspezifisch-repräsentative Stichprobe liegt vor, wenn die Zusammensetzung der Stichprobe hinsichtlich einiger relevanter Merkmale der Populationszusammensetzung entspricht. Dies wird typischerweise durch eine nicht-probabilistische Quotenstichprobe erreicht.

Der **Quotenplan** für die Stichprobe von $n = 400$ sieht somit vor, 248 bzw. rund 250 Studierende aus den Ingenieurwissenschaften (ca. 62 %) und jeweils rund 50 Studierende (ca. 12–13 %) aus den anderen drei Fächergruppen zu befragen. Die entsprechende Anzahl an Paper-Pencil-Fragebögen wird in verschiedenen fachspezifischen Lehrveranstaltungen verteilt und nach dem Ausfüllen direkt wieder eingesammelt. Die merkmalspezifisch-repräsentative Stichprobe basiert also auf einem vergleichsweise unaufwändigen nicht-probabilistischen Quotenverfahren. Dieses schützt leider nicht vor deutlichen **Verzerrungen der Stichprobe**. So kann im beschriebenen Beispiel das Studierenden-Sample in verschiedener Hinsicht von der Population abweichen: Beispielsweise könnte die Stich-



■ **Abb. 9.2** Bei Verhaltensstichproben ist oft unklar, wie repräsentativ sie eigentlich sind. (Mod. nach Goscinny & Sempé, 1976. Mit freundlicher Genehmigung von Diogenes)

probe unbeabsichtigt einen viel geringeren Anteil an politisch interessierten Studierenden enthalten als die Population. Die Umfrage würde dann die Akzeptanz von und Nachfrage nach politikwissenschaftlichen Lehrangeboten im Studium Generale unterschätzen. Ob dies der Fall ist, weiß man nicht. Wird mit merkmalspezifisch-repräsentativen Stichproben gearbeitet, so ist es wichtig, die für die Untersuchungsfrage wichtigen Populationsmerkmale zu identifizieren, anhand derer dann die Quotierung vorgenommen werden soll. Zudem muss die Tatsache, dass eine merkmalspezifisch-repräsentative Stichprobe von zahlreichen Verzerrungen hinsichtlich nicht erfasster Merkmale betroffen sein kann, in die Ergebnisinterpretation einbezogen werden.

Global-repräsentative Stichprobe

Leider kann man nur selten ausschließen, dass neben den bekannten, mit dem untersuchten Merkmal (hier: Themeninteressen der Studierenden bzgl. Lehrangebot im Studium Generale) kovariierenden Merkmalen (im Beispiel hier: Studienfach der Studierenden) auch noch andere Variablen das untersuchte Merkmal beeinflussen. Besonders gravierend ist dieses Problem bei Untersuchungen, die ein neuartiges Produkt, eine technische Innovation oder bisher unerprobte Vorschriften und Richtlinien evaluieren oder bei Studien, in denen eine Population gleichzeitig bzgl. vieler, sehr unterschiedlicher Merkmale beschrieben werden soll (sog. **Omnibus-Untersuchungen** = Mehr-Themen-Untersuchungen). Da man nicht weiß, welche weiteren Variablen mit dem untersuchten Merkmal zusammenhängen oder da man – wie bei Omnibus-Untersuchungen – davon ausgehen muss, dass sich mangelnde Repräsentativität der Stichprobe auf die vielen untersuchten Merkmale in unterschiedlicher Weise auswirkt, wird man eine Stichprobe bevorzugen, die der Population in möglichst allen Merkmalen entspricht: eine

Stichprobe, die für die Population **global-repräsentativ** ist.

Global-repräsentative Stichprobe – Eine global-repräsentative Stichprobe liegt vor, wenn die Zusammensetzung der Stichprobe in allen Merkmalen und Merkmalskombinationen der Populationszusammensetzung entspricht. Dies kann nur durch probabilistische Stichprobenverfahren sichergestellt werden, sofern gleichzeitig ein Mindeststichprobenumfang eingehalten wird.

Globale Repräsentativität kann nur durch eine probabilistische Stichprobenziehung sichergestellt werden, bei der die Stichprobe nach einem blinden Zufallsverfahren zusammengestellt wird. Zudem muss die Stichprobe eine Mindestgröße haben, da nur bei entsprechend großen Stichproben die Zufallsauswahl greift (Gesetz der großen Zahl) bzw. eine präzise statistische Schätzung von Populationsparametern ebenfalls vom Stichprobenumfang abhängt. Betrachtet man den Spezialfall von bevölkerungsrepräsentativen Stichproben (Zielpopulation: Wohnbevölkerung in Deutschland), so haben global-repräsentative Stichproben, die auf probabilistischen Auswahlverfahren basieren, typischerweise einen Mindestumfang von rund 2 000 Befragten.

Vielen Laien ist unklar, was „Repräsentativität“ im stichprobentheoretischen Sinne bedeutet und sie glauben oft, dass große Stichproben per se Repräsentativität garantieren (■ Abb. 9.2). Immer wieder wird das Etikett „repräsentativ“ auch pauschal zum Gütemerkmal von Wissenschaft erklärt. Eine „repräsentative Studie“, so wird suggeriert, sei besonders „seriös“ und „aussagekräftig“. Dabei bezieht sich Repräsentativität lediglich auf das Verhältnis von Population und Stichprobe, ist somit ein **Qualitätsmerkmal allein der Stichprobe**. Auch eine Studie, die mit einer in diesem Sinne perfekt global-repräsentativen Stichprobe operiert, kann wissenschaftlich wertlos sein, etwa wenn die interessierenden Merkmale unzureichend operationalisiert wurden, so dass die

Messwerte fehlerhaft sind. Umgekehrt stellt eine nicht-repräsentative Stichprobe keineswegs pauschal den Wert einer empirischen Studie infrage. Explorative und explanative Studien – darunter fast die gesamte Experimentalforschung (zum Experiment ► Abschn. 7.6.1) – sind von vorne herein gar nicht darauf angelegt, Merkmalsausprägungen in der gesamten Bevölkerung oder in anderen großen Populationen möglichst genau zu schätzen und arbeiten so gut wie nie mit repräsentativen bzw. probabilistischen Stichproben. Einer empirischen Studie darf also nur dann gerechtfertigt die „mangelnde Repräsentativität“ ihrer Stichprobe(n) als Qualitätsmangel vorgeworfen werden, wenn sie als deskriptive Studie tatsächlich das primäre **Erkenntnisziel der Populationsbeschreibung** verfolgt (► Abschn. 7.5) und/oder wenn ihre Ergebnisse unzulässig auf die Allgemeinheit bezogen werden.

➤ **Die Repräsentativität ist ein Qualitätsmerkmal, das sich lediglich auf das Verhältnis zwischen Stichprobe und Population bezieht, nicht auf die wissenschaftliche Qualität der gesamten Studie. Global-repräsentative bzw. probabilistische Stichproben sind bei populationsbeschreibenden Studien notwendig, um die Ausprägung oder Verbreitung von Merkmalen in der Population präzise anhand der Stichprobendaten schätzen zu können.**

Die meisten empirischen Studien in der quantitativen empirischen Sozialforschung, die nicht auf Populationsbeschreibung, sondern z. B. auf Theoriebildung oder Hypothesenprüfung angelegt sind, arbeiten mit nicht-global-repräsentativen bzw. nicht-probabilistischen Stichproben und haben dennoch einen hohen wissenschaftlichen Wert.

Wann immer Stichproben für „repräsentativ“ erklärt werden, sollte man detailliert nach der Definition der Zielpopulation und dem Verfahren der Stichprobenauswahl fragen, um dann Möglichkeiten und Grenzen der Generalisierbarkeit der Stichprobenergebnisse auf die Populationsverhältnisse abzuwägen. Für derartige Beurteilungen ist in der Regel Sachkenntnis über den Untersuchungsgegenstand notwendig. Wenn z. B. die Lebenssituation von jungen Müttern anhand einer Stichprobe von Teilnehmerinnen von Mutter-Kind-Kursen untersucht wird, sind Aussagen über die Generalisierbarkeit der Befunde v. a. dann möglich, wenn man aus anderen Quellen zuverlässig weiß, inwiefern sich die Teilnehmerinnen von Mutter-Kind-Kursen von Müttern unterscheiden, die nicht an solchen Kursen teilnehmen (z. B. hinsichtlich Bildung, Familienstand, sozialer Integration oder Erwerbstätigkeit).

Dass eine Zufallssteuerung der Stichprobenauswahl aus der Population ab einem Mindeststichprobenumfang eher zu einer unverzerrten bzw. global-repräsentativen

Stichprobe führt als eine nicht-zufallsgesteuerte Auswahl, ist unstrittig. Unklarheit besteht aber oft dahingehend, ob und inwiefern echte **Zufallsstichproben eine notwendige Voraussetzung für statistische Signifikanztests** sind. In Statistik-Lehrbüchern wird zu den gängigen statistischen Signifikanztests in der Regel die „Zufallsstichprobe“ als eine der notwendigen Voraussetzungen zur sachgerechten Durchführung des Tests angegeben (Bortz & Schuster, 2010; Eid, Gollwitzer & Schmitt, 2013). Echte Zufallsstichproben sind jedoch, wie bereits angesprochen, in der akademischen Sozialforschung eher der Ausnahmefall. Ganze Forschungsfelder – wie etwa die psychologische Experimentalforschung – arbeiten nahezu ausschließlich mit nicht-probabilistischen Stichproben und werten ihre Daten dennoch mit parametrischen statistischen Signifikanztests (z. B. *t*-Test, Varianzanalyse, Regressionsanalyse) aus (Hager, 2013, S. 258).

Zu dieser Situation gibt es v. a. drei unterschiedliche Positionen:

1. **Pragmatische Position:** Bislang in der Forschungspraxis am häufigsten anzutreffen ist die – meist implizite (also nicht ausdrücklich diskutierte) – Annahme, die verwendeten nicht-probabilistischen Stichproben könnten bei der inferenzstatistischen Auswertung einfach **wie Zufallsstichproben behandelt** und somit mit herkömmlichen **parametrischen (verteilungsgebundenen) Signifikanztests** ausgewertet werden (Westermann, 2000). Diese Annahme ist natürlich problematisch. Zu ihrer Begründung wird zuweilen argumentiert, dass auch für eine Gelegenheitsstichprobe im Nachhinein eine (**fiktive**) **Inferenzpopulation** konstruiert werden kann, aus der die vorhandenen Fälle eine Zufallsauswahl bilden. Wird z. B. ein Experiment mit freiwilligen studentischen Versuchspersonen durchgeführt, die sich auf einen Teilnahmeaufruf gemeldet haben, so kann diese Gelegenheits- bzw. Selbstselektionsstichprobe (nicht-zufallsgesteuerte Auswahl aus der Population sämtlicher Studierenden) auch als Zufallsstichprobe aus der Inferenzpopulation speziell derjenigen Studierenden verstanden werden, die prinzipiell ebenfalls an genau diesem Experiment freiwillig teilnehmen würden. Ergebnisse dürften dann auch nur im Hinblick auf diese konstruierte Inferenzpopulation interpretiert werden.
2. **Mathematisch-statistische Position:** Aus mathematisch-statistischer Sicht wird dagegen betont, dass parametrische Signifikanztests nur dann sinnvoll sind, wenn tatsächlich eine zufallsgesteuerte Stichprobenauswahl erfolgt und die Daten somit als Ausprägungen von Zufallsvariablen gelten können. In Studien, in denen keine probabilistischen Stichproben vorliegen, dürfen dementsprechend parametrische Signifikanztests nicht eingesetzt werden, sondern statt-

dessen müssen **nicht-parametrische (verteilungs-freie) Signifikanztests** zum Einsatz kommen, etwa **Resampling-Verfahren** wie die **Randomisierungstests** (Bortz, Lienert, & Boehnke, 2008; Edgington & Onghena, 2007). Diese arbeiten nicht mit theoretischen Verteilungsannahmen über die Population, sondern bilden empirische Prüfverteilungen aus den vorliegenden Stichprobenergebnissen. Sie erlauben somit auch keinen Rückschluss auf eine Population, sondern nur eine kritische Bewertung des Stichprobenergebnisses. Manche Autoren fordern, dass die gängige statistische Auswertungspraxis angesichts überwiegend fehlender Zufallsstichproben weitgehend von parametrischen auf non-parametrische statistische Verfahren umgestellt wird.

3. **Erkenntnistheoretische Position:** Eine dritte, hier als „erkenntnistheoretisch“ etikettierte Position plädiert dafür, die Frage nach der Angemessenheit statistischer Auswertungsverfahren und der Interpretierbarkeit statistischer Befunde nicht zu stark auf mathematisch-statistische Aspekte und die Frage der Zufälligkeit von Stichproben zu verengen. Die Verwendung parametrischer Signifikanztests bei nicht-zufallsgesteuerten Stichproben lässt sich nämlich insofern gut rechtfertigen, als **parametrische Signifikanztests** als **Approximationen der nicht-parametrischen Tests** (Randomisierungstests) gelten können (Hager, 2013, S. 258). Das eigentliche Problem liegt aus Sicht der erkenntnistheoretischen Position nicht bei der Frage, ob Daten aus nicht-probabilistischen Stichproben nun parametrisch oder lieber non-parametrisch ausgewertet werden sollten (was eben in der Regel zu vergleichbaren Resultaten führt), sondern bei der viel grundlegenden Frage, inwiefern empirisch-statistische Ergebnisse (die sich auf **statistische Hypothesen** beziehen) begründet und nachvollziehbar der eigentlichen Zielsetzung hypothesenprüfender Studien dienen können. Diese Zielsetzung besteht gemäß der Wissenschaftstheorie des Kritischen Rationalismus darin, Aussagen über den **Bewährungsgrad von inhaltlichen bzw. substanziellen Hypothesen** und somit von **Theorien** zu treffen, die wiederum v.a. Aussagen über die Existenz von Ursache-Wirkungs-Relationen treffen (► Abschn. 2.2.4). Wenn also in einer psychologischen Experimentalstudie hinsichtlich der interessierenden abhängigen Variablen in hypothesenkonformer Richtung ein Mittelwertunterschied zwischen Experimental- und Kontrollgruppe gefunden wird, der sich im parametrischen *t*-Test (verstanden als Approximation an einen nicht-parametrischen Test) als statistisch signifikant erweist, so heißt dies nicht, dass dieser Effekt die **Höhe des Populationseffektes** schätzt (auf die genauen Populationsverhältnisse ist

kein Rückschluss möglich, wenn es sich um nicht-probabilistische Stichproben handelt). Stattdessen ist die statistisch signifikante Effektgröße als grober Hinweis auf den **Bewährungsgrad der substanziellen Hypothese bzw. Theorie** zu interpretieren, die den gefundenen Gruppenunterschied kausal vorhergesagt hatte (Hager, 2013, S. 261). Wenn sich der Effekt in weiteren Studien replizieren lässt, steigt der Bewährungsgrad der Theorie. Fragen der Stichprobenziehung und statistische Auswertung sowie deren Voraussetzungen sind also stärker in eine **erkenntnistheoretische Gesamtreflexion des Forschungsprozesses einzubetten** (ausführlicher dazu Hager, 2013).

Zur Notwendigkeit global-repräsentativer bzw. probabilistischer Stichproben in der quantitativen Sozialforschung bleibt festzuhalten: **Populationsbeschreibende Untersuchungen** erfüllen nur dann ihren Zweck, wenn man sich vor der Stichprobenziehung darüber Klarheit verschafft, über welche Population Aussagen formuliert werden sollen. Erst nachdem die Merkmale der Zielpopulation präzise festgelegt sind, erfolgt die Entwicklung eines Stichprobenplans, der die zufallsgesteuerte Ziehung einer **global-repräsentativen Stichprobe** gewährleistet, auf deren Basis die Ausprägungen von Populationsparametern geschätzt werden. In der populationsbeschreibenden Umfrageforschung wird jedoch auch verstärkt über die Einsatzmöglichkeiten nicht-probabilistischer Stichproben diskutiert (Baker et al., 2013), insbesondere vor dem Hintergrund der hohen Kosten und oft mangelnden Realisierbarkeit (z. B. wegen Nichterreichbarkeit oder Teilnahmeverweigerung) von echten Zufallsstichproben. Anstelle globaler Repräsentativität von Stichproben im Hinblick auf Populationen wird als angemesseneres Kriterium z. B. in der Markt- und Meinungsforschung u. a. das **Kriterium der Zweckmäßigkeit** („fit for purpose“) vorgeschlagen, d. h. es muss vorab definiert und dann im Forschungsprozess sichergestellt werden, dass Stichproben in der Weise gezogen (und ggf. noch gewichtet) werden, dass sie die Population in – hinsichtlich des Forschungsproblems – zweckmäßiger Weise abbilden (dabei kann es sich dann z. B. um lediglich merkmalspezifisch-repräsentative bzw. nicht-probabilistische Stichproben wie Quotenstichproben handeln).

Bei **gegenstandserkundenden und hypothesenprüfenden Studien** kann sinnvoll mit nicht-probabilistischen Stichproben gearbeitet werden, denn hier sollen nicht Populationsparameter möglichst präzise geschätzt, sondern Theorien über Ursache-Wirkungs-Verhältnisse gebildet und hinsichtlich ihres Bewährungsgrades geprüft werden. Effektgrößen sind hier als Indikatoren des Bewährungsgrades zu deuten. Die wichtigsten probabilistischen sowie nicht-probabilistischen Stichprobenverfahren für quanti-

tative Studien werden in ► Abschn. 9.3 noch genauer vorgestellt.

In der qualitativen Forschung wird ein grundlegend anderes Stichprobenverständnis zugrunde gelegt, bei dem zufallsgesteuerte Auswahl sowie Repräsentativität im Hinblick auf eine exakt definierte Zielpopulation keine Rolle spielen, sondern es darauf ankommt, durch die gezielte Auswahl von einigen wenigen Fällen die Bandbreite der Erscheinungsformen des untersuchten sozialen Sachverhaltes möglichst erschöpfend abzubilden (► Abschn. 9.2).

9.2 Stichproben für qualitative Studien

Qualitative Studien arbeiten meist mit relativ **kleinen Stichproben** im ein- bis zwei-, selten im dreistelligen Bereich. So liegt die untersuchte Fallzahl bei qualitativen Doktorarbeiten in den Sozialwissenschaften oft zwischen 20 und 30 (Mason, 2010; für eine Übersicht der Stichprobengestaltung der in der Zeitschrift *Health Education and Behavior* publizierten qualitativen gesundheitswissenschaftlichen Studien s. Safman & Sobal, 2004). Da die differenzierte interpretative Rekonstruktion der einzelnen Fälle im Kontext ihrer Lebenswelt angestrebt wird und damit ein hoher Arbeitsaufwand für jeden einzelnen Fall verbunden ist, wären größere Stichprobenumfänge bei qualitativen Beobachtungs-, Befragungs- oder Dokumentenanalyse-Studien aus forschungsökonomischen Gründen gar nicht zu bewältigen. Während bei großen Stichproben eine Auswahl nach dem statistischen Zufallsprinzip der beste Garant für globale Repräsentativität des Samples ist, greift dieser Mechanismus bei kleinen Stichproben nicht. Hier würde eine blinde Zufallsauswahl zu verzerrten und wenig aussagekräftigen Stichproben führen. Nicht zuletzt deswegen hat sich im qualitativen Forschungsansatz anstelle der Zufallsauswahl die nicht-zufällige Auswahl – und zwar speziell die **bewusste bzw. absichtsvolle Auswahl von Fällen** („purposive/purposeful sampling“) – als Stichprobenverfahren eingebürgert (Glaser & Strauss, 1999; Marshall, 1996). Damit ist gemeint, dass auf der Basis theoretischer und empirischer Vorkenntnisse gezielt solche Fälle in das Sample aufgenommen werden, die besonders aussagekräftig für die Fragestellung sind.

Wird keine Vollerhebung, sondern eine Teilerhebung durchgeführt, so sind im qualitativen Ansatz drei absichtsvolle Sampling-Strategien zu unterscheiden: Die Theoretische Stichprobe, die Fallauswahl gemäß einem qualitativen Stichprobenplan sowie die gezielte Auswahl bestimmter Arten von Fällen (Flick, 2010, S. 155ff.; Glaser & Strauss, 1999, S. 244ff.).

9.2.1 Theoretische Stichprobe

Das Verfahren der Theoretischen Stichprobenbildung („theoretical sampling“), bei dem am Ende eine sog. Theoretische Stichprobe („theoretical sample“) vorliegt, wurde von den Soziologen Barney Glaser und Anselm Strauss (1967, 1999) im Rahmen ihrer **Grounded-Theory-Methodologie** (GTM) entwickelt, einem umfassenden qualitativen Forschungsansatz, der auf eine gegenstandsverankerte Theoriebildung hinausläuft (zur Theoriebildung gemäß GTM ► Abschn. 6.3.5; zur Datenauswertung mittels Kodierung von vorgefundenen Dokumenten oder Interviewtranskripten gemäß GTM ► Abschn. 10.6.2 „Kodierung gemäß Grounded-Theory-Methodologie“). Die Besonderheit der Theoretischen Stichprobe besteht darin, dass die Fallauswahl ganz bewusst so gesteuert wird, dass ein **maximaler theoretischer Erkenntniswert** resultiert. Dabei erfolgt die Fallauswahl schrittweise im Zuge der Datenerhebung und Datenauswertung: Anhand der bisherigen Ergebnisse der Studie wird jeweils entschieden, welche *weiteren* Fälle (Personen, Gruppen, Institutionen, Ereignisse) *als nächstes* in das Sample aufzunehmen sind. Stichprobenauswahl, Datenerhebung und Datenanalyse werden nicht linear, sondern **mehrfach zirkulär abgearbeitet**, was der Logik des qualitativen Forschungsprozesses entspricht. Bei der Theoretischen Stichprobenbildung können verschiedene und wechselnde Kriterien zur bewussten Auswahl von Fällen herangezogen werden. Die Stichprobenziehung wird im Idealfall erst dann abgeschlossen, wenn auf der Basis der bereits erarbeiteten Zwischenergebnisse der Studie der Eindruck entstanden ist, dass weitere Fälle keinen neuen Informationsgehalt für die Theoriebildung versprechen (**theoretische Sättigung**). Aus forschungsökonomischen Gründen kann jedoch auch ein früheres Abschließen der Studie geboten sein, etwa wenn sich im Laufe der Erhebung herausstellt, dass viele weitere Kriterien für das untersuchte Phänomen wichtig sind, die notwendigen Ressourcen für eine entsprechend große Stichprobe jedoch nicht zur Verfügung stehen (Glaser & Strauss, 1999, S. 245). Damit ist die theoriebildende Aussagekraft der betreffenden Studie dann jedoch deutlich reduziert.

Die konkreten **Auswahlkriterien**, die bei der Theoretischen Stichprobenziehung zur Anwendung kommen, werden für jede einzelne Studie gemäß interessierendem (Untersuchungs-)Feld und Forschungsproblem individuell festgelegt. Allgemein lässt sich das Vorgehen so beschreiben:

- » Auswahlstrategien beschreiben Wege, um ein Feld zu erschließen. Dies kann – ausgehend von **Extremfällen, negativen, kritischen oder abweichenden Fällen** –

von seinen Rändern aus geschehen. Ein Feld kann auch von innen heraus erschlossen werden – ausgehend von **besonders typischen oder besonders entwickelten Fällen**. Es lässt sich von seiner vermuteten Struktur ausgehend erschließen – indem etwa möglichst **unterschiedliche Fälle** in ihrer Variationsbreite einbezogen werden. (Flick, 2010, S. 169f.; Hervorhebungen hinzugefügt; für qualitative Auswahlstrategien s. auch Patton, 1990)

Haben die Forschenden entschieden, welche Merkmale die als nächstes in das Sample aufzunehmenden Fälle (z. B. zwei weitere Fälle) haben sollen, so müssen die dafür passenden **Rekrutierungsstrategien** gewählt werden (z. B. Aushang in einer Institution, Vermittlung durch Kontaktpersonen im Feld, Aufruf in einem Internetforum, Anfrage im Bekanntenkreis etc.). Im Sinne der Wissenschaftlichkeit muss bei einer Theoretischen Stichprobe im Rahmen der jeweiligen Studie und ihres Forschungsproblems nachvollziehbar begründet werden, warum bei einer bestimmten Fallzahl schließlich von „theoretischer Sättigung“ ausgegangen wird.

In der Forschungspraxis wird das Prinzip der theoretischen Sättigung bislang offenbar **nicht immer konsequent umgesetzt**. So finden sich überzufällig viele qualitative Studien (z. B. unter Doktorarbeiten), die angeben, mit Theoretischen Stichproben zu arbeiten, deren Stichprobenumfänge aber **exakt bei 10, 20 oder 30 Fällen** liegen (Mason, 2010). Diese „glatten“ Zahlen deuten darauf hin, dass ein bestimmter, **überzeugend wirkender Stichprobenumfang** vorab festgelegt wurde und nicht das Prinzip der theoretischen Sättigung entscheidend war. Dieses würde nämlich häufiger auch zu ungeraden Stichprobenumfängen – etwa von 9, 27 oder 31 – führen. Für eine wissenschaftlich überzeugende Theoretische Stichprobenbildung ist es deswegen notwendig, die im Laufe des Forschungsprozesses erarbeiteten Auswahlkriterien für die jeweils nächsten Fälle inhaltlich zu begründen und insbesondere das Erreichen theoretischer Sättigung nachvollziehbar zu erläutern. Andernfalls droht die Gefahr, dass auch eine unreflektierte Auswahl beliebiger Fälle (also eine **Gelegenheitsstichprobe**; ► Abschn. 9.3.1) als „Theoretische Stichprobe“ ausgegeben wird.

9.2.2 Stichprobe gemäß qualitativem Stichprobenplan

Während bei der Theoretischen Stichprobenbildung die Stichprobe ergebnisoffen erst nach und nach auf der Basis theoretischer Überlegungen vervollständigt wird, arbeiten andere qualitative Stichproben mit einem vorab festgelegten **qualitativen Stichprobenplan** („qualitative

Studienbeispiel

Theoretische Stichprobe

Eine qualitative Studie aus dem Bereich der **Pflegeforschung** ging der Frage nach, inwiefern **Angehörige von Demenzpatienten** gegenüber deren Pflegeeinrichtungen eine aktive Fürspercherrolle einnehmen (Legault & Ducharme, 2009).

Da es vor allem die Töchter sind, die ihre alten, demenzkranken Eltern pflegen, konzentrierte sich die Studie auf weibliche Befragungspersonen. Die Studie beschränkte sich zudem auf ein konkretes kanadisches Pflegeheim, das zur Kooperation bereit war.

Im ersten Schritt sollten für die Stichprobe **3 Töchter** rekrutiert werden, deren demenzkranker Elternteil seit mehr als sechs Monaten und weniger als sechs Jahren in der Einrichtung lebte. Die Töchter wurden zunächst über das Pflegepersonal von der Studie informiert. Die teilnahmebereiten Töchter wurden dann von der Studienleitung telefonisch kontaktiert und zu einem Interviewtermin eingeladen.

Nach der Auswertung der ersten 3 Interviews wurden schrittweise weitere Untersuchungspersonen rekrutiert, dabei wurden dann – auf der Basis der sich entwickelnden Theorie – ganz gezielt Töchter mit unterschiedlichem **sozioökonomischem Status**, mit unterschiedlicher **Besuchshäufigkeit**, mit unterschiedlichem **Problemverhalten auf Seiten des demenzkranken Elternteils** und schließlich auch mit unterschiedlicher **Zufriedenheit mit der Pflegeeinrichtung** hinzugezogen – denn diese Merkmale wurden im Verlauf des Forschungsprozesses aus theoretischen Überlegungen im Hinblick auf die Fallauswahl für relevant erachtet. Dabei erwies es sich als besonders wichtig – und gleichzeitig besonders schwierig – Töchter in das Sample aufzunehmen, die mit dem Pflegeheim unzufrieden waren und nur selten zu Besuch kamen.

Theoretische Sättigung wurde aus Sicht der Forschenden erreicht, nachdem **14 Töchter** interviewt worden waren.

sampling plan“). Dieses Vorgehen ähnelt der Quotenstichprobe im quantitativen Forschungsansatz. Doch während bei der Quotenstichprobe (► Abschn. 9.3.1) mittels Quotenplan die quantitative Zusammensetzung der Population nachgebildet wird, soll gemäß qualitativem Stichprobenplan die Fallauswahl am Informationsgehalt der Fälle ausgerichtet sein. Der qualitative Stichprobenplan wird bewusst so zusammengestellt, dass möglichst alle für den untersuchten Sachverhalt besonders wichtigen Merkmale und Merkmalskombinationen im Sample vorkommen. Bei Personenstichproben wird der Stichprobenplan meist an **soziodemografischen Kriterien** orientiert (z. B. Alter, Geschlecht, Beruf, Nationalität, Studiengang, Krankheitsbild etc.). Üblicherweise werden in einem qualitativen Stichprobenplan ca. **3 Merkmale** berücksichtigt. Je nach Anzahl von deren Ausprägungen ergibt sich daraus bereits eine Kreuztabelle mit recht vielen Zellen; (► Tab. 9.2). Pro Zelle werden meist **1–3 Fälle** untersucht. Es ist auch möglich, einzelne Zellen unbesetzt zu lassen, etwa wenn die entsprechende Merkmalskombination in der Praxis un-

Studienbeispiel

Stichprobe gemäß qualitativem Stichprobenplan

In einer qualitativen **Befragungsstudie zur Wahrnehmung technischen Wandels** wurde ein **Stichprobenplan mit 3 Merkmalen** der Befragungspersonen zugrunde gelegt: 1.) Herkunftsland (Ost- und Westdeutschland sowie Frankreich), 2.) Geschlecht (weiblich, männlich) sowie 3.) Beruf (Informatiker, Sozialwissenschaftler, Lehrer; Flick, 1996; ■ Tab. 9.2). Somit ergibt sich anhand der Ausprägungen der drei nominalskalierten Merkmale im Stichprobenplan eine **Kreuztabelle mit $3 \times 2 \times 3 = 18$ Zellen**. Jede Zelle wurde mit 2 Fällen besetzt, so dass der **Gesamtstichprobenumfang bei 36 Befragungspersonen** liegt. Zur Umsetzung dieses Stichprobenplans müssen den Forschenden Rekrutierungsstrategien zur Verfügung stehen, um z. B. französische Informatikerinnen und Informatiker zu erreichen. Zudem müssen bei einer solchen internationalen Studie mögliche Sprachbarrieren überwunden werden (z. B. Einsatz französischsprachiger Interviewer), was normalerweise mit zusätzlichem Zeit- und Kostenaufwand verbunden ist (■ Tab. 9.2).

9

wichtig ist. Der **Gesamtstichprobenumfang** muss mit den vorhandenen Ressourcen (Zeit, Geld, Personal) zu bewältigen sein. Zudem müssen für die laut Stichprobenplan vorgesehenen Fälle passende **Rekrutierungsstrategien** existieren.

Im Rahmen einer **qualitativen Studie zur Krebsprävention** wurden 64 heterosexuelle nordamerikanische Männer in qualitativen Interviews zu ihrem Wissen und ihren Einstellungen gegenüber Prostatakrebs- und Darmkrebs-Vorsorgeuntersuchungen befragt (Winterich et al., 2009). Die Stichprobe wurde bewusst hinsichtlich Bildungsstand (gering, mittel, hoch), Hautfarbe (weiß, schwarz) sowie Wohnort (Land, Stadt) quotiert ($3 \times 2 \times 2 = 12$ Zellen), so dass Männer all dieser Gruppen im Sample etwa zu gleichen Teilen vertreten waren (ca. 5 Befragungspersonen pro Zelle). Die Rekrutierung vor Ort erfolgte jeweils über Kirchen, soziale Dienste, Männergruppen etc.

Der **Nachteil** der Vorabfestlegung der Samplestruktur besteht darin, dass auf diese Weise nur Unterschiede oder Gemeinsamkeiten zwischen den vorausgewählten Gruppen sichtbar gemacht werden können. Die Einschränkung der Vergleichsgruppen in der Stichprobe kann dazu führen, dass für die Theoriebildung ggf. wichtige Merkmalskombinationen, die in der Population existieren, übersehen werden. So könnte in der Beispielsstudie zur Technikwahrnehmung anstelle des Geschlechts möglicherweise das Lebensalter ein viel ausschlaggebenderer Faktor sein. Wenn verschiedene Altersgruppen jedoch bei der Vorabfestlegung der Samplestruktur nicht systematisch berücksichtigt wurden, können diese Effekte nach der Erhebung

gegebenenfalls nicht mehr untersucht werden (z. B. weil dann zu wenige ältere Personen im Sample vorhanden sind). Das Arbeiten mit qualitativen Stichprobenplänen ist vor allem dann indiziert, wenn über den Untersuchungsgegenstand schon **genügend Vorwissen** existiert, so dass die Auswahl der Merkmale für den Stichprobenplan gut fundiert erfolgen kann. Entsprechende Vorabauswahlen gemäß einem qualitativen Stichprobenplan sind nicht nur bei Personenstichproben zu treffen, sondern z. B. auch bei Feldbeobachtungen (Auswahl von Beobachtungsorten und -zeiten) oder bei Stichproben von vorgefundenen Dokumenten (z. B. Auswahl von Zeitschriften-Covern, Briefen, Akten oder Sitzungsprotokollen).

Bei qualitativen Studien mit größeren Stichproben kommen anstelle von qualitativen Stichprobenplänen teilweise auch **Quotenpläne** zum Einsatz (zur Quotenstichprobe ► Abschn. 9.3.1).

9.2.3 Gezielte Auswahl bestimmter Arten von Fällen

Das Theoretical Sampling sowie die Stichprobenziehung auf der Basis eines qualitativen Stichprobenplans laufen beide auf eine bewusst heterogene Stichprobenstruktur hinaus: Eine Vielfalt an Fällen soll berücksichtigt werden, um das interessierende Untersuchungsfeld breit abzubilden. Demgegenüber werden auch oft Fragestellungen untersucht, die sich auf eine ganz **spezielle Zielgruppe** beziehen. In diesen Studien wird dann meist eine **gezielte Auswahl bestimmter Arten von Fällen** eingesetzt. Dabei sind zwei Varianten gezielter Stichproben zu unterscheiden:

1. **Homogene gezielte Stichproben:** Die Untersuchungseinheiten aus der interessierenden Zielgruppe werden über einen einzigen oder über wenige **Rekrutierungswege** angesprochen und es wird ein **relativ kleines Sample** zusammengestellt. Beispiele: Eine qualitative Studie ging der offenen Forschungsfrage nach, welche psychologischen Folgen eine schwere Kopfverletzung für jugendliche Patienten und ihre Familien hat. Dazu wurde zu einem einzigen Zeitpunkt in einer einzelnen Klinik in den USA ein Sample von 12 entsprechenden Patienten rekrutiert (Berland & Thomas, 1991). Eine andere Studie beschäftigte sich mit der offenen Forschungsfrage, wie Überlebende des Atombombenabwurfs über Hiroshima und Nagasaki im Jahr 1945 die Ereignisse verarbeiteten. Dazu wurden im Jahr 2010 – also 65 Jahre später – insgesamt 8 Überlebende befragt, die mittlerweile in den USA ansässig waren (Knowles, 2011).
2. **Heterogene gezielte Stichproben:** Die Fälle aus dem interessierenden Feld werden breit gestreut über

■ **Tabelle 9.2** Beispiel einer Vorabfestlegung der Samplestruktur für eine qualitative Stichprobe. (Mod. nach Flick, 2010, S. 157. Aus: Uwe Flick, „Qualitative Sozialforschung. Eine Einführung“ Copyright © 1995, 2002 Rowohlt Taschenbuch Verlag GmbH, Reinbek bei Hamburg)

	Westdeutschland		Ostdeutschland		Frankreich		Summe
	weiblich	männlich	weiblich	männlich	weiblich	männlich	
Informatiker/in	2	2	2	2	2	2	12
Sozialwissenschaftler/in	2	2	2	2	2	2	12
Lehrer/in	2	2	2	2	2	2	12
Summe	6	6	6	6	6	6	36

unterschiedliche Rekrutierungswege angesprochen und es wird ein **relativ großes Sample** zusammengestellt. Auf diese Weise erhält man eine Auswahl von Fällen, die sich zwar alle hinsichtlich des Zielmerkmals ähneln, deren sonstige **interne Heterogenität** man jedoch im Vorfeld nicht genau kennt und auch nicht bewusst steuert, im Nachhinein dann jedoch im Zuge der Datenanalyse näher untersuchen kann (z. B. durch Typenbildung). Beispiel: Welche Erfahrungen lesbische Frauen mit dem Gesundheitssystem sammeln, wurde anhand einer qualitativen schriftlichen Befragung von 128 Respondentinnen erkundet, die sich selbst als lesbisch identifizierten und über unterschiedliche Onlinekanäle rekrutiert wurden (Bjorkman & Malterud, 2009). Die Fallzahl bei qualitativen Stichproben wird meist nicht mit „ n “ angegeben.

2. die **Quotenstichprobe**, die auf einer bewussten Auswahl gemäß einem auf die Populationszusammensetzung abgestimmten Quotenplan basiert, sowie
3. die **Stichprobenziehung mittels spezieller Verfahren**, die vor allem für schwer erreichbare Populationen entwickelt wurden.

Nicht-probabilistische (nicht zufallsgesteuerte) Stichprobe – Bei einer **nicht-probabilistischen (nicht zufallsgesteuerten) Stichprobe** erfolgt die Auswahl der Untersuchungsobjekte **willkürlich** und/oder **bewusst/absichtsvoll** durch die Forschenden. Bei nicht-probabilistischen Stichproben ist unbekannt, welche Auswahlwahrscheinlichkeit die einzelnen Elemente der Population haben bzw. in welcher Weise und in welchem Ausmaß die Stichprobenzusammensetzung gegenüber einer Zielpopulation verzerrt ist.

Gelegenheitsstichprobe

Die **willkürliche Auswahl** bzw. **Gelegenheitsstichprobe** („convenience sample“) ist ein in der quantitativen (und qualitativen) akademischen Sozialforschung sehr häufig anzutreffender Stichprobentyp. Denn Gelegenheitsstichproben sind definitionsgemäß mit dem geringsten Aufwand verbunden: Man bezieht einfach diejenigen Personen oder Objekte in die Studie ein, bei denen gerade eine **günstige Gelegenheit** besteht, sie ohne großen Aufwand zu untersuchen (Higginbottom, 2004). In der akademischen Forschungspraxis wird deswegen bekanntlich oft auf Studierende zurückgegriffen: Einen Fragebogen in der Vorlesung zu verteilen oder per Aushang im Institutsflur studentische Versuchspersonen für ein Experiment anzuwerben, ist die beste Gelegenheit, um schnell an Untersuchungspersonen zu kommen. Im Internet veröffentlichte Umfragen, die sich an beliebige Internetnutzende richten, oder Zeitungsanzeigen, mit denen Teilnehmende angeworben werden, haben ebenfalls den Charakter von Gelegenheitsstichproben. Auch qualitative Studien, bei denen willkürlich und ohne theoretische Reflexion einzelne Fälle untersucht werden, auf die man im Rahmen der Forschungstätigkeit im Feld gerade stößt, haben den Charakter einer Gelegenheitsstichprobe und stellen keine Theoretische Stichprobe dar (► Abschn. 9.2.1).

9.3 Stichproben für quantitative Studien

Stichprobenumfänge für quantitative Studien sind im Allgemeinen deutlich größer als die für qualitative Studien und liegen meist im drei- bis vierstelligen Bereich oder noch höher. In der Forschungspraxis werden häufig nicht-probabilistische Stichproben mit eingeschränkter Repräsentativität genutzt (► Abschn. 9.3.1). Sofern inhaltlich sinnvoll und forschungsökonomisch realisierbar, kommen natürlich auch probabilistische Stichproben zum Einsatz, die der beste Garant für globale Repräsentativität eines Samples sind (► Abschn. 9.3.2).

9.3.1 Nicht-probabilistische Stichproben

Drei Typen nicht-probabilistischer Stichproben spielen in der quantitativen Forschung eine wichtige Rolle:

1. die **Gelegenheitsstichprobe**, die auf einer willkürlichen Auswahl von Fällen basiert, die gerade verfügbar sind;

Gelegenheitsstichprobe oder Ad-hoc-Stichprobe – Bei einer Gelegenheitsstichprobe oder Ad-hoc-Stichprobe („convenience sample“, „ad hoc sample“, „accidental sample“, „haphazard sample“, „opportunistic sample“), manchmal auch: **anfallende/angefallene Stichprobe**, werden willkürlich Personen oder Objekte für die Stichprobe einer Untersuchung ausgewählt, die gerade zur Verfügung stehen oder leicht zugänglich sind (z. B. öffentliche Online-Befragung, Passantenbefragung).

Zuweilen werden willkürliche Auswahlen, bei denen nicht gezielt einzelne Personen angesprochen werden (z. B. durch personalisierte E-Mails), sondern ein **allgemeiner Teilnahmeaufruf** veröffentlicht wird (z. B. per Aushang, Presseanzeige, Newsletter oder Online-Banner) als **Selbstselektions-Stichproben** bezeichnet. Dahinter steht die Überlegung, dass bei nicht-personalisierten Einladungen zur Studienteilnahme überwiegend besonders motivierte Personen aus eigenem Antrieb teilnehmen (z. B. aufgrund ihres besonderen Interesses am Thema), während die persönliche Einladung durch ihren direkteren Aufforderungscharakter auch weniger motivierte Personen erreicht. Skeptische Personen können bei persönlicher Einladung ggf. zur Teilnahme überzeugt werden. Es ist jedoch zu beachten, dass eine gewisse Selbstselektion bei allen empirischen Studien vorliegt, da die Teilnahme aus forschungsethischen Gründen grundsätzlich freiwillig ist und es somit bei einem Teil der zu rekrutierenden Personen wegen fehlender Bereitschaft von vorne herein zu einer Teilnahmeverweigerung oder währenddessen zu einem Untersuchungsabbruch kommt (zur Bedeutung der Freiwilligkeit in der Forschungsethik ► Abschn. 4.1.1).

Eine Sonderform der **Selbstselektions-Stichprobe** stellt das **Access-Panel** dar. Ein Access-Panel besteht aus einer Menge an Personen, deren soziodemografische Merkmale und Kontaktdaten bekannt sind, und die sich aufgrund eines Teilnahmeaufrufs bereit erklärt haben, regelmäßig an wissenschaftlichen Studien teilzunehmen (Selbstselektion). Prinzipiell lassen sich durch aktive Rekrutierung auch probabilistische Panels zusammenstellen, in der Praxis sind aus ökonomischen Gründen aber oft nicht-probabilistische bzw. Selbstselektions-Panels anzutreffen. Markt- und Meinungsforschungsinstitute verfügen in der Regel über große, teilweise internationale Access-Panels, aus denen sie für aktuelle Studien jeweils Stichproben ziehen können. Access-Panels lassen sich im Rahmen der akademischen Forschung auch selbst aufbauen (z. B. Access-Panel befragungsbereiter Studierender). Access-Panels werden heute meist über Online-Plattformen verwaltet und müssen sorgfältig gepflegt werden (z. B. regelmäßige Kontaktierung und Incentivierung der Panel-Mitglieder = Panelisten), um Ausfälle (Panelmortalität) gering zu halten. Zudem müssen für ausgefallene Mitglieder regelmäßig neue Panelisten nachrekrutiert werden. Zum Umgang mit Access-Panels in der Markt-, Meinungs- und Sozialforschung existiert

die ÖNORM ISO 26362:2010. Es sind bereits einige **Access-Panel-Anbieter** in Österreich und Deutschland nach dieser Norm zertifiziert (u. a. GfK Austria, Respondi, marketagent). Aus einem Access-Panel lassen sich nicht nur willkürliche Auswahlen ziehen, sondern durch entsprechende Steuerung der Einladungen auch Quotenstichproben zusammenstellen oder sehr spezielle Zielgruppen rekrutieren (sofern das Panel groß genug ist und das betreffende Merkmal in der Datenbank vorliegt, z. B. Stichprobe von Diabetespatienten).

Die Aussagekraft von Gelegenheitsstichproben im Rahmen der **quantitativen Forschung** ist naturgemäß eng begrenzt. Ihnen liegen in der Regel **keine** klare Definition der Zielpopulation und auch kein definierter Auswahlrahmen zugrunde. Stattdessen kann nur **indirekt auf die Inferenzpopulation zurückgeschlossen** werden: Greift man etwa für Interviews in der Fußgängerzone die ersten $n = 50$ Passanten heraus, die Teilnahmebereitschaft zeigen, so repräsentieren diese die Grundgesamtheit aller Personen, die an diesem Tag an diesem Ort mittags durch die Fußgängerzone schlendern und zu einem Interview bereit sind. Keineswegs handelt es sich bei dieser willkürlichen Auswahl um eine Zufallsauswahl der lokalen Einwohner. Denn für eine Zufallsauswahl müssten alle Bürgerinnen und Bürger des Ortes die gleiche (oder zumindest eine bekannte) Chance haben, in das Sample zu gelangen. Das willkürliche – **im Laienverständnis vermeintlich „zufällige“** – Herausgreifen von Passanten schließt aber systematisch viele Einwohnerinnen und Einwohner aus (z. B. alle, die tagsüber zu Hause oder im Betrieb arbeiten, in der Schule oder am Ausbildungsplatz sind, es sehr eilig haben etc.). Leider werden Gelegenheitsstichproben oft fälschlich als „Zufallsstichproben“ bezeichnet. Wer für eine Studie, die nur auf einer Gelegenheitsstichprobe basiert, in Anspruch nimmt, eine „Zufallsstichprobe“ erhoben zu haben, begeht einen schwerwiegenden „Etikettenschwindel“. Denn die Gelegenheitsstichprobe weist die **niedrigste**, die Zufallsstichprobe dagegen die **höchste** Stichprobenqualität im Sinne globaler Repräsentativität auf.

Methodisch bestehen die Einschränkungen der Gelegenheitsstichprobe in der willkürlichen Gestaltung des „Auswahlrahmens“ (der „Auswahlrahmen“ liegt hier nicht als vollständige Liste der Elemente der Auswahlpopulation vor, sondern wird indirekt durch die Wahl der Untersuchungsumstände festgelegt) sowie in der willkürlichen Gestaltung des Auswahlprozesses. Beide Aspekte können durch entsprechende Vorüberlegungen verbessert werden, indem nicht nur die „beste“ und damit die forschungsökonomisch sparsamste Gelegenheit ins Auge gefasst wird, sondern auch eine Erhöhung der Aussagekraft der Stichprobe im Sinne einer Annäherung an die relevante Zielpopulation berücksichtigt wird. Handelt es

sich bei der Zielpopulation um die Einwohnerschaft eines Ortes, so ist eben zu überlegen, ob die Fußgängerzone zur Mittagszeit der beste „Auswahlrahmen“ ist, oder ob es vielleicht andere/ergänzende Rahmenbedingungen gibt, unter denen ein besserer Querschnitt der Bevölkerung des Ortes zu erreichen ist (z. B. Volksfest, Supermarkt). Zudem ist beim Herausgreifen der einzelnen Befragten sicherzustellen, dass keine grundlegende Verzerrung vorliegt (z. B. indem nur besonders freundlich und aufgeschlossen wirkende Personen, die alleine unterwegs sind, angesprochen werden). Hier kann die Festlegung eines „blinden“ Auswahlschemas (z. B. angesprochen wird jede 10. Person, die den Supermarkt verlässt, oder jede 5. Person, die im Restaurant sitzt) entsprechende Verzerrungen aufgrund des persönlichen Eindrucks vermeiden.

➤ Da man umgangssprachlich sagt, man habe „nach dem Zufallsprinzip“ einige Personen (z. B. Kurs Teilnehmer, Passanten etc.) befragt, werden *Ad-hoc-Stichproben* leider allzu oft fälschlich als „Zufallsstichproben“ bezeichnet. Bei Gelegenheitsstichproben handelt es sich jedoch um nicht-probabilistische Stichproben, es wird kein „blindes“ statistisches Zufallsverfahren auf eine vollständige Populationsliste (oder einen anderen geeigneten Auswahlrahmen) angewendet, sondern die Forschenden wählen aus forschungsökonomischen Gründen willkürlich bestimmte Erhebungsbedingungen aus und greifen unter diesen Bedingungen willkürlich Untersuchungseinheiten heraus, was mit diversen (und überwiegend unbekannten) Verzerrungen im Hinblick auf die Populationsverhältnisse verbunden ist.

In der Forschungspraxis fehlen oft die finanziellen, personellen und zeitlichen Ressourcen für aufwändige probabilistische Stichproben. Anstatt in dieser Situation gänzlich auf empirische Forschung zu verzichten und somit das Themengebiet allein der Spekulation zu überlassen, wird pragmatisch mit Gelegenheitsstichproben gearbeitet. Dahinter steht die Überzeugung, dass eine – wenn auch schwache – Datenbasis immer noch besser ist als keinerlei empirische Basis. Damit das Arbeiten mit Gelegenheitsstichproben wissenschaftlichen Kriterien entspricht, müssen die engen Grenzen ihrer Aussagekraft deutlich gemacht und Überinterpretationen strikt vermieden werden. Insbesondere dürfen Daten aus Gelegenheitsstichproben nicht zur Populationsbeschreibung herangezogen werden. Für Theoriebildung und Hypothesenprüfung können sie indessen nützlich sein. Manche Autoren lehnen jedoch alle Gelegenheitsstichproben als unwissenschaftlich und für die Sozialforschung untauglich ab (z. B. Schnell, Hill, & Esser, 2008, S. 298).

Quotenstichprobe

Bei der Quotenstichprobe („quota sample“) greift man ganz bewusst bzw. systematisch in den Auswahlprozess ein, so dass sich eine bestimmte Zusammensetzung der Stichprobe im Sinne **merkmalsspezifischer Repräsentativität** ergibt (► Abschn. 9.1.3). Typischerweise werden als Quotierungsmerkmale einige für das Untersuchungsthema besonders wichtige **soziodemografische Merkmale** herangezogen (z. B. Alter, Beruf, Bildungsstand, Geschlecht etc.). Indem man mittels Quotierungsverfahren z. B. eine Personenstichprobe so zusammenstellt, dass ihre Alters- und Bildungsverteilung derjenigen der Zielpopulation entspricht, erhofft man sich, dass auch die Stichprobenergebnisse hinsichtlich des Untersuchungsthemas (z. B. Mediennutzungsverhalten, Drogenkonsum) den Populationsverhältnissen besser entsprechen als z. B. bei einer Gelegenheitsstichprobe. Quotenstichproben sind nicht nur für Personenstichproben, sondern auch für andere Objektbereiche geeignet (z. B. Quotenstichprobe von Spielfilmen unterschiedlicher Filmgattungen und unterschiedlicher nationaler Herkunft).

Quotenverfahren – Die Stichprobenziehung nach dem **Quotenverfahren** („quota sampling“) setzt voraus, dass man die Populationszusammensetzung hinsichtlich einiger, für die Untersuchungsfrage wichtiger soziodemografischer Merkmale kennt. Diese Merkmalsverteilung wird anhand des **Quotenplans** („quota plan“) dann durch bewusste bzw. systematische Auswahl von **Personen oder Objekten** mit bestimmten Merkmalen (oder Merkmalskombinationen) in der Stichprobe nachgebildet. Die resultierende Quotenstichprobe („quota sample“) kann als nicht-probabilistische Stichprobe zwar keine globale Repräsentativität, durch den Quotenplan aber merkmalspezifische Repräsentativität (hinsichtlich der Quotierungsmerkmale) beanspruchen.

Voraussetzung für Quotenstichproben sind Kenntnisse darüber, a) welche soziodemografischen oder sonstigen einfachen manifesten Merkmale für das Untersuchungsthema besonders wichtig sind und b) wie die Ausprägungen dieser Merkmale in der Population verteilt sind. Wie man aus der Population aller Studierenden einer Hochschule eine Quotenstichprobe ziehen kann, wurde in ► Abschn. 9.1.3 „Merkmalsspezifisch-repräsentative Stichprobe“ bereits anhand eines einzelnen Quotierungsmerkmals, nämlich der Fächergruppe, dargestellt. Quotenstichproben beschränken sich meist nicht auf ein Merkmal, sondern nutzen mehrere Quotierungsmerkmale gleichzeitig. In dem genannten Beispiel, in dem es um das Interesse der Studierenden an unterschiedlichen Themen im Studium Generale geht (► Abschn. 9.1.3 „Merkmalsspezifisch-repräsentative Stichprobe“), könnte neben der Fächergruppe auch die Herkunft der Studierenden für die Stichprobenzusammensetzung wichtig sein. Innerhalb der einzelnen Fächergruppen sind deswegen noch anteilmäßig internationale Studierende zu berücksichti-

gen. Man spricht von **kombinierten Quoten**, wenn die Quotierungsvorgaben Merkmalskombinationen betreffen (z. B. genau 7 % der Stichprobenmitglieder sollen internationale Studierende sowie gleichzeitig Studierende der Ingenieurwissenschaften sein). Dementsprechend würden **unabhängige Quoten** (Randquoten) jeweils separate Vorgaben für die einzelnen Quotierungsmerkmale machen (z. B. insgesamt 62 % der Stichprobenmitglieder sollen Studierende der Ingenieurwissenschaften und insgesamt 10 % internationale Studierende sein – für die Merkmalskombination wird keine exakte Vorgabe gemacht). Kombinierte Quotenpläne bilden die Populationsverhältnisse besser ab, erhöhen aber den Rekrutierungsaufwand. Betrachtet man den kombinierten Quotenplan in ■ Tab. 9.3, so fällt auf, dass bei einem Stichprobenumfang von $n = 400$ Studierenden manche Zellen eine sehr geringe Besetzung aufweisen. Auf der Basis von nur 4 internationalen Studierenden der Fächergruppe Mathematik und Naturwissenschaften wird man kaum tragfähige Schlussfolgerungen über die spezifischen Themenwünsche dieser Gruppe ziehen können. Eine Vergrößerung der Zellbesetzung ist auf zwei Wegen möglich:

1. Man erhöht den **Stichprobenumfang**: Werden statt 400 insgesamt 1 200 Studierende befragt, steigt bei einer Quote von 1 % die Zellenbesetzung von 4 auf 12. Dies ist jedoch mit großem Zusatzaufwand verbunden.
2. Man arbeitet anstelle der in ■ Tab. 9.3 verwendeten kombinierten **proportionalen Quotenstichprobe** mit einer **nicht-proportionalen Quotenstichprobe**. Das bedeutet, dass man den prozentualen Anteil der internationalen Studierenden in der Stichprobe höher ansetzt als in der Population (z. B. 5 % statt 1 % in bestimmten Fächergruppen), um – bei konstantem Stichprobenumfang von $n = 400$ – genügend Befragte aus dieser Gruppe im Sample zu haben. Vor der quantitativen Datenanalyse kann eine nicht-proportionale Quotenstichprobe im Zuge der Datenaufbereitung durch Gewichtungungsverfahren wieder den Populationsverhältnissen angeglichen werden (► Abschn. 11.4.4).

In der **Online-Forschung** werden Quotenstichproben neuerdings als sog. **River Samples** erstellt (DMS Research, 2008): Dazu werden Umfrageeinladungen möglichst breit im Internet gestreut. Aus dem ständigen **Zustrom an neuen Internet-Nutzenden**, die auf die Studieneinladungen stoßen (deswegen River Sampling, River = Fluss, Strom), entscheiden sich einige für die Studienteilnahme. Würde man alle Freiwilligen nun im Rahmen einer einzigen Studie untersuchen, wäre das eine einfache Gelegenheits- oder Selbstselektions-Stichprobe. Stattdessen werden beim River-Sampling die Teilnehmenden aber zunächst nach ihren soziodemografischen Angaben ge-

fragt und auf dieser Basis vom Umfrage-Server gemäß entsprechende Quoten-Plänen verschiedenen Studien zugeordnet, die dann mit Quoten-Stichproben arbeiten. Im Unterschied zur Nutzung von **Quoten-Stichproben aus Online-Access-Panels** (die aus selbst-selektierten dauerhaften Panelisten bestehen), sollen River-Panels durch die ständig „frischen“ Befragten aussagekräftiger sein. Systematische Methodenstudien, die die verschiedenen Stichprobentechniken vergleichen, stehen jedoch noch aus.

Stichprobenverfahren für spezielle Populationen

Um im Rahmen quantitativer (oder auch qualitativer) Studien ganz spezielle Populationen zu untersuchen, bei denen bereits das Auffinden einzelner Populationsmitglieder schwierig sein kann, wird zur Rekrutierung und Zusammenstellung der Stichprobe am häufigsten das sog. **Schneeballverfahren** eingesetzt, bei dem die sozialen Netzwerke der Populationsmitglieder genutzt werden („snowball sampling“; Goodman, 1961). Eine Weiterentwicklung des Schneeballverfahrens in Richtung erhöhter Repräsentativität stellt das **„Respondent Driven Sampling“** (RDS) dar. Für schwer erreichbare Populationen, deren Mitglieder untereinander nicht besonders eng vernetzt sind, bietet sich die Methode des **„Targeted Samplings“** an, bei der man bevorzugte Aufenthalts- oder Wohnorte der Mitglieder der Zielpopulation identifiziert und dann systematisch vor Ort rekrutiert.

Schneeballverfahren – Die Stichprobenziehung nach dem **Schneeballverfahren** („snowball sampling“, „chain sampling“, „chain-referral sampling“, „referral sampling“) ist geeignet für Populationen, die für die Forschenden schwer erreichbar, deren Mitglieder untereinander jedoch gut vernetzt sind. Hierbei werden einzelne Mitglieder aus der Population gebeten, über ihre persönlichen sozialen Netzwerke weitere Untersuchungspersonen zu rekrutieren. Die **Schneeballstichprobe** („snowball sample“) kann als nicht-probabilistische Stichprobe keine Repräsentativität beanspruchen.

Beim Schneeballverfahren wird auf die **persönlichen sozialen Netzwerke** einzelner Populationsmitglieder zurückgegriffen, die von den Forschenden gebeten werden, weitere Untersuchungspersonen zu benennen bzw. zu rekrutieren. Je nach Datenerhebungsmethode werden im Schneeballverfahren entweder zunächst nur Kontaktdaten gesammelt (z. B. für spätere Interviews) oder es wird gleich ein Fragebogeninstrument weitergegeben. Ein Schneeballverfahren kann einstufig erfolgen, indem eine Gruppe von Populationsmitgliedern gebeten wird, im eigenen Bekanntenkreis einmalig eine bestimmte Anzahl an Untersuchungspersonen zu rekrutieren. Es ist aber auch möglich, dass die über die Kontaktpersonen angeworbenen Untersuchungspersonen ihrerseits gebeten werden, weitere Populationsmitglie-

■ **Tabelle 9.3** Zusammensetzung der Population der Studierenden der TU Ilmenau im WS 2010/2011 laut Hochschulstatistik sowie Quotenplan für eine Stichprobe im Umfang von $n = 400$ nach den Quotierungsmerkmalen Fächergruppe und Herkunft der Studierenden

Zusammensetzung der Population ($N = 6\,439$)				Kombinierter proportionaler Quotenplan ($n = 400$)	
Fächergruppe	Anteil	Herkunft	Anteil	Anzahl	Anzahl
	[%]		[%]		
Ingenieurwissenschaften	62	DE	55	248	220
		INT	7		28
Mathematik und Naturwissenschaften	13	DE	12	52	48
		INT	1		4
Rechts-, Wirtschafts- und Sozialwissenschaften	13	DE	12	52	48
		INT	1		4
Sprach- und Kulturwissenschaften	12	DE	11	48	44
		INT	1		4
Gesamt	100		100	400	400

DE Deutschland; INT International

der anzuwerben. Auf diese Weise wächst die Zahl der Untersuchungsteilnehmenden wie ein **rollender Schneeball**.

Beim Schneeballverfahren kann es in Abhängigkeit von der Auswahl der ersten Kontaktpersonen und ihrer Netzwerkwerkzusammensetzung zu zahlreichen Verzerrungen kommen. Die rekrutierenden Populationsmitglieder müssen möglichst gut instruiert werden, welche und wie viele Personen sie auswählen (hier können bei Bedarf auch Quoten vorgegeben werden), was sie über das Forschungsprojekt mitteilen und wie sie den Datenschutz sichern sollen. Wenn die Schneeballrekrutierung in einer kleinen Population unprofessionell abläuft, kann dies zu Misstrauen, Widerstand oder gar öffentlichem Protest (z. B. kritische Internet- und Pressebeiträge über die Studie) führen.

Studien, die auf Schneeballstichproben basieren, untersuchten z. B.

- britische Bürger/innen, die gegen Mobilfunkmasten politisch protestieren (Drake, 2011),
- israelische Computerhacker/innen (Turgeman-Goldschmidt, 2005),
- afroamerikanische Crack-Konsumentinnen mit Kindern (Lam et al., 2007) oder
- kenianische Prostituierte, die unter Alkoholabhängigkeit leiden (Chersich et al., 2007).

Bei der Schneeballstichprobe handelt es sich um eine willkürliche, nicht-probabilistische Stichprobe, deren Repräsentativität für die Zielpopulation unbekannt ist.

Es gibt jedoch eine Weiterentwicklung des Verfahrens, mit deren Hilfe die Repräsentativität deutlich gesteigert werden kann: Der als **Respondent Driven Sampling** (RDS) eingeführte Ansatz arbeitet mit einem bestimmten Incentivierungsverfahren, dokumentiert den Rekrutierungsprozess detailliert und nutzt ein mathematisches Modell zur Gewichtung der rekrutierten Personen (Heckathorn, 1997; <http://www.respondentdrivensampling.org>). Bestandteil des RDS-Ansatzes ist es, dass die rekrutierten Personen jeweils immer nur eine begrenzte Zahl weiterer Personen rekrutieren dürfen und für jede tatsächlich rekrutierte Person incentiviert werden. Zudem wird erfragt, wie viele Personen aus der Zielpopulation ein Stichprobenmitglied jeweils kennt, um Auswahlwahrscheinlichkeiten zu schätzen. Vorliegende Studien mit Stichprobenziehung gemäß RDS beziehen sich z. B. auf spezielle Populationen wie Drogenkonsumenten oder Jazzmusiker. Der Ansatz lässt sich auch im Rahmen der Online-Forschung einsetzen (Wejnert & Heckathorn, 2008).

Respondent Driven Sampling – Die Stichprobenziehung nach der Methode des Respondent Driven Sampling (RDS) stellt eine Weiterent-

wicklung der Schneeballstichprobe dar und ist wie diese geeignet für Populationen, die für die Forschenden schwer erreichbar, deren Mitglieder untereinander jedoch gut vernetzt sind. Das Respondent Driven Sampling ist wesentlich aufwändiger als das Schneeballverfahren, garantiert jedoch durch eine bessere Kontrolle der Rekrutierungswege eine höhere Repräsentativität des Samples.

Der **Targeted-Sampling-Ansatz** (Watters & Biernacki, 1989) ist indiziert, wenn Stichproben aus schwer erreichbaren Populationen gezogen werden sollen, deren Mitglieder nicht stark untereinander vernetzt sind. Um z. B. eine Stichprobe von Drogenkonsumenten zu ziehen, werden gemäß Target Sampling in einer oder mehreren Städten zunächst typische Aufenthaltsorte der entsprechenden Drogenszene identifiziert und aufgelistet (z. B. durch eigene Beobachtung, Hinweise von Szenemitgliedern, Auskünfte der Polizei etc.). Dann wird ein Stichprobenplan entworfen, dem zu entnehmen ist, wie viele Personen pro Treffpunkt zu rekrutieren sind (Peterson et al., 2008). Um eine Stichprobe von Paaren zu ziehen, in denen es zu häuslicher Gewalt kommt, wurden gemäß dem Ansatz des „Targeted Neighborhood Sampling“ zunächst Stadtviertel identifiziert, aus denen besonders häufig Notrufe wegen häuslicher Gewalt bei der Polizei eingehen. In den betreffenden Vierteln wurden dann Flyer in den Briefkästen verteilt, um Betroffene zu rekrutieren (Christopher, Pflieger, Canary, Guerrero, & Holtzworth-Munroe, 2008).

Targeted Sampling – Die Stichprobenziehung nach der Methode des Targeted Sampling ist besonders geeignet für schwer erreichbare Populationen, deren Mitglieder untereinander nicht besonders eng vernetzt sind. Sie stellt eine ortsbasierte Alternative zu netzwerkbasierten Verfahren wie dem Schneeballverfahren und dem Respondent Driven Sampling dar.

9.3.2 Probabilistische Stichproben

Probabilistische Stichproben basieren auf einem **Auswahlrahmen** („sampling frame“) und einem darauf angewendeten statistischen Zufallsverfahren. Sie sind aufwändig und teuer, aber insbesondere für populationsbeschreibende Untersuchungen unerlässlich, in denen es auf globale Repräsentativität der Stichprobe ankommt. Zufallsgesteuerte Stichproben werden am häufigsten im Rahmen der Meinungs- und Wahlforschung, aber auch in der Gesundheitsforschung sowie der Politikwissenschaft und Soziologie eingesetzt, wo es eben um die präzise Feststellung von Merkmalsausprägungen in großen Grundgesamtheiten wie etwa der Wohnbevölkerung ganzer Länder geht. Man unterscheidet vier Arten der probabilistischen Stichprobe: die **einfache Zufallsstichprobe**, die **geschichtete Zufallsstichprobe**, die **Klumpenstichprobe** und die **mehrstufige Stichprobe**.

Der **Prozess der probabilistischen Stichprobenziehung** muss zwingend mit einer präzisen Definition der Zielpopulation („target population“) sowie der Spezifizierung des Auswahlrahmens beginnen (► Abschn. 9.1.2). Dann folgt die Stichprobenkonstruktion im Sinne der Wahl der Auswahlmethode („sampling method“), der Festsetzung des Stichprobenumfangs („sample size“) und Festlegung des Stichprobenplans („sampling plan“). Dabei sind Informationen über die Zusammensetzung der Zielpopulation zu berücksichtigen, aber auch forschungsökonomische Aspekte im Blick zu behalten. Nach oder gleichzeitig mit der Rekrutierung der Untersuchungspersonen oder Untersuchungsobjekte gemäß Stichprobenplan erfolgt die Datenerhebung und damit die eigentliche praktische Stichprobenziehung.

Einfache Zufallsstichprobe

Die Ziehung einer **einfachen Zufallsstichprobe** („simple random sample“) setzt voraus, dass jedes zur finiten Population gehörende Untersuchungsobjekt einzeln identifizierbar ist bzw. dass als Auswahlrahmen eine vollständige Populationsliste vorliegt. So ist das Einwohnermelderegister ein vollständiger Auswahlrahmen für die lokale Wohnbevölkerung. Ebenso stellt das Immatrikulationsverzeichnis einen vollständigen Auswahlrahmen der Zielpopulation aller Studierenden einer Hochschule dar. Für die Qualität der Stichprobe ist es von Bedeutung, dass die Entscheidung darüber, welche Untersuchungsobjekte zur Stichprobe gehören und welche nicht, ausschließlich vom Zufall abhängt. Besteht eine finite Population aus N Objekten und sollen sich in der Stichprobe n Objekte befinden, können gemäß dem Binomialkoeffizienten insgesamt (Gl. 9.1)

$$C = \binom{N}{n} = \frac{N!}{n! \cdot (N - n)!} \quad (9.1)$$

verschiedene Stichproben bzw. Kombinationen (C) gezogen werden (der mittlere Ausdruck wird gelesen als „N über n“; Erläuterung dieser Formel durch ► Chancengleichheit für Skatspieler – Ziehung einer einfachen Zufallsstichprobe). Die Menge aller zu einer Population gehörenden Stichproben nennt man **Stichprobenraum** („sample space“). Setzen wir voraus, dass die Wahrscheinlichkeit, in die Stichprobe aufgenommen zu werden, für jedes Untersuchungsobjekt gleich ist („equal probability selection method“, EPSEM), hat jede der C verschiedenen Stichproben die gleiche Auswahlwahrscheinlichkeit. Derartige Stichproben werden **Zufallsstichproben** genannt.

Idealerweise geht man bei der Entnahme einer einfachen Zufallsstichprobe wie folgt vor: Die gesamte Populationsliste bzw. der vollständige Auswahlrahmen wird von 1 bis N durchnummeriert. Mithilfe von **Zufallszahlen**

Chancengleichheit für Skatspieler – Ziehung einer einfachen Zufallsstichprobe

$N = 5$ Skatspieler (wir nennen sie einfachheitshalber A, B, C, D und E) treffen sich in einem Lokal und wollen Skat spielen. Da für eine Skatrunde jedoch nur $n = 3$ Spieler benötigt werden, müssen zwei Personen zusehen. Man einigt sich darauf, die drei aktiven Spieler auszulosen. Alle fünf Anwesenden schreiben ihren Namen auf einen Zettel und werfen den zusammengefalteten Zettel in ein leeres Bierglas (in der Statistik wird hierfür – zumindest symbolisch – eine Urne verwendet). Nach gründlichem Durchmischen werden nacheinander die Zettel B, E und D gezogen. Die Skatrunde steht fest. Zunächst einmal ist es einleuchtend, dass sich an dieser Runde nichts geändert hätte, wenn die gleichen Zettel in einer anderen Reihenfolge, etwa E, D und B gezogen worden wären. Alle möglichen $3! = 3 \cdot 2 \cdot 1 = 6$ verschiedenen Reihenfolgen (BDE, BED, DBE, DEB, EBD und EDB) bilden die gleiche Stichprobe. Insgesamt hätten sich zehn verschiedene Skatrunden bilden können.

$$\binom{N}{n} = \frac{N!}{n! \cdot (N-n)!} = \frac{5!}{3! \cdot (5-3)!} \\ = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1) \cdot (2 \cdot 1)} = 10$$

Ein Spieler fragt sich nun, ob dieses Verfahren gerecht sei, ob jeder der fünf anwesenden Skatspieler tatsächlich die gleiche Chance hatte, in die Runde aufgenommen zu werden. Er argumentiert in folgender Weise: Bei der ersten Zettelentnahme besteht für jeden der fünf Spieler eine Auswahlwahrscheinlichkeit von $1/5$. Steht aber der 1. Spieler fest, erhöht sich für die verbleibenden Spieler bei der 2. Zettelentnahme die Auswahlwahrscheinlichkeit auf $1/4$ und für die 3. Zettelentnahme auf $1/3$. Also hat offensichtlich nicht jeder Spieler die gleiche Chance, in die Skatrunde aufgenommen zu werden.

Diese Argumentation ist unvollständig. Es wurde übersehen, dass es sich bei den Auswahlwahrscheinlichkeiten der 2. und 3. Ziehung um sog. **bedingte Wahrscheinlichkeiten** handelt. Die Wahrscheinlichkeit, bei der 2. Ziehung ausgewählt zu werden, beträgt $1/4$, vorausgesetzt, man wurde in der 1. Ziehung nicht berücksichtigt. Diese Wahrscheinlichkeit hat den Wert $4/5$. Die Wahrscheinlichkeit, dass jemand in der 1. Ziehung nicht ausgewählt und in der 2. Ziehung ausgewählt wird, lautet nach dem Multiplikationstheorem der Wahrscheinlichkeiten

$$\frac{N-1}{N} \cdot \frac{1}{N-1} = \frac{1}{N} = \frac{1}{5} \quad (9.2)$$

Für die 3. Ziehung sind die Wahrscheinlichkeiten, sowohl bei der 1. als auch bei der 2. Ziehung nicht ausgewählt worden zu sein, zu beachten. Für diese Wahrscheinlichkeiten erhält man die Werte $4/5$ und $3/4$. Damit ergibt sich zusammengenommen für die 3. Ziehung eine Trefferwahrscheinlichkeit von

$$\frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{1}{N-2} = \frac{1}{N} = \frac{1}{5}$$

Alle fünf anwesenden Skatspieler hatten damit beim **Ziehen ohne Zurücklegen** die gleiche Chance, in die Skatrunde aufgenommen zu werden.

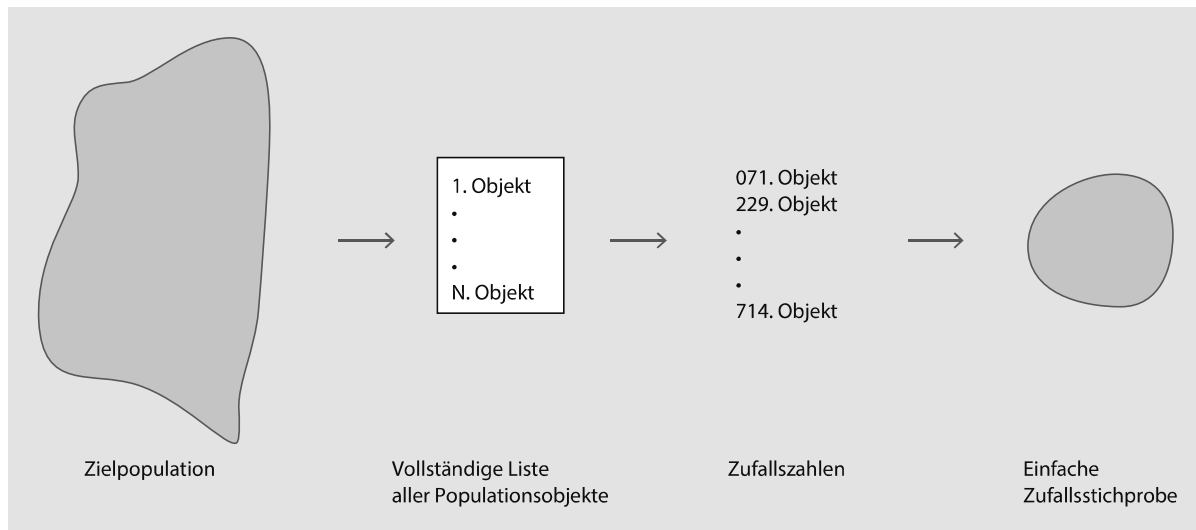
■ **Tabelle 9.4** Zufallszahlen. (Aus The Rand Corporation, 1955)

88473	86062	26357
00677	42981	84552
25227	51260	14800
15386	68200	21492
42021	40308	91104
63058	06498	49339
32548	69104	89073
03521	52177	24816
39975	90626	35889
58252	56687	60412

(diese sind Tabellenwerken oder Zufallszahlengeneratoren wie <http://www.random.org> zu entnehmen) werden aus dieser Liste n Nummern bzw. die dazugehörigen Untersuchungsobjekte ausgewählt (zur Bestimmung von Zufallszahlen Billeter, 1970, S. 15ff.). Soll z. B. aus einer Population von $N = 8\,000$ Untersuchungsobjekten eine Stichprobe des Umfangs $n = 100$ gezogen werden, würde man mit den in ■ Tab. 9.4 aufgeführten Zufallszahlen die Untersuchungsobjekte auswählen, wobei von den fünfstelligen Zufallszahlen dann nur jeweils maximal 4 Ziffern verwendet werden, da der Stichprobenumfang ebenfalls vierstellig ist.

Unter Berücksichtigung der **ersten 4 Ziffern** der Zufallszahlen in der ersten Spalte müsste als erstes Untersuchungsobjekt die Nummer 8847 ausgewählt werden. Da die Population jedoch nur 8000 Elemente enthält, wird diese Nummer ausgelassen. Das erste Untersuchungsobjekt hat dann die Nummer 67, das zweite die Nummer 2522 und so fort. Eine gleichwertige Zufallsstichprobe würde resultieren, wenn man die Auswahl z. B. anhand der letzten 4 Ziffern der zweiten Zahlenkolonne oder anderer Viererkombinationen von Einzelzahlen zusammengestellt hätte. Jede beliebige Auswahl von Zufallszahlen garantiert eine Zufallsstichprobe, vorausgesetzt, eine bereits ausgewählte Zufallszahl wird nicht wieder verwendet. Man nennt dies eine Stichprobenentnahme **ohne Zurücklegen**.

Einfache Zufallsstichprobe – Man zieht eine einfache Zufallsstichprobe („simple random sample“), indem man aus einer vollständigen Liste aller Objekte der finiten Zielpopulation (Auswahlrahmen, „sample frame“) nach einem „blinden“ statistischen Zufallsprinzip (z. B. mittels Zufallszahlen) eine Anzahl von Objekten auswählt, wobei die Auswahlwahrscheinlichkeiten aller Objekte gleich groß sein müssen.



■ **Abb. 9.3** Ziehung einer einfachen Zufallsstichprobe

An dieser Stelle könnte man vermuten, dass sich die Auswahlwahrscheinlichkeiten mit sukzessiver Entnahme von Untersuchungsobjekten ändern, dass sie also nicht – wie gefordert – für alle Untersuchungsobjekte mit einem „**Auswahlsatz**“ von n/N konstant sind. Wächst die Wahrscheinlichkeit, dass ein bestimmtes Untersuchungsobjekt ausgewählt wird, nicht mit fortschreitender Stichprobenentnahme? Dass dem nicht so ist, erläutert ► Chancengleichheit für Skatspieler – Ziehung einer einfachen Zufallsstichprobe (■ Abb. 9.3).

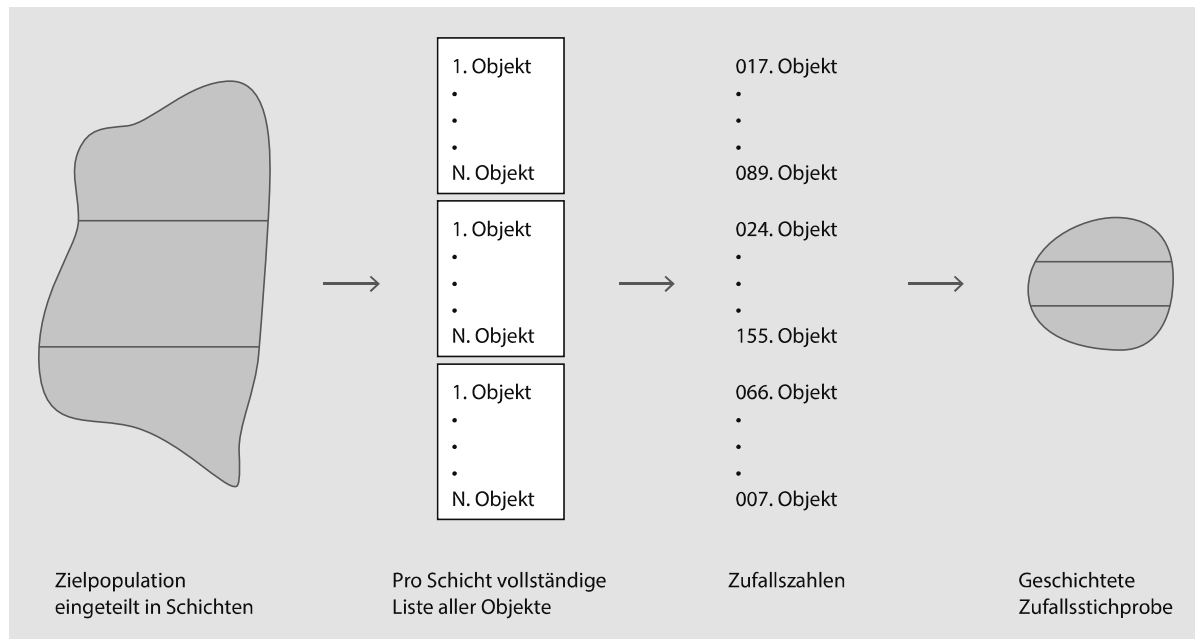
Die Ziehung einer einfachen Zufallsstichprobe setzt voraus, dass jedes Untersuchungsobjekt einer finiten Population erfasst ist und nach dem **Zufallszahlenprinzip** – oder einem anderen Auswahlverfahren, das ebenfalls eine zufällige Auswahl garantiert – ausgewählt werden kann. Eine praktisch häufig eingesetzte Alternative zum Zufallszahlenprinzip ist eine **systematische Auswahl**, bei der wiederum „blind“ (d.h. ohne Ansehen der konkreten Untersuchungseinheit) **jede n -te Einheit** ausgewählt wird. Dieses Verfahren hat den Vorteil, dass es auch auf infinite Populationen anwendbar ist. Beispiel: Soll eine Zufallsstichprobe aus der (ständig wachsenden) Population aller Besucher einer konkreten Website online befragt werden, so kann man über mehrere Tage hinweg z. B. jedem 10. Besucher der Website die Befragungseinladung präsentieren. Die infinite Zielpopulation (alle aktuellen und zukünftigen Website-Besucher) wird bei diesem Verfahren eingeschränkt auf eine Auswahlpopulation all jener Website-Besucher, die die Website im Erhebungszeitraum anklicken. Die **systematische Stichprobe** („systematic sample“) wird als gute Annäherung an die einfache Zufallsstichprobe gemäß Zufallszahlenprinzip aufgefasst.

Systematische Stichprobe – Man zieht eine systematische Stichprobe („systematic sample“), indem man ab einem definierten Anfangspunkt jedes n -te Element der Population zieht. Die in diesem Sinne systematische Stichprobenauswahl stellt eine pragmatische Alternative zur einfachen Zufallsstichprobe dar und ist auch auf infinite Populationen anwendbar.

Neben der systematischen Stichprobe sind **weitere quasi-zufallsgesteuerte Auswahlverfahren aus infiniten Populationen** möglich. So kann z. B. eine Zufallsauswahl von Kunden eines Fast-Food-Restaurants gezogen werden, indem man immer genau die Person herausgreift (z. B. für eine Kundenbefragung), die nach einem Kunden an der Reihe ist, der einen Coupon eingelöst hat. Da das Einlösen des Coupons durch einen Kunden völlig unabhängig davon geschieht, welcher Kunde als nächstes an der Reihe ist, kann das geschilderte Vorgehen als Annäherung an eine Zufallsauswahl für die infinite Population der Kunden verstanden werden. Eine solche quasi-zufallsgesteuerte Auswahl nach einer festen Regel ist auf jeden Fall unverzerrter als eine willkürliche Auswahl, bei der dann unbewusst z. B. besonders sympathische Kunden bevorzugt befragt würden.

Geschichtete Zufallsstichprobe

Bei der geschichteten Zufallsstichprobe erfolgt die Zufallsauswahl der Untersuchungseinheiten nicht direkt aus der Gesamtpopulation, sondern jeweils aus den Schichten der Population, die sich ergeben, wenn man die Population gemäß eines für das Forschungsproblem relevanten soziodemografischen Merkmals strukturiert: Anhand des Merkmals Einkommen mit den drei Ausprägungen niedriges, mittleres und hohes Einkommen lässt sich eine Population in drei Schichten bzw. Teilpopulationen gruppieren. Wird aus jeder dieser drei Schichten ei-



■ **Abb. 9.4** Ziehung einer geschichteten Zufallsstichprobe

ne separate Zufallsstichprobe gezogen, so bilden diese drei einfachen Zufallsstichproben der Schichten zusammen eine geschichtete Stichprobe der Gesamtpopulation (■ Abb. 9.4).

Im Vergleich zu einfachen Zufallsstichproben wird in populationsbeschreibenden Studien die Parameterschätzung präziser, wenn es gelingt, die Population nach einem Merkmal zu schichten, von dem bekannt ist, dass es mit dem untersuchten Merkmal hoch korreliert (für entsprechende Berechnungsformeln ► Webangebot zu diesem Buch). Typische Schichtungsmerkmale sind neben dem Einkommen auch der sozioökonomische Status oder der Wohnort. Der Vorteil präziser Parameterschätzungen geschichteter Stichproben gegenüber einfachen Zufallsstichproben kommt erst dann voll zum Tragen, wenn zusätzlich zum Schichtungsmerkmal die Größen der Teilpopulationen sowie deren Streuungen bekannt sind.

Schichtungsmerkmale sollten nicht nur mit dem untersuchten Merkmal hoch korrelieren, sondern zugleich einfach erhebbar sein. Intelligenz- und Einstellungsvariablen sind z. B. Merkmale, die zwar mit vielen sozialwissenschaftlich interessierenden Sachverhalten zusammenhängen; der Aufwand, der zu ihrer Erfassung erforderlich ist (Einsatz psychologischer Skalen und Tests; ► Abschn. 10.4), macht sie jedoch als Schichtungsmerkmale praktisch unbrauchbar. Für die meisten human- und sozialwissenschaftlichen Studien, die Personen als Merkmalsträger untersuchen, erweisen sich **soziodemografische Merkmale** (Alter, Geschlecht, soziale Schicht, Bildung etc.)

als günstige Schichtungsmerkmale. Hat ein Schichtungsmerkmal z. B. k Ausprägungen, wird für jede Ausprägung eine Zufallsstichprobe des Umfanges n_j benötigt ($j = 1, 2, \dots, k$). Bei einem gegebenen Schichtungsmerkmal entscheidet allein die Aufteilung der Gesamtstichprobe auf die einzelnen Schichten, also die Größe n_j der Teilstichproben, über die Präzision der Parameterschätzung. Bei günstiger Aufteilung erreicht man durch eine geschichtete Stichprobe **genauere Parameterschätzungen** als durch eine einfache Zufallsstichprobe.

Geschichtete bzw. stratifizierte Stichprobe – Man zieht eine geschichtete bzw. stratifizierte Zufallsstichprobe („stratified random sample“), indem man die Zielpopulation auf der Basis eines Merkmals oder mehrerer Merkmale in Teilpopulationen (Schichten) einteilt – pro Merkmalsausprägung bzw. Merkmalskombination entsteht eine Teilpopulation – und aus jeder dieser Schichten eine einfache Zufallsstichprobe entnimmt. Ziel der geschichteten zufälligen Stichprobenauswahl ist es, gegenüber der einfachen Zufallsstichprobe im Rahmen populationsbeschreibender Studien die Genauigkeit von Parameterschätzungen zu erhöhen.

Bisher gingen wir davon aus, dass die Schichtung nur in Bezug auf **ein** Merkmal vorgenommen wird. Diese Beschränkung ist jedoch keineswegs erforderlich und bei Untersuchungen, die gleichzeitig mehrere Sachverhalte untersuchen (Omnibus-Untersuchungen), auch nicht sehr sinnvoll. Hängen nämlich die einzelnen zu untersuchenden Sachverhalte mit jeweils verschiedenen Schichtungsmerkmalen zusammen, können bei einer nur nach

einem Merkmal geschichteten Stichprobe natürlich nur für diejenigen Sachverhalte genauere Parameterschätzungen berechnet werden, die mit dem Schichtungsmerkmal korrelieren. Die Parameterschätzungen zu den übrigen Sachverhalten fallen dann genauso exakt aus wie bei einer einfachen Zufallsstichprobe. Wenn eine Studie also mehrere Sachverhalte untersucht oder wenn der interessierende Sachverhalt mit **mehreren Schichtungsmerkmalen** zusammenhängt, empfiehlt es sich, die Schichtung gleichzeitig nach mehreren Merkmalen vorzunehmen (analog dem kombinierten Quotenplan bei einer Quotenstichprobe, ► Abschn. 9.3.1). Wären z. B. sowohl der Landesteil (Ost-/Westdeutschland) als auch das Einkommen (geringes/mittleres/hohes Einkommen) der Untersuchungspersonen wichtige Schichtungsmerkmale, hätte man Stichproben aus 6 Teilgesamtheiten bzw. Schichten, die als Kombinationen dieser beiden Merkmale resultieren, zu untersuchen. Dieser Aufwand zur genaueren Parameterschätzung lohnt sich allerdings nur, wenn wiederum auch die Umfänge und Streuungen dieser Teilpopulationen bekannt sind.

Grundsätzlich lassen sich zwei Varianten der geschichteten Stichprobe anhand des Auswahlgesetzes (d. h. der Anzahl der Stichprobenelemente in der jeweiligen Schicht im Verhältnis zur Anzahl der Populationselemente in der Schicht: n_j / N_j) unterscheiden:

- **Proportional geschichtete Stichprobe:** Wenn die Stichprobenumfänge n_j zu ihren jeweiligen Teilpopulationen N_j proportional sind, sprechen wir von einer proportional geschichteten Stichprobe. In diesem Falle haben alle Objekte der Gesamtpopulation – wie bei der einfachen Zufallsstichprobe – dieselbe Auswahlwahrscheinlichkeit $n_j / N_j = \text{konst.}$
- **Disproportional geschichtete Stichprobe:** Bei disproportional geschichteten Stichproben sind die Auswahlwahrscheinlichkeiten für die einzelnen Teilpopulationen unterschiedlich. So können dann z. B. aus einer eher kleinen Populationsschicht (z. B. gesellschaftliche Elite) überproportional viele Personen in die Stichprobe gezogen werden, wenn insbesondere über diese Teilpopulation genaue Aussagen getroffen werden sollen. In ihrer Gesamtzusammensetzung weist die disproportional geschichtete Stichprobe systematische – aber aktiv hergestellte und somit bekannte – Abweichungen von der Zielpopulation auf. Werden diese bei der statistischen Parameterschätzung berücksichtigt, resultieren präzisere Ergebnisse als bei der proportional geschichteten Stichprobe.

Die geschichtete Stichprobe ist als aufwändiges probabilistisches Verfahren nicht mit der Quotenstichprobe, ihrem nicht-probabilistischen Pendant, zu verwechseln.

Beide Verfahren verfolgen die Zielsetzung, eine Stichprobe entlang zentraler soziodemografischer Merkmale (Schichtungs- bzw. Quotierungsmerkmale) auf die Zielpopulation hin zu optimieren. Bei der geschichteten Stichprobe werden dabei jedoch aus den vordefinierten Populationsschichten jeweils echte Zufallsstichproben gezogen. Bei der Quotenstichprobe dagegen wird für jede Ausprägung der Quotierungsmerkmale eine bestimmte Zahl von Personen willkürlich ausgewählt.

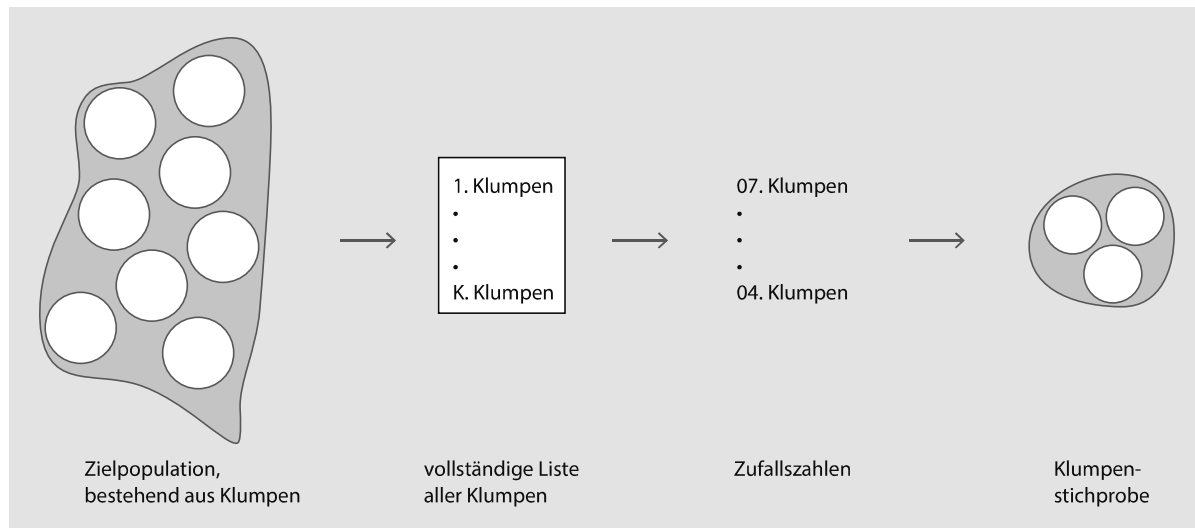
Klumpenstichprobe

Wann immer eine Population aus vielen Gruppen oder natürlich zusammenhängenden Teilkollektiven (Klumpen) besteht, bietet sich die Ziehung einer Klumpenstichprobe („cluster sample“) an. Bei der Klumpenstichprobe wird eine **vollständige Liste der Klumpen** als Auswahlrahmen zugrunde gelegt. Die aus dieser Liste zufällig ausgewählten Klumpen werden dann jeweils vollständig untersucht (► Abb. 9.5).

Klumpenstichprobe – Man zieht eine Klumpenstichprobe („cluster sample“), indem man aus einer in natürliche Gruppen (Klumpen) gegliederten Population nach dem Zufallsprinzip eine ausreichende Anzahl von Klumpen auswählt und diese Klumpen dann vollständig untersucht.

Bei einer Untersuchung von Schulkindern würde man z. B. die Klumpenstichprobe aus mehreren zufällig ausgewählten Schulklassen (= Klumpen) zusammensetzen und bei einer Untersuchung von Betriebsangehörigen einzelne Betriebe oder Abteilungen per Zufall aus der Gesamtliste auswählen und die Klumpen dann vollständig erheben. Untersucht man Krankenhauspatienten, könnte man z. B. alle Patienten einiger zufällig ausgewählter Krankenhäuser zu einer Klumpenstichprobe zusammenfassen. Man beachte: Werden willkürlich einige Schulklassen untersucht, so handelt es sich **nicht** um eine Klumpenstichprobe, sondern um eine Gelegenheitsstichprobe (► Abschn. 9.3.1). Die Klumpenstichprobe als probabilistisches Verfahren basiert auf einer echten Zufallsauswahl einer ausreichenden Mindestzahl von Klumpen aus einer **Liste aller Klumpen**, aus denen sich die geklumpete Population zusammensetzt.

Die Klumpenstichprobe erfordert **weniger organisatorischen Aufwand als die einfache Zufallsstichprobe**. Während eine einfache Zufallsstichprobe strenggenommen voraussetzt, dass alle Untersuchungsobjekte der Population einzeln im Auswahlrahmen erfasst sind und einzeln rekrutiert werden, benötigt die Klumpenstichprobe lediglich eine vollständige Liste aller in der Population enthaltenen Klumpen (z. B. aller Krankenhäuser, aller Betriebe, aller Wahlkreise etc.), die dann auch oft als Gruppe für die Datenerhebung rekrutiert werden können. Eine



■ **Abb. 9.5** Ziehung einer Klumpenstichprobe

vollständige Liste aller Klumpen einer geklumpten Population ist in der Regel einfacher anzufertigen als eine vollständige Liste aller einzelnen Populationsmitglieder.

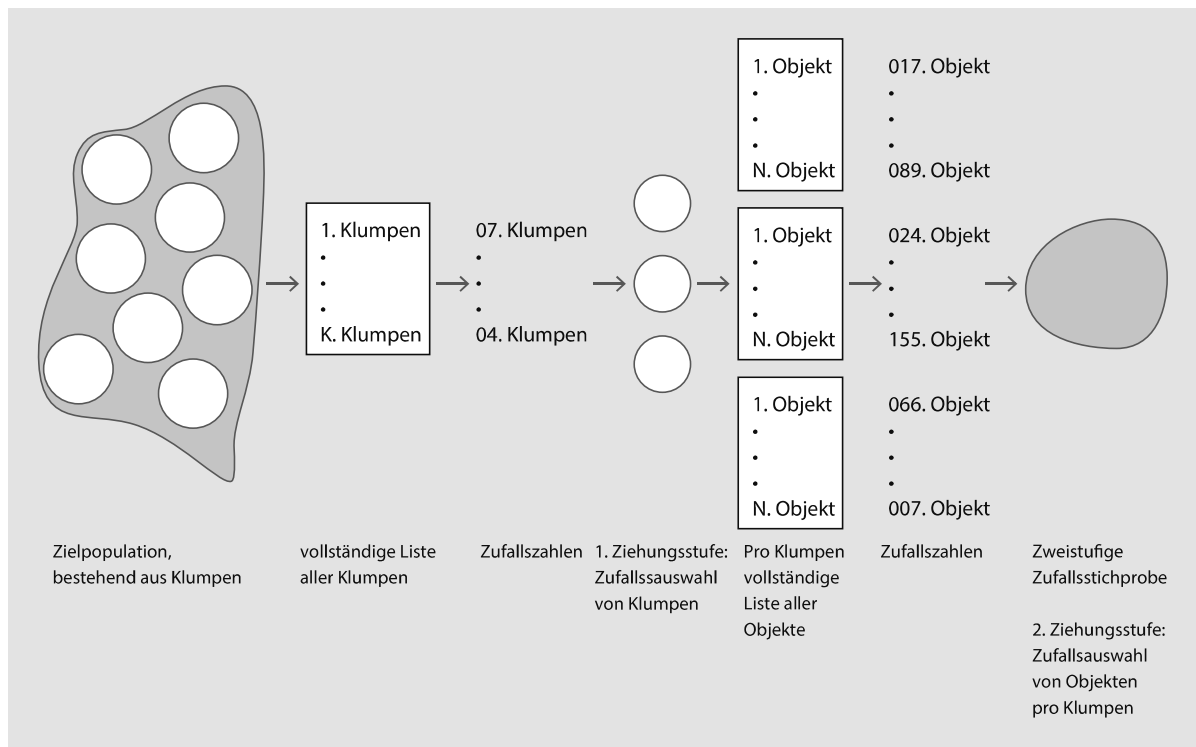
Der Auswahlvorgang bezieht sich bei der Klumpenstichprobe nicht, wie bei der einfachen Zufallsstichprobe, auf die einzelnen Untersuchungsobjekte, sondern auf die Klumpen, wobei sämtliche ausgewählten Klumpen vollständig, d. h., mit allen Untersuchungsobjekten erfasst werden ($n_j = N_j$). Die Auswahlwahrscheinlichkeit ist für jeden Klumpen gleich. Es ist darauf zu achten, dass jedes Untersuchungsobjekt nur einem Klumpen angehört, dass sich also die Klumpen nicht wechselseitig überschneiden (■ Abb. 9.5).

Die Schätzung von Populationsparametern (z. B. Populationsmittelwert μ , Populationsanteilstwert π) auf der Basis von Stichprobenkennwerten ist bei Klumpenstichproben präziser als bei einfachen Zufallsstichproben, sofern **Gleichheit aller Klumpen** und **Verschiedenartigkeit der Untersuchungsobjekte innerhalb der Klumpen** vorausgesetzt werden können (für entsprechende Berechnungsformeln ► Webangebot zu diesem Buch). Wird eine Klumpenstichprobe von z. B. zehn Arbeitsteams eines Großbetriebes gezogen, so bedeutet Gleichheit der Klumpen, dass sich die Teams untereinander ähneln (z. B. hinsichtlich Mitgliederzahl, Arbeitsaufgaben, Qualifikationsprofil der Mitglieder). In sich sind die Klumpen (Arbeitsteams) jedoch heterogen, etwa hinsichtlich Persönlichkeitsmerkmalen, Einstellungen, Hobbys der einzelnen Teammitglieder. Sind diese Voraussetzungen verletzt, kommt es zu einem unerwünschten **Klumpungseffekt** bzw. Klumpeneffekt („cluster effect“), der die Repräsentativität der Stichprobe reduziert: Wird z. B. in einer inhaltsanalytischen Studie von TV-Nach-

richtensendungen mit einer Klumpenstichprobe gemäß der „künstlichen Woche“ gearbeitet (d. h. alle Nachrichtensendungen eines zufällig ausgewählten Montags, eines zufällig ausgewählten Dienstags, eines zufällig ausgewählten Mittwochs etc. werden erhoben), so entsteht ein starker Klumpungseffekt. Denn die Nachrichtensendungen eines Tages (= Klumpens) ähneln sich untereinander hinsichtlich ihrer Themen sehr stark, gleichzeitig weichen sie von jedem anderen Klumpen deutlich ab (Jandura, Jandura, & Kuhlmann, 2005). Bei diesem Beispiel ist aufgrund der Problematik des Klumpungseffekts lieber auf eine einfache Zufallsauswahl zurückzugreifen (d. h. direkte Auswahl einzelner Nachrichtensendungen über das gesamte Jahr hinweg anstelle der Auswahl ganzer Tage als Klumpen).

Mehrstufige Zufallsstichprobe

Eine Klumpenstichprobe – so zeigte der vorige Abschnitt – setzt sich aus mehreren per Zufallsprinzip aus der Grundgesamtheit gezogenen Klumpen zusammen, die alle vollständig untersucht werden. In der Praxis kommt es häufig vor, dass die angetroffenen Klumpen zu groß sind, um sie vollständig zu erheben. In dieser Situation zieht man statt einer einfachen Klumpenstichprobe eine **zwei- oder mehrstufige Zufallsstichprobe** („multistage sampling“). Die erste Stufe betrifft die Zufallsauswahl der Klumpen aus dem Auswahlrahmen und die zweite Ziehungsstufe die Zufallsauswahl der Untersuchungsobjekte aus den einzelnen Klumpen. Damit erfasst eine zweistufige Stichprobe im Unterschied zur einstufigen Klumpenstichprobe die einzelnen Klumpen nicht vollständig, sondern nur in zufälligen Ausschnitten. Soll z. B. eine probabilistische Stichprobe von Schulkindern



■ **Abb. 9.6** Ziehung einer zweistufigen Zufallsstichprobe

einer Großstadt gezogen werden, so kann man in der ersten Ziehungsstufe eine Zufallsauswahl von Schulen (Klumpen) ziehen, in der zweiten Ziehungsstufe werden aus allen ausgewählten Schulen jeweils einfache Zufallsstichproben von Schüler/innen ausgewählt. Im Ergebnis liegt eine **zweistufige Klumpen- bzw. Zufallsstichprobe** vor (■ Abb. 9.6). Mit wachsender Klumpenanzahl (bzw. abnehmender Klumpengröße) verbessern in sich heterogene, aber untereinander homogene Klumpen die Parameterschätzung bei einer zweistufigen Stichprobe.

Eine **dreistufige Zufallsstichprobe** von Grundschulkindern erhält man z.B., wenn aus der Liste aller zur Population gehörenden Grundschulen eine Zufallsauswahl gezogen wird (1. Stufe), aus diesen Schulen zufällig Schulklassen (2. Stufe) und aus den Klassen wiederum zufällig eine festgelegte Anzahl von Schulkindern ausgewählt werden (3. Stufe). Ein populäres Beispiel für ein dreistufiges Stichprobensystem, das Repräsentativität für die Gesamtbevölkerung anstrebt, ist das Stichprobensystem des Arbeitskreises Deutscher Marktforschungsinstitute (**ADM-Mastersample**). Dieses Stichprobensystem kommt u. a. beim **ALLBUS** (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften; <http://www.gesis.org/allbus/>) zum Einsatz, mit dem seit 1980 in zweijährigem Abstand Mehrthemen-Umfragen (Omnibus-Umfragen) zu Bereichen wie Arbeit, Soziales, Umwelt, Politik

etc. durchgeführt werden (Lipsmeier, 1999, S. 102f.). Bei bevölkerungsrepräsentativen mehrstufigen Stichproben sind die Klumpen in der ersten Ziehungsstufe meist geografische Flächen (Wahlkreise). In der zweiten Ziehungsstufe werden aus den Wahlkreisen jeweils Haushalte (wiederum Klumpen) ausgewählt. Auf der dritten Ziehungsstufe schließlich werden per Zufallsprinzip aus den Haushalten (z. B. mittels des sog. Schweden-Schlüssels oder der Next-Birthday-Methode; ► Abschn. 10.2.3 „Vollstrukturiertes persönliches Interview“) die Befragungspersonen ausgewählt.

Mehrstufige Stichprobe – Man zieht eine mehrstufige Zufallsstichprobe („multi-stage random sample“), indem man zunächst eine Klumpenstichprobe mit großen Klumpen zieht (1. Ziehungsstufe). Diese Klumpen werden nicht vollständig untersucht, sondern nur in zufälligen Ausschnitten. Werden aus den Klumpen mittels einfacher Zufallsstichprobe die Untersuchungsobjekte gezogen (2. Ziehungsstufe), so liegt eine zweistufige Klumpen- bzw. Zufallsstichprobe vor („two stage cluster/random sample“). Zieht man auf der zweiten Stufe wieder eine Klumpenstichprobe, ergibt sich durch Ziehung einer Zufallsstichprobe aus diesen Klumpen eine 3. Ziehungsstufe und somit eine dreistufige Klumpen- bzw. Zufallsstichprobe („three stage cluster/random sample“).

Diverse **kombinierte Stichprobenpläne** mit mehreren Ziehungsstufen und unterschiedlichen Auswahlverfahren sind möglich (z. B. auch stratifizierte mehrstufige Zufallsstichproben). Allerdings steigt mit der Komplexität der

Auswahlstufen auch die Komplexität der Berechnungsmodelle für Parameterschätzungen. Genauere Hinweise und Formeln zur Schätzung von Populationsparametern auf der Basis unterschiedlicher probabilistischer Stichproben sind dem ► Webangebot zu diesem Buch zu entnehmen.

? Lernquiz

1. Was versteht man unter einer Vollerhebung und wann ist sie der Stichprobenziehung vorzuziehen?
2. Erklären Sie die Unterschiede zwischen Ziel-, Auswahl- und Interferenzpopulation!
3. In der Presse ist oft von „repräsentativen Studien“ die Rede. Warum ist diese Bezeichnung irreführend?
4. Was versteht man unter einer „Theoretischen Stichprobe“ und wann wird sie eingesetzt?
5. Wie ist die „einfache Zufallsstichprobe“ definiert?
6. Ist die Gelegenheitsstichprobe eine Zufallsstichprobe? Begründen Sie Ihre Antwort!
7. Was versteht man unter „probabilistischen Stichproben“, wozu werden sie eingesetzt und welche Arten von probabilistischen Stichproben sind zu unterscheiden?
8. Worin unterscheidet sich die Klumpenstichprobe von der geschichteten Stichprobe?
9. In der Zeitung lesen Sie unter der Überschrift „Haschisch macht müde und faul“ folgende Meldung: „Wie eine amerikanische Repräsentativstudie zeigt, haben 70 % aller Haschischkonsumenten unterdurchschnittliche Schulleistungen. Gleichzeitig schlafen sie überdurchschnittlich lange. Diese Befunde belegen eindrücklich, wie gefährlich eine liberale Drogenpolitik ist.“ In dieser Nachricht sind 5 Fehler versteckt. Welche?
10. Nennen Sie Vor- und Nachteile probabilistischer Stichprobenverfahren.

? Übungsaufgaben

In einem privaten Verkehrsbetrieb sind 20 Busfahrer und Busfahrerinnen beschäftigt (Population), deren Unfallzahlen in der ■ Tab. 9.5 abgedruckt sind (Geschlecht: 1 = Frau, 2 = Mann).

Ziehen Sie aus dieser Population verschiedene Arten von Stichproben (aus didaktischen Gründen wird hier mit einem sehr kleinen Datensatz gearbeitet):

- eine **einfache Zufallsstichprobe** von $n = 10$ (verwenden Sie dazu die Zufallszahlen aus ■ Tab. 9.4 in ► Abschn. 9.3.2 „Einfache Zufallsstichprobe“, und zwar die 2. Kolonne von rechts, beginnend mit „85734“; streichen Sie von den 5-stelligen Zahlen die hinteren 3 Ziffern weg und suchen Sie von oben nach unten 10 Zufallszahlen im

■ **Tabelle 9.5** Unfallzahlen von allen Busfahrern und Busfahrerinnen eines Verkehrsbetriebs ($N = 20$)

Person	Geschlecht	Unfallzahl
1	1	0
2	1	3
3	2	2
4	2	0
5	1	0
6	1	1
7	2	0
8	1	1
9	2	3
10	2	2
11	2	0
12	2	4
13	1	0
14	1	1
15	1	1
16	1	2
17	1	0
18	2	3
19	2	2
20	2	0

Wertebereich 1–20 aus, doppelte Zahlen werden übersprungen);

- eine **systematische Stichprobe** von $n = 4$ (ziehen Sie jedes 5. Objekt der Population, so dass Sie $n = 4$ Objekte erhalten),
- eine **Quotenstichprobe** von $n = 10$ mit der Vorgabe 80:20 für das Verhältnis von Frauen und Männern (greifen Sie bewusst die ersten 8 Frauen und die ersten 2 Männer heraus).
- Da sich in der Population prozentual gleichviele Männer und Frauen befinden (50:50), ist eine Stichprobe mit einem Geschlechterverhältnis 80:20 unglücklich gewählt. Durch Höhergewichtung der unterrepräsentierten Männer und Heruntergewichtung der überrepräsentierten Frauen in der obigen Quotenstichprobe lässt sich im

Nachhinein ein Verhältnis von 50:50 erzeugen (**gewichtete Quotenstichprobe** von $n = 10$). Berechnen Sie die notwendigen Gewichtungsfaktoren für alle Stichprobenmitglieder als einfache Soll/Ist-Gewichte (Soll: angestrebte Anzahl der Fälle mit bestimmter Merkmalsausprägung dividiert durch Ist: tatsächliche Anzahl von Fällen in der ungewichteten Stichprobe; Kontrolle: die Summe der Soll/Ist-Gewichte über alle Stichprobenelemente hinweg muss 10 ergeben).

- Berechnen Sie für die Population und für alle 4 Stichproben jeweils die durchschnittliche Unfallzahl und vergleichen Sie die Ergebnisse.

9

Weiterführende Literatur

- Kauermann, G. & Küchenhoff, H. (2010). *Stichproben: Methoden und praktische Umsetzung mit R*. Berlin: Springer.
- Schreier, M. (2011). Qualitative Stichprobenkonzepte. In: G. Naderer & E. Balzer (Hrsg.). *Qualitative Marktforschung in Theorie und Praxis. Grundlagen, Methoden und Anwendungen* (2. Aufl., S. 241–256). Wiesbaden: Gabler
- Tashakkori, A. & Teddlie, C. (2009). Sampling strategies for mixed methods research. In: Teddlie, C. & Tashakkori, A. (Eds.). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences* (pp. 168–196). Thousand Oaks: Sage.

Literatur

- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., et al. (2013). *Report of the AAPOR Task Force on non-probability sampling*: American Association for Public Opinion Research AAPOR. Retrieved October 4, 2013, from <http://www.aapor.org/AM/Template.cfm?Section=Reports1&Template=/CM/ContentDisplay.cfm&ContentID=5963>.
- Behnke, J. (2005). Lassen sich Signifikanztests auf Vollerhebungen anwenden? Einige essayistische Anmerkungen. *Politische Vierteljahresschrift*, 46(1), 2–15.
- Berland, M. A. & Thomas, K. R. (1991). Psychosocial issues following severe head injury in adolescence: Individual and family perceptions. *Rehabilitation Counseling Bulletin*, 35(1), 5–22.
- Billeter, E. P. (1970). *Grundlagen der repräsentativen Statistik*. Wien: Springer.
- Bjorkman, M. & Malterud, K. (2009). Lesbian women's experiences with health care: A qualitative study. *Scandinavian Journal of Primary Health Care*, 27(4), 238–243.
- Bortz, J., Lienert, G. A., & Boehnke, K. (2008). *Verteilungsfreie Methoden in der Biostatistik* (3rd ed.). Heidelberg: Springer.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th. ed.). Berlin, Heidelberg: Springer.
- Broscheid, A. & Geschwend, T. (2005). Zur statistischen Analyse von Vollerhebungen. *Politische Vierteljahresschrift*, 46(1), 16–26.
- Chersich, M. F., Luchters, S. M. F., Malonza, I. M., Mwarogo, P., King'ola, N., & Temmerman, M. (2007). Heavy episodic drinking among Kenyan female sex workers is associated with unsafe sex, sexual violence and sexually transmitted infections. *International Journal of STD and AIDS*, 18(11), 764–769.
- Christopher, F. S., Pflieger, J. C., Canary, D. J., Guerrero, L. K., & Holtzworth-Munroe, A. (2008). Targeted neighborhood sampling: A new approach for recruiting abusive couples. *Journal of Family Violence*, 23(2), 89–100.
- Diekmann, A. (2005). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Reinbeck: Rowohlt.
- DMS Research. (2008). A Deep Dive into the River – Comparing Multiple Streams of River Respondents. Retrieved September 9, 2013, from http://www.usamp.com/objects/pdf/pr/DMS/DMS_River_Deep_Dive_White_Paper.pdf.
- Drake, F. (2011). Protesting mobile phone masts: Risk, neoliberalism, and governmentality. *Science, Technology and Human Values*, 36(4), 522–548.
- Edgington, E. & Onghena, P. (2007). *Randomization Tests* (4th ed.). Boca Raton, FL: Taylor & Francis.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2013). *Statistik und Forschungsmethoden* (3rd ed.). Weinheim: Beltz.
- Flick, U. (1996). *Psychologie des technisierten Alltags – Soziale Konstruktion und Repräsentation technischen Wandels in verschiedenen kulturellen Kontexten*. Opladen: Westdeutscher Verlag.
- Flick, U. (2010). *Qualitative Sozialforschung. Eine Einführung* (3rd ed.). Reinbeck: Rowohlt.
- Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics*. New York: Norton.
- Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.
- Glaser, B. G. & Strauss, A. L. (1999). *The discovery of grounded theory* (8th ed.). Chicago: Aldine.
- Goodman, L. A. (1961). Snowballsampling. *Annals of Mathematical Statistics*, 32(1), 148–170.
- Goscinnny, R. & Sempé, J.-J. (1976). *Der kleine Nick und die Mädchen*. Zürich: Diogenes.
- Hager, W. (2013). The statistical theories of Fisher and of Neyman and Pearson: A methodological perspective. *Theory and Psychology*, 23(2), 251–270.
- Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44(2), 174–199.
- Higginbottom, G. M. A. (2004). Sampling issues in qualitative research. *Nurse Researcher*, 12(1), 7–19.
- Jandura, G., Jandura, O., & Kuhlmann, C. (2005). Stichprobenziehung in der Inhaltsanalyse. Gegen den Mythos der künstlichen Woche. In V. Gehrau, B. Fretwurst, B. Krause & G. Daschmann (Hrsg.), *Auswahlverfahren in der Kommunikationswissenschaft* (S. 71–116). Köln: Halem.
- Knowles, A. (2011). Resilience among Japanese atomic bomb survivors. *International Nursing Review*, 58(1), 54–60.
- Kozak, M. (2005). Finite and infinite populations in biological statistics: Should we distinguish them? *The Journal of American Science* 4(1), 59–62.

- Lam, W. K. K., Cance, J. D., Eke, A. N., Fishbein, D. H., Hawkins, S. R., & Williams, J. C. (2007). Children of African-American mothers who use crack cocaine: Parenting influences on youth substance use. *Journal of Pediatric Psychology*, 32(8), 877–887.
- Legault, A. & Ducharme, F. (2009). Advocating for a parent with dementia in a long-term care facility: The process experienced by daughters. *Journal of Family Nursing*, 15(2), 198–219.
- Lipsmeier, G. (1999). Standard oder Fehler? Einige Eigenschaften von Schätzungsverfahren bei komplexen Stichprobenplänen und aktuelle Lösungsansätze. *ZA-Information*, 44, 96–117.
- Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, 13(6), 522–525.
- Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 11(3). Retrieved October 4, 2013, from <http://www.qualitative-research.net/index.php/fqs/article/view/1428>.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks: Sage.
- Peterson, J. A., Schacht Reisinger, H., Schwartz, R. P., Mitchell, S. G., Kelly, S. M., Brown, B. S., et al. (2008). Targeted sampling in drug abuse research: A review and case study. *Field Methods*, 20(2), 155–170.
- Safman, R. M. & Sobal, J. (2004). Qualitative sample extensiveness in health education research. *Health Education and Behavior*, 31(1), 9–21.
- Scheuch, E. K. (1974) Auswahlverfahren in der Sozialforschung. *Handbuch der empirischen Sozialforschung, Bd. 3a Grundlegende Methoden und Techniken*. Stuttgart: Enke.
- Schnell, R., Hill, P. B., & Esser, E. (2008). *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- The Rand Corporation. (1955). *A million random digits with 100 000 normal deviates*. Glencoe: Free Press.
- Turgeman-Goldschmidt, O. (2005). Hackers' Accounts: Hacking as a Social Entertainment. *Social Science Computer Review*, 23(1), 8–23.
- Watters, J. K. & Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. *Social Problems*, 36(4), 416–430.
- Wejnert, C. & Heckathorn, D. D. (2008). Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research. *Sociological Methods and Research*, 37(1), 105–134.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik. Ein Lehrbuch zur psychologischen Methodenlehre*. Göttingen: Hogrefe.
- Winterich, J. A., Quandt, S. A., Grzywacz, J. G., Clark, P. E., Miller, D. P., Acuña, J., et al. (2009). Masculinity and the body: How African American and white men experience cancer screening exams involving the rectum. *American Journal of Men's Health*, 3(4), 300–309.