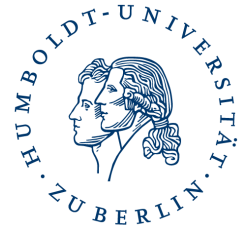


HUMBOLDT-UNIVERSITÄT ZU BERLIN



Methodenbericht

*Zur Erhebung "Großstadt und soziale Ungleichheit" im Rahmen
der Lehrveranstaltung Empirische Sozialforschung I+II im
Wintersemester 2019/2020*

Dr. Markus Schrenker
Lehrbereich Empirische Sozialforschung
Institut für Sozialwissenschaften

Berlin, März 2020

Inhalt

1 Vorbemerkung.....	3
2 Projektskizze.....	3
3 Erhebungsinstrument.....	4
4 Pretest	4
5 Feldzeit und Interviewdauer.....	4
6 Interviewer*innen	6
7 Stichprobendesign und Ausschöpfung.....	6
7.1 Stichprobendesign.....	7
7.2 Ausschöpfung	8
8 Gewichtung.....	5
8.1 Berechnung der Designgewichte.....	5
8.2 Redressmentgewichtung.....	6
9 Abschließende Bemerkungen.....	8
Literatur.....	10

1 Vorbemerkung

Im Rahmen der jährlich stattfindenden Lehrveranstaltung 'Empirische Sozialforschung I+II' haben Studierende der Sozialwissenschaften der Humboldt-Universität eine telefonische Befragung der deutschsprachigen Berliner Wohnbevölkerung zum Thema 'Großstadt und soziale Ungleichheit' durchgeführt. Die Interviews wurden vom 20. 11. 2019 bis 14. 12. 2019 im Telefonlabor des Instituts für Sozialwissenschaften mittels Computer Assisted Telephone Interviewing (CATI) erhoben. Ziel der Lehrveranstaltung ist es sowohl methodische Kompetenzen als auch praktische Erfahrungen bei der Konzeption und Umsetzung empirischer Untersuchungen zu vermitteln. Die Studierenden durchliefen den gesamten Forschungsprozess von der Theoriebildung und Hypothesengenerierung, über die Konzeptspezifikation und Erstellung eines praxistauglichen Fragebogens, bis hin zur Datenerhebung und -auswertung nach aktuellen wissenschaftlichen Standards. Grundlage der Stichprobenziehung war eine Festnetzstichprobe nach dem Gabler-Häder-Verfahren. Vorgelegt wird hier die Methodendokumentation für die CATI-Befragung. Der Bericht informiert über die Einzelheiten der Stichprobenanordnung, der Durchführung der Feldarbeit, der Datenaufbereitung und der Gewichtung. Zur besseren Übersichtlichkeit wird dem Bericht eine kurze Projektskizze mit den wichtigsten Rahmendaten vorangestellt.

2 Projektskizze

Tabelle 1: Projektübersicht

Grundgesamtheit:	Deutschsprachige Wohnbevölkerung über 18 Jahre in Berlin, die über Festnetzanschlüsse erreichbar ist
Art der Befragung:	Telefonisches Interview mit CATI
Realisierte Stichprobengröße:	N = 892
Erhebungszeitraum:	20. 11. 2019 bis 14. 12. 2019
Stichprobe:	Festnetzstichprobe nach dem Gabler-Häder-Verfahren
Auswahlverfahren:	Last-Birthday
Kontaktregel:	Mindestens 8 Anrufversuche
Mittlere Interviewdauer (in Min.):	38.7
Durchschnittliche Interviewdauer (in Min.):	41.6
Adressbrutto:	N = 34273
Bereinigter Stichprobenansatz:	N = 6663
Bereinigte Ausschöpfung in Prozent:	13.4
Anzahl Interviewer*innen:	125
Vollständige Interviews pro Interviewer*in:	7

3 Erhebungsinstrument

Das für die telefonische Befragung verwendete standardisierte Erhebungsinstrument basiert auf der Vorarbeit der 29 Arbeitsgruppen des Seminars "Empirische Sozialforschung I+II" im Sommersemester 2019. Innerhalb des vorgegebenen Themenbereichs "Großstadt und soziale Ungleichheit" entwickelte jede Arbeitsgruppe eigene Fragestellungen, dazu passende theoretische Konstrukte und deren empirische Operationalisierungen. Diese wurden schließlich in eigenen standardisierten Fragebögen umgesetzt, welche neben den konzeptspezifischen Items auch standarddemografische Fragen beinhalteten. Die Kursverantwortlichen destillierten daraus ein synoptisches Gesamtinstrument, wobei die von den Arbeitsgruppen konstruierten Instrumente auf ein handhabbares Maß gekürzt, redigiert und in einer sinnvollen Fragebogendramaturgie angeordnet wurden. Aufgrund der Vielzahl der Arbeitsgruppen deckt das Gesamtinstrument viele unterschiedliche thematische Aspekte wie Wohnzufriedenheit, soziale Teilhabe im Kiez, Gentrifizierung, Verdrängung, räumliche Mobilität, Kriminalitätswahrnehmung, Wohnsituation, Einstellungen zu Wohnungspolitik, zu Fremden und zu Homosexuellen ab. Trotz eines für telefonische Befragungen ungewöhnlich großen Umfangs des Erhebungsinstruments, mussten die Messungen einzelner Konzepte dadurch zwangsläufig sehr sparsam operationalisiert werden. Das Gesamtinstrument wurde schließlich mit Hilfe von "Command Center" der Firma Voxco als CATI-Fragebogen implementiert.

4 Pretest

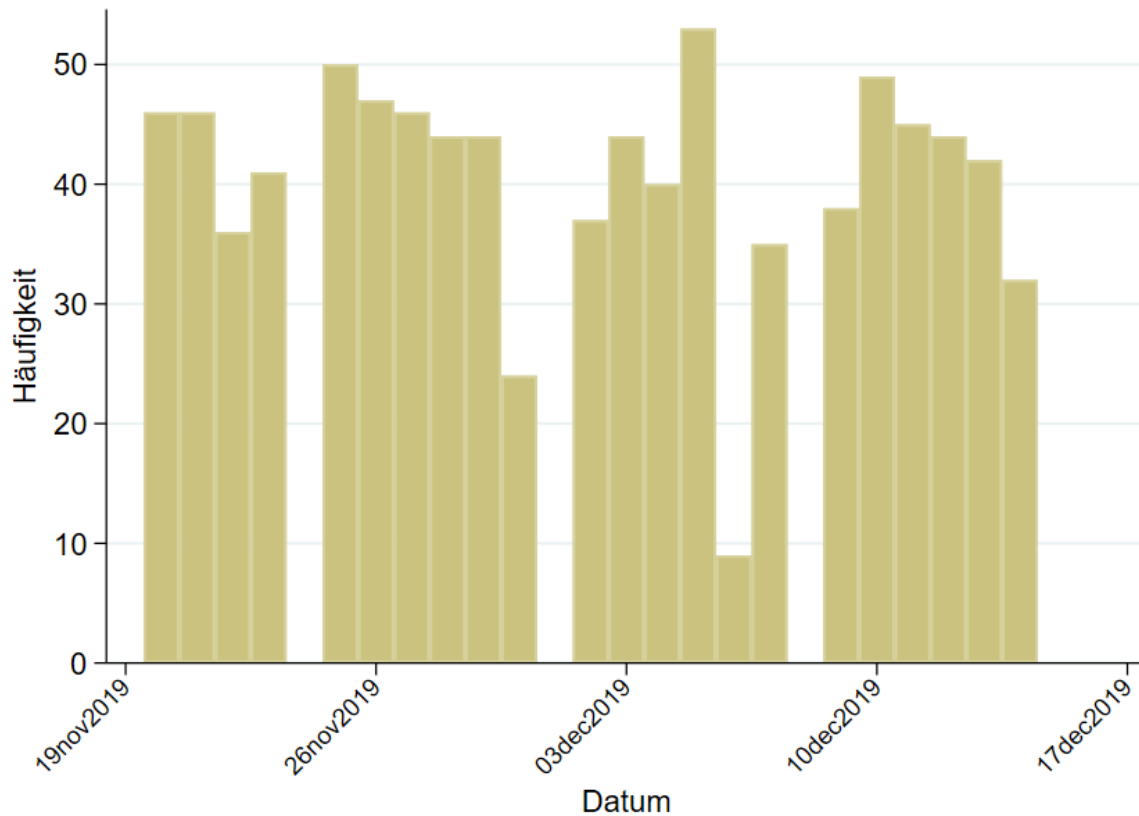
Nach der Programmierung des Erhebungsinstruments wurde vom 6. 11. 2019 bis 23. 11. 2019 ein Pretest unter Feldbedingungen durchgeführt. Der Zweck des Pretests bestand in erster Linie darin, das Erhebungsinstrument im Hinblick auf den flüssigen Interviewablauf, die Interviewdauer und das Frageverständnis zu evaluieren und kleinere Filter- und Programmierfehler zu beheben. Insgesamt wurden hierzu Kontaktversuche auf 3098 verschiedenen Telefonnummern unternommen, die zu 78 vollständigen Interviews führten. Das arithmetische Mittel der Interviewdauer betrug 46 Minuten (Median = 44 Minuten). Insgesamt gab es kaum Sprach- und Verständnisprobleme, lediglich bei 10 Prozent der Interviews kam es zu geringen Problemen bei einzelnen Fragen. Die Antwortbereitschaft war überwiegend gut, nur bei 4 Prozent der Interviews verschlechterte sich die Bereitschaft zum Ende des Interviews, was bei einer Dauer von 46 Minuten sehr positiv zu bewerten ist. Lediglich bei einem Interview wurden die Angaben als weniger zuverlässig eingeschätzt. Im Interesse einer Verringerung der Interviewdauer wurden nach dem Pretest 18 Items gestrichen. Als Kriterium für die Kürzung von Items, dienten hoher Itemnonresponse (über 10 Prozent), geringe Varianz in den Antworten sowie Items die Verständnisschwierigkeiten oder sozial erwünschtes Antwortverhalten produzierten. Bei einzelnen Items wurde von diesen Kriterien abgewichen, wenn Fragen als theoretisch besonders wichtig eingeschätzt wurden. Vier Items wurden ergänzt, um theoretische Lücken zu schließen. Umfangreichere Skalen wurden jedoch auf ihre Markeritems reduziert. Zudem wurden einzelne Frageformulierungen und Antwortskalen modifiziert bzw. vereinfacht. Insgesamt konnte so eine Zeitersparnis von ca. 5 Minuten realisiert werden, womit sich die Dauer des Interviews auf etwa 40 Minuten reduzieren ließ. Hinweise auf Reihenfolgeeffekte berücksichtigten wir durch kleinere Änderungen im Frageablauf oder Randomisierungen der Reihenfolge.

5 Feldzeit und Interviewdauer

Die Feldzeit der Haupterhebung begann am 20. 11. 2019 und endete am 14. 12. 2019. In zweistündigen Schichten wurde an Wochentagen jeweils von 15.00 bis 21.00 Uhr und an Samstagen

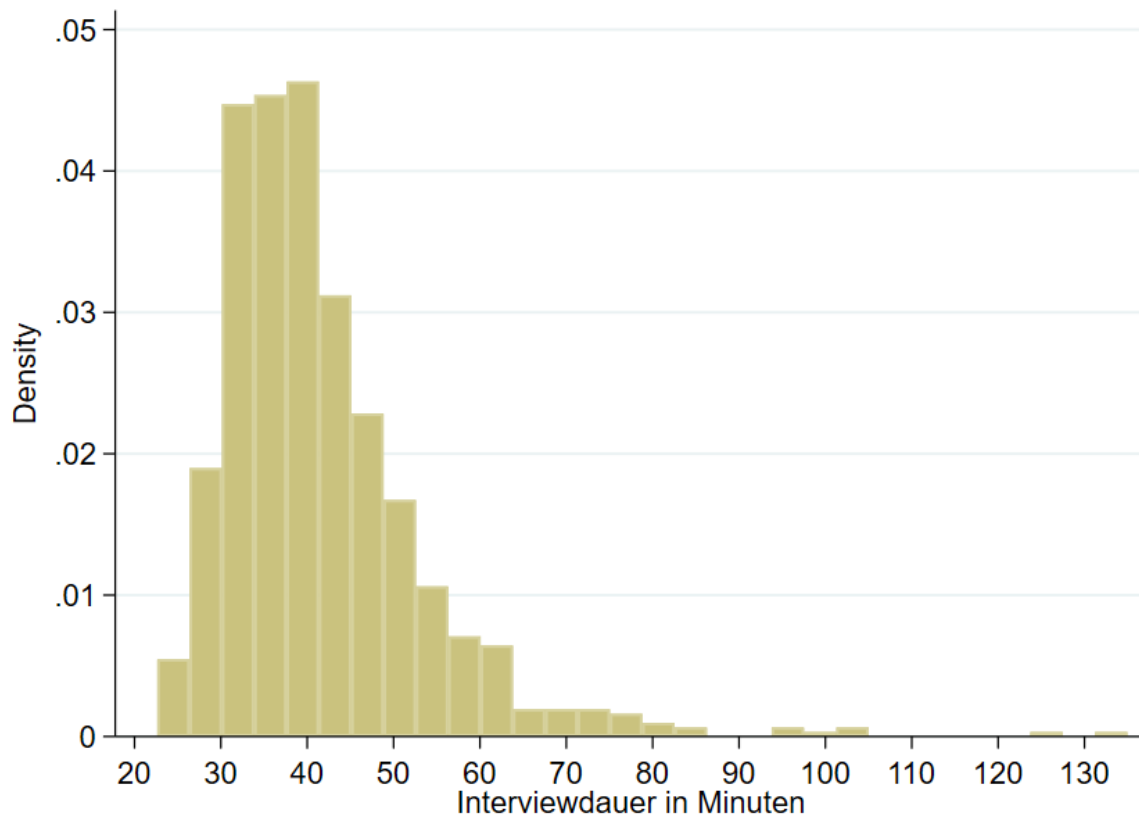
von 12.00 bis 18.00 Uhr telefoniert. Jede Nummer der Bruttostichprobe wurde bis zu acht Mal zu unterschiedlichen Uhrzeiten und an unterschiedlichen Tagen kontaktiert, um eine möglichst hohe Ausschöpfung zu erreichen.

Abbildung 1: Realisierte Interviews über die Feldzeit



Das arithmetische Mittel der Interviewdauer betrug ca. 41.6 Minuten, mit einer Standardabweichung von 12.1 Minuten. Der Median liegt mit 39 Minuten etwas darunter. In 80 Prozent aller Fälle dauerte das Interview zwischen 30 und 56 Minuten. Das Interview war damit für eine CATI-Befragung außergewöhnlich umfangreich. Einzelne extrem lange Interviews zeugen außerdem von einem erhöhten Gesprächsbedarf auf Seiten der Befragungspersonen bei dem Thema der Befragung. Relativ viele Interviews (168) mussten aufgrund der Länge unterbrochen werden, konnten aber fast alle zu einem späteren Zeitpunkt fortgesetzt werden. Der Anteil abgebrochener Interviews ist vor dem Hintergrund der langen Interviewdauer als nicht besonders hoch einzuschätzen. So kommen auf 892 vollständige Interviews zwar 115 Abbrüche, von denen ereigneten sich aber zwei Drittel bereits bei den ersten Fragen und sind insofern eher als Verweigerungen zu zählen.

Abbildung 2: Verteilung der Interviewdauer



6 Interviewer*innen

Für die Befragung wurden insgesamt 125 Interviewer*innen eingesetzt, die im Durchschnitt 7 Interviews durchführten. Alle Interviewer*innen nahmen am Seminar Empirische Sozialforschung teil und wurden sowohl im Hinblick auf allgemeine Prinzipien telefonischer Befragungen und das geforderte Interviewverhalten geschult, z. B. zur Vermeidung von Abbrüchen, als auch speziell in Bezug auf das Instrument der vorliegenden Untersuchung instruiert. Die Schulung umfasste die ausführliche Auseinandersetzung mit dem Fragebogen sowie mindestens zwei Probeinterviews. Die Arbeit der Interviewer*innen in der Feldphase unterlag außerdem fortlaufender Supervision. Das Team der ESF-Tutorinnen und Tutoren betreute während der gesamten Datenerhebung die Studierenden bei auftretenden Schwierigkeiten und Problemen in der Interviewdurchführung – egal ob technisch, inhaltlich oder kommunikativ – und gab Feedback zum erforderlichen Interviewverhalten.

7 Stichprobendesign und Ausschöpfung

Die Grundgesamtheit der vorliegenden Erhebung ist die in Privathaushalten lebende deutschsprachige Wohnbevölkerung über 18 Jahre in Berlin, die potenziell über Festnetzanschlüsse erreichbar ist. Personen, die lediglich über Mobilfunkanschlüsse erreichbar sind oder, die nicht über für die Beantwortung der Fragen erforderlichen Deutschkenntnisse verfügen, sind somit nicht Teil der Inferenzpopulation. Die Verwendung eines sog. Dual-Frame-Ansatzes, bei dem die Festnetzstichprobe durch eine Mobilfunkstichprobe ergänzt wird, kam angesichts des regionalen Bezugs der

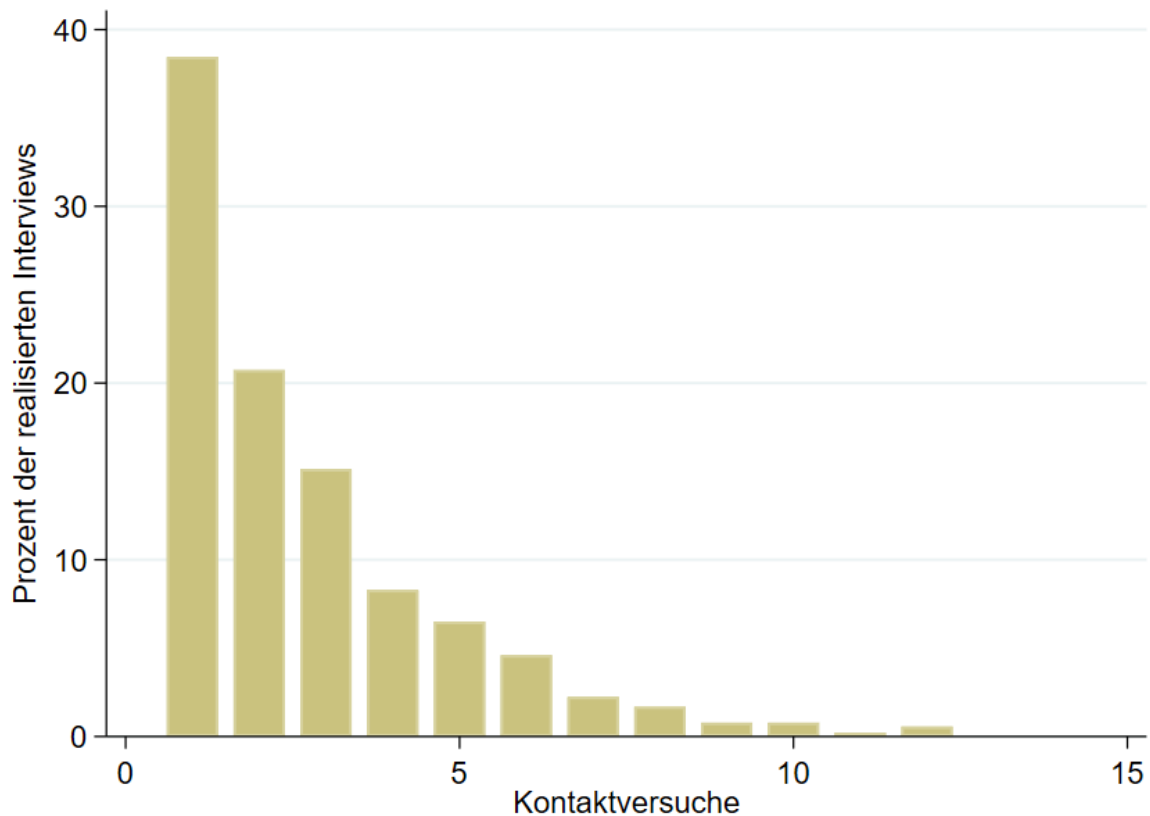
Untersuchung nicht in Frage, da Mobilfunknummern nicht regionalisierbar sind. Daraus resultierende Verzerrungen der demografischen Zusammensetzung der Stichprobe von der Berliner Bevölkerungsstruktur werden im Abschnitt Gewichtung diskutiert.

7.1 Stichprobendesign

Entscheidend für die Absicherung der Repräsentativität der Stichprobe ist eine Zufallsauswahl der Befragungspersonen aus der Grundgesamtheit. Da keine vollständige Liste aller zur Grundgesamtheit gehörenden Personen inklusive ihrer Telefonnummern existiert, aus der mit einfacher Zufallsauswahl ausgewählt werden könnte, haben sich in der empirischen Sozialforschung Verfahren etabliert, mit denen eine möglichst gute Annäherung an eine einfache Zufallsauswahl sichergestellt werden kann. In der vorliegenden Untersuchung wurde ein zweistufiges Auswahlverfahren angewendet, wobei im ersten Schritt eine Festnetzstichprobe nach dem sog. Gabler-Häder-Verfahren gezogen wird (Gabler und Häder 2009). In bekannten 100er Blöcken von Telefonnummern werden die fehlenden Nummern aufgefüllt, so dass im Nummernuniversum auch nicht in Telefonverzeichnissen eingetragene Nummern enthalten sind. Aus diesem Universum wird dann per einfacher Zufallsauswahl ausgewählt (sog. Randomized Last Digit, RLD). Festnetzanschlüsse, die sich nach dem ersten Kontakt als nicht zur Grundgesamtheit gehörend herausstellen, weil es sich zum Beispiel um Firmenanschlüsse handelt, werden aussortiert. Da man so zunächst eine Stichprobe der privaten Haushalte erhält, wird in einem zweiten Auswahlschritt mit der sog. Last-Birthday-Methode eine zufällige Auswahl unter den über 18jährigen Haushaltspersonen sichergestellt. Befragungsperson innerhalb des Haushaltes ist dann die Person über 18 Jahre, die zuletzt Geburtstag hatte und über ausreichende Deutschkenntnisse verfügt. Ist die Person gerade nicht im Haushalt, werden Termine vereinbart und weitere Kontaktversuche unternommen. Ein Rückgriff auf eine andere Person des Haushaltes ist nicht gestattet, um systematische Verzerrungen in der Stichprobe zu vermeiden, wie z. B. eine Überrepräsentation der nicht-erwerbstätigen Bevölkerung. Die Terminierung von Anrufversuchen vor allem in den Abendstunden unterstützt die Erreichbarkeit der Erwerbsbevölkerung.

Pro Rufnummer wurden mindestens acht Kontaktversuche zu unterschiedlichen Tageszeiten und an unterschiedlichen Wochentagen vorgenommen, bis ein Kontakt zustande kam. Das Ziel dieses aufwändigen Verfahrens ist es, die Ausschöpfung zu optimieren und Stichprobenverzerrungen aufgrund einer disproportionalen Erreichbarkeit verschiedener Bevölkerungsgruppen zu minimieren. Im Bedarfsfall wurde über dieses Abbruchkriterium auch deutlich hinausgegangen, wenn z. B. Terminvereinbarungen vorlagen. Abbildung 3 illustriert, dass über 50 Prozent der vollständigen Interviews bereits mit dem zweiten Kontaktversuch realisiert wurden, knapp 75 Prozent mit dem dritten Versuch.

Abbildung 3: Notwendige Kontaktversuche bei vollständigen Interviews



7.2 Ausschöpfung

Für die Beurteilung der Repräsentativität einer Befragung ist die Ausschöpfungsquote ein erster Anhaltspunkt. Je höher die Ausschöpfung ausfällt, umso geringer ist das Potenzial für Stichprobenverzerrungen. Allerdings können auch Befragungen mit einer geringen Ausschöpfung eine hohe Stichprobenqualität aufweisen, solange die zugrunde liegenden Ausfallprozesse zufällig sind. Umgekehrt können außergewöhnlich hohe Ausschöpfungsquoten auf unzulässige Abweichungen vom Auswahlverfahren hinweisen.

Für die Berechnung der Ausschöpfung der CATI-Befragung muss nach unterschiedlichen Gründen für das Nichtzustandekommen eines Interviews differenziert werden. Man unterscheidet zwischen qualitätsneutralen Ausfallgründen und nicht neutralen (oder systematischen) Gründen. Während die neutralen Ausfallgründe die Zufälligkeit der Auswahl unbeeinflusst lassen (z. B. „kein Anschluss“), können die nicht neutralen Ausfallgründe zu Stichprobenverzerrungen führen (z. B. „Befragungsperson verweigert“). Es ist daher sinnvoll, die Ausschöpfung nicht als allgemeine, sondern als bereinigte Ausschöpfung zu berechnen, d. h. als Anteil der vollendeten Interviews an dem durch die neutralen Ausfälle bereinigten Bruttoansatz. Allerdings ist dabei zu beachten, dass für einen Teil der Anschlüsse nicht bestimmbar ist, ob es sich um neutrale oder nicht-neutrale Ausfälle handelt, z. B. wenn trotz wiederholter Anrufe niemand erreicht wurde. Aus diesem Grund wird der Anteil der neutralen Ausfälle unter den unklaren Ausfallgründen geschätzt. Dazu werden die bekannten neutralen Ausfälle durch den Bruttoansatz abzüglich der unklaren Ausfälle geteilt:

$$\text{Schätzer} = \text{Neutrale Ausfälle} / (\text{N(Brutto)} - \text{Unklare Ausfälle})$$

$$0.806 = 22801 / (34273 - 5970)$$

Im Fall der Festnetzstichprobe beträgt der Schätzer für den Anteil neutraler Ausfälle an den unklaren Ausfällen 0.806. Der Schätzer gilt als konservativ, da damit tendenziell die Anzahl neutraler Ausfälle an den Ausfällen insgesamt unterschätzt wird (AAPOR 2011). Die auf dieser Basis bereinigte Ausschöpfungsquote stellt somit eine Untergrenze dar. Der hohe Anteil neutraler Ausfälle erklärt sich vor allem aus der Vielzahl nicht genutzter bzw. geschalteter Nummern im Auswahlrahmen (Kein Anschluss, Faxnummer etc.). Während die Anzahl genutzter Nummernblöcke in den letzten Jahren zugenommen hat, hat die Nutzungsdichte innerhalb der Blöcke abgenommen (Sand 2014: 4). Der um neutrale Ausfälle bereinigte Stichprobenansatz beläuft sich daher auf insgesamt 6663 Festnetzanschlüsse.

Tabelle 2: Ausschöpfungsprotokoll

Effektiver Bruttoansatz	34273
Kein Anschluss, Faxnummer etc.	20265
Kontaktregel überschritten ohne Kontakt	1047
Anschluss gehört nicht zur Zielpopulation	1228
Sprachprobleme/Auslandsnummer	261
<i>Neutrale Ausfälle</i>	<u>22801</u>
Kont. Frei-/Besetzzeichen ohne Kontakt	3727
Kont. Anrufbeantworter/Mailbox	2221
Sofort weggedrückt	22
<i>Unklare Ausfälle</i>	<u>5970</u>
Schätzer Neutral/Unklar:= .806	
<i>Neutrale Ausfälle unter unklaren Ausfällen (geschätzt)</i>	<u>4809</u>
Neutrale Ausfälle insgesamt	<u>27610</u>
Bereinigter Stichprobenansatz	6663
Zielperson verweigert	1138
Kontaktperson verweigert	2393
Offene Termine mit ZP/in Feldzeit nicht erreichbar	909
Offene Termine mit KP	55
Abgebrochene Interviews	115
<i>Nicht neutrale Ausfälle</i>	<u>4610</u>
<i>Nicht neutrale Ausfälle unter unklaren Ausfällen (geschätzt)</i>	<u>1161</u>
Nicht neutrale Ausfälle insgesamt	<u>5771</u>
Auswertbare Interviews	892
Bereinigte Ausschöpfung	<u>0.134</u>

Setzt man die Anzahl der auswertbaren Interviews von 892 ins Verhältnis zum bereinigten Bruttoansatz, so ergibt sich die bereinigte Ausschöpfungsquote. Diese liegt für die vorliegende Stichprobe liegt bei insgesamt 13.4 Prozent, was angesichts des sehr langen und stellenweise anspruchsvollen Interviews sowie dramatisch sinkender Ausschöpfungsquoten in telefonischen Bevölkerungsumfragen in den letzten Jahren (vgl. PEW 2019) als absolut zufriedenstellend einzuschätzen ist. Die derartig berechnete Ausschöpfungsquote entspricht der sog. Response Rate 3 nach AAPOR Standards (AAPOR 2011).

Die Qualität der Stichprobe sollte jedoch nicht allein anhand der Höhe der Ausschöpfung evaluiert werden. Entscheidend für die Repräsentativität der Angaben ist, ob die Verweigerungen überwiegend zufällig zustande kamen oder systematisch mit interessierenden Untersuchungsmerkmalen zusammenhängen. Um dies einschätzen zu können, bietet sich ein Vergleich der Stichprobenzusammensetzung mit bekannten demografischen Merkmalen der Grundgesamtheit an, worauf im nächsten Abschnitt unter „Redressmentgewichtung“ genauer eingegangen werden soll. Ausgewählte Stichprobenverteilungen werden dafür mit der aktuellen Bevölkerungsfortschreibung des Statistischen Bundesamts verglichen. Vorab sind aber disproportionale Auswahlwahrscheinlichkeiten durch geeignete Designgewichte zu korrigieren, die sich aus Abweichungen von der einfachen Zufallsauswahl beim Stichprobendesign ergeben haben.

8 Gewichtung

Voraussetzung für statistische Schlüsse von der Stichprobe auf die Grundgesamtheit ist eine Auswahl aller Stichprobenelemente aus der Grundgesamtheit mit gleicher Wahrscheinlichkeit. Wie in den vorherigen Abschnitten dargelegt, ist in der vorliegenden Untersuchung sowohl aufgrund des Auswahlverfahrens als auch aufgrund der typischen Erreichbarkeit bestimmter Bevölkerungsgruppen über Festnetzanschlüsse von ungleichen Auswahl- und Teilnahmewahrscheinlichkeiten auszugehen, was die Verwendung personenspezifischer Gewichte zur Korrektur der ungleichen Wahrscheinlichkeiten bei der Auswertung unerlässlich macht. Grundlage der Gewichtung bildet (1) ein Designgewicht, welches unterschiedliche Auswahlwahrscheinlichkeiten der Befragungspersonen aufgrund von Haushaltsgröße und Erreichbarkeit berücksichtigt und (2) ein Redressmentgewicht zum Ausgleich von Abweichungen der Stichprobenzusammensetzung von bekannten Merkmalsverteilungen in der Grundgesamtheit. Die Berechnung der Gewichte wird im Folgenden erläutert.

8.1 Berechnung der Designgewichte

Bei der Generierung der Designgewichte ist zu berücksichtigen, dass einzelne Personen aus der Grundgesamtheit eine unterschiedlich hohe Wahrscheinlichkeit haben, kontaktiert zu werden. Diese variiert mit der Anzahl der Festnetzanschlüsse, über die die Person potenziell erreicht werden kann sowie, aufgrund des zweiten Auswahltrittes über das Last-Birthday-Verfahren, mit der Anzahl der Personen im Haushalt, die ebenfalls zur Grundgesamtheit gehören. Personen, die über mehrere Festnetzanschlüsse erreichbar sind, haben somit eine höhere Auswahlwahrscheinlichkeit, während z. B. Personen aus Mehrpersonenhaushalten eine geringe Auswahlwahrscheinlichkeit gegenüber Personen aus Ein-Personen-Haushalten aufweisen. Die Auswahlwahrscheinlichkeit ergibt sich demnach aus:

$$\pi_i = \frac{m^F k_i^F}{M^F z_i}$$

wobei:

m^F Zahl der über Festnetznummern realisierten Interviews in der Stichprobe

M^F Nummern im Festnetzauswahlrahmen

k_i^F Zahl der Festnetznummern mit denen der Haushalt erreichbar ist

z_i Anzahl der zur GG gehörenden Personen im Haushalt

Da die Designgewichtung mit der inversen der Auswahlwahrscheinlichkeit erfolgt und der Quotient aus realisierter Stichprobe zur Größe der Grundgesamtheit für alle Stichprobenelemente konstant ist, genügt es mit dem befragtenspezifischen Quotienten (z/k) zu gewichten. Die Gewichte werden dabei anhand der Fallzahl normiert, so dass der Erwartungswert der Gewichte Eins entspricht. Im Ergebnis führt die Designgewichtung dazu, dass Personen aus Single-Haushalten, die über viele Nummern erreichbar sind, ein geringeres Gewicht bekommen, während Personen aus Mehrpersonenhaushalten mit nur einem Festnetzanschluss ein größeres Gewicht bekommen. Um eine zu starke Variation der Gewichte zu vermeiden, wurde als Maximum der berücksichtigten Anschlüsse pro Person das 99-Perzentil festgelegt, was neun Anschlüssen entspricht. Für Personen, die zur Anzahl der Anschlüsse keine Angabe gemacht haben oder machen konnten, wurde das arithmetische Mittel eingesetzt. Bei der Haushaltsgröße erfolgte die Ersetzung fehlender Angaben mit dem Median von zwei Haushaltspersonen.

8.2 Redressmentgewichtung

Es ist aufgrund der forschungspraktisch unabdingbaren Nutzung einer reinen Festnetzstichprobe davon auszugehen, dass bestimmte Bevölkerungsgruppen bereits in der Bruttostichprobe unterrepräsentiert sind, da der Auswahlrahmen die Grundgesamtheit nur unzureichend abdeckt. So legen neuere Untersuchungen nahe, dass der Anteil der sog. Mobile Onlys, also Personen, die nur noch über Mobilfunkanschlüsse zu erreichen sind, in urbanen Zentren inzwischen bei etwa 30 Prozent liegt (infas 2019). Ausschließlich über Mobilfunk erreichbare Personengruppen sind jünger, im ehemaligen Ostteil überproportional vertreten, tendenziell eher männlich, verfügen über geringere Einkommen und haben häufiger keinen Schulabschluss wie eigene Untersuchungen belegen (Schrenker 2019). Dementsprechend ist für die Stichprobe eine Überrepräsentation von Älteren, höher Gebildeten und weiblichen Personen im Vergleich zur Grundgesamtheit der Berliner Wohnbevölkerung zu erwarten. Diese Stichprobenverzerrung sollte gegebenenfalls durch eine geeignete Gewichtung ausgeglichen werden. Für den Abgleich zwischen Stichprobenverteilung und Grundgesamtheit wird im Folgenden die Bevölkerungsfortschreibung des Statistischen Bundesamts für Berlin zum Stichtag 30. 12. 2018 auf Basis des Zensus 2011 genutzt.

Tatsächlich weist ein Vergleich der Stichprobenverteilung mit der Geschlechterverteilung der Bevölkerungsfortschreibung auf eine deutliche Abweichung zwischen Stichprobe und Grundgesamtheit hin. Frauen sind mit fast vier Prozentpunkten überrepräsentiert. Auch bezüglich der Altersverteilung ergeben sich die erwarteten Abweichungen, wobei insbesondere die Gruppe der 18-

bis 39jährigen in der Stichprobe stark unterrepräsentiert ist, während Personen die 58 Jahre und älter sind, stark überrepräsentiert sind.

Tabelle 3: Geschlechterverteilung

	Männer	Frauen
Ist	.4485	.5506
Soll	.4875	.5125

Tabelle 4: Altersverteilung

	18-29	30-39	40-49	50-57	58-64	65-74	75+
Ist	.0774	.0874	.1078	.1326	.2024	.1975	.1949
Soll	.1808	.2007	.1509	.1417	.0964	.1078	.1217

Ein Blick auf die allgemeinbildenden Schulabschlüsse offenbart außerdem, dass Personen mit höheren allgemeinen Schulabschlüssen überproportional an der Befragung teilgenommen haben, während Personen ohne Abschluss oder mit Hauptschulabschluss deutlich seltener teilnahmen. So ist der Anteil der Befragten mit (Fach-)Abitur in der Stichprobe fast 16 Prozentpunkte höher als in der Grundgesamtheit.

Tabelle 5: Verteilung der Schulbildung

	Ohne Abschluss	Hauptschule	Realschule	(Fach-)Abitur
Ist	.0104	.0904	.2725	.6267
Soll	.0523	.2460	.2012	.4673

Auch bei der Zusammensetzung der Stichprobe nach Stadtbezirken ergeben sich einige auffällige Abweichungen: So ist beispielsweise die Bevölkerung Neuköllns mit fast vier Prozentpunkten unterrepräsentiert. Als Ursachen hierfür können der hohe Anteil von Personen mit Migrationshintergrund und häufigere Sprachprobleme, die eine Teilnahme an einer Telefonumfrage verhindern, vermutet werden, sowie die Beliebtheit des Stadtteils bei eher jungen und sozioökonomisch weniger gut ausgestatteten Bevölkerungsgruppen. Eher bürgerliche Viertel wie Steglitz-Zehlendorf oder Charlottenburg-Willmersdorf oder Viertel mit geringem Ausländeranteil wie Treptow-Köpenick sind dagegen stärker in der Stichprobe vertreten.

Tabelle 6: Verteilung nach Stadtbezirken

	Ist	Soll
Mitte	.1064	.1030
Friedrichshain-Kreuzberg	.0689	.0778
Pankow	.1110	.1065
Charlottenburg.-Willmersdorf	.1311	.0942
Spandau	.0553	.0637
Steglitz-Zehlendorf	.1275	.0825
Tempelhof-Schöneberg	.0756	.0946
Neukölln	.0504	.0879
Treptow-Köpenick	.1103	.0724
Marzahn-Hellersdorf	.0594	.0704
Lichtenberg	.0423	.0772
Reinickendorf	.0617	.1030

Durch die Redressmentgewichtung werden die Designgewichte so transformiert, dass sich die Stichprobenverteilungen weitgehend an die Randverteilungen der Grundgesamtheit angleichen. Die beschriebenen Stichprobenverzerrungen sollen damit behoben werden. Alle in diesem Abschnitt berücksichtigten Merkmalsverteilungen fungieren dabei simultan als Soll-Größen. Neben der Bildungsverteilung und der Verteilung nach Bezirken fließt die Altersverteilung differenziert nach Geschlecht in die Gewichtung ein. Die personenspezifischen Redressmentgewichte werden schließlich durch einen iterativen „proportional fitting“ Algorithmus (sog. Raking) gebildet (vgl. Deming und Stephan 1940). Von der Designgewichtung ausgehend werden die Stichprobenanteile in den einzelnen Zellen durch eine Soll-Ist-Gewichtung wiederholt an die Anteile der Grundgesamtheit angepasst, bis sich eine weitgehende Annäherung an die Soll-Größen aus der Grundgesamtheit einstellt. Die resultierenden Gewichte haben einen Erwartungswert von 1 und durften weder kleiner als 0,2 noch größer als 5 ausfallen, da sich mit zunehmender Varianz der Gewichte, die effektive Stichprobengröße verringert.

Bei der Anwendung der Gewichtungsvariablen in Analysen ist Folgendes zu beachten: Während es unerlässlich ist, die Designgewichte zu verwenden, hängt die Brauchbarkeit der Redressmentgewichte davon ab, ob damit tatsächlich die wesentlichen die Teilnahbereitschaft determinierenden Merkmale berücksichtigt wurden. Je nach Fragestellung kann es noch weitere Merkmale geben, die sowohl die Teilnahbereitschaft beeinflussen als auch mit der interessierenden abhängigen Variablen korreliert sind. In diesem Fall müssen gegebenenfalls eigens auf diese Fragestellungen zugeschnittene Gewichtungsvariablen erstellt werden, die die entsprechenden Merkmale berücksichtigen.

9 Abschließende Bemerkungen

Bevor die erhobenen Daten von den Studierenden im Rahmen der Lehrveranstaltung Empirische Sozialforschung ausgewertet werden durften, wurde sichergestellt, dass eine Re-Identifikation einzelner Befragter unter keinen Umständen möglich ist. Hierzu wurden insbesondere offene Angaben zur beruflichen Tätigkeit anonymisiert, wenn diese so spezifisch waren, dass sie Rückschlüsse auf konkrete Personen erlaubt hätten. Der Datensatz wurde außerdem mit Makroindikatoren zur

Mietpreisentwicklung aus Wohnungsmarktreports (CBRE 2013, 2019) und weiteren soziodemografischen Angaben auf Bezirksebene angereichert.

Literatur

- AAPOR*, 2011: Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th Edition. Lenexa, Kansas. The American Association for Public Opinion Research.
- CBRE*, 2019: Wohnungsmarktreport Berlin 2019. Berlin. Berlin Hyp AG.
- CBRE*, 2013: Wohnungsmarktreport Berlin 2013. Berlin. GSW Immobilien AG.
- Deming, W. Edwards, und Frederick F. Stephan*, 1940: On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics* 11: S. 427-444.
- Gabler, Siegfried, und Sabine Häder*, 2009: Die Kombination von Mobilfunk- und Festnetzstichproben in Deutschland. S. 239-252 in: *Weichbold, Martin, Johann Bacher und Christof Wolf* (Hg.), *Umfrageforschung: Herausforderungen und Grenzen*. Österreichische Zeitschrift für Soziologie: Sonderheft, 9. Wiesbaden: VS Verlag.
- infas*, 2019: Regionalisierte Marktforschung: Basisstudie zur Verteilung der Mobile Onlysin Deutschland. <https://infas360.de/service/news/neue-studie-anteil-der-mobile-onlys-liegt-je-nach-region-bei-ueber-40/> (13. 3. 2020).
- Kennedy, Courtney, und Hannah Hartig*, 2019: Response rates in telephone surveys have resumed their decline. <https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/> (3. 3. 2020).
- Sand, Matthias*, 2014: Dual-Frame-Telefonstichproben. Entwicklung, Handhabung und Gewichtung. *GESIS-Technical Reports*. GESIS. Mannheim. GESIS - Leibniz-Institut für Sozialwissenschaften. 2014/02.
- Schrenker, Markus*, 2019: Methodenbericht Ost-Migrantische Analogien. Berlin. Humboldt-Universität.