

Grundlagen der empirischen Sozialforschung

Sitzung 5 - Auswahlverfahren

Jan Finsel

Lehrstuhl für empirische Sozialforschung
Prof. Dr. Petra Stein

17. November 2008

Online-Materialien

- ▶ Die Materialien zur Vorlesung finden Sie auf der Homepage http://www.uni-due.de/soziologie/stein_lehre.php
- ▶ Die ganze Vorlesung ist dort auch als Stream verfügbar
- ▶ Es gibt eine Übung von Dawid Bekalarczyk um 14 bis 16 Uhr am Montag in Raum S-E 005 hier in Essen
- ▶ Im Sekretariat von Frau Werner in Raum R12 R06 A30 können CDs bzw. DVDs erstanden werden
- ▶ Meine Materialien finden Sie auf meiner Seite <http://www.uni-due.de/soziologie/finsel.php>
- ▶ Mail: jan_finsel@uni-due.de

Vorlesungsblock Datenerhebung - sechs Sitzungen

Ablauf

1. Grundlagen der empirischen Sozialforschung - Der Forschungsprozess
2. Definitionen und Hypothesen in der Wissenschaft
3. Messen und Skalieren
4. Forschungsdesigns und Untersuchungsformen
5. Auswahlverfahren und Stichprobendesigns
6. Das Datenerhebungsverfahren der Befragung

Der Plan für heute I

- 1 Struktur des Vorlesungsblockes Datenerhebung
 - Vorlesungsblock Datenerhebung - sechs Sitzungen
- 2 Auswahlverfahren - Definitionen
 - Grundgesamtheit und Untersuchungseinheit
 - Vollerhebung, Teilerhebung und Stichprobe
 - Auswahlregeln für Stichproben
 - Begrifflichkeiten zur Unterscheidung, ob über Stichproben oder die Grundgesamtheit geredet wird
 - Vor- und Nachteile von Voll- und Teilerhebungen
 - Grafik zur Veranschaulichung der Stichprobenziehung
 - Angestrebte Grundgesamtheit, Auswahlgesamtheit und Inferenzpopulation
 - Coverage
- 3 Zufallsstichproben
- 4 Willkürliche und bewusste Auswahlen
 - Bewusste Auswahlen I

Der Plan für heute II

- Bewusste Auswahlen II
- Quotaverfahren
- Kritikpunkte am Quotaverfahren

5 Nonresponse

- Problem von Nonresponse
- Ursachen und Folgen von Unit-Nonresponse

Auswahlverfahren

Verfahren, welche die prinzipiellen Regeln zur Konstruktion von Stichproben angeben

Grundgesamtheit und Untersuchungseinheit

Untersuchungseinheit

Objekt an dem Messungen vorgenommen werden

Grundgesamtheit (Population)

Menge von Elementen (Untersuchungseinheiten), über die im Rahmen einer Untersuchung Aussagen gemacht werden sollen

Schwierigkeit

Für beschreibende Studien ist sie relativ einfach festzulegen. Bei Theorie testenden Fragestellungen ist es oft nicht genau klar, auf welche Objekte sich die Hypothesen beziehen

Vollerhebung, Teilerhebung und Stichprobe

Vollerhebung

Daten zu allen Elementen einer Grundgesamtheit werden erhoben

Teilerhebung

Nur eine Teilmenge von Objekten der Grundgesamtheit wird untersucht

Stichprobe

Die Elemente der Teilerhebung werden durch Regeln bestimmt, die vor der Untersuchung festgelegt werden

Auswahlregeln für Stichproben

Auswahlregeln

können auf zufälligen, willkürlichen oder bewussten Auswahlprozessen fußen

Warnung

Lediglich Zufallsstichproben erlauben die Anwendung inferenzstatistischer Techniken, d. h. nur für sie sind die Fehler, die beim Schluss von der Stichprobe auf die Grundgesamtheit entstehen, berechenbar!

Begrifflichkeiten zur Unterscheidung, ob über Stichproben oder die Grundgesamtheit geredet wird

Statistiken

Maßzahlen, welche sich auf die Stichprobe beziehen

Parameter

Maßzahlen, welche sich auf die Grundgesamtheit beziehen

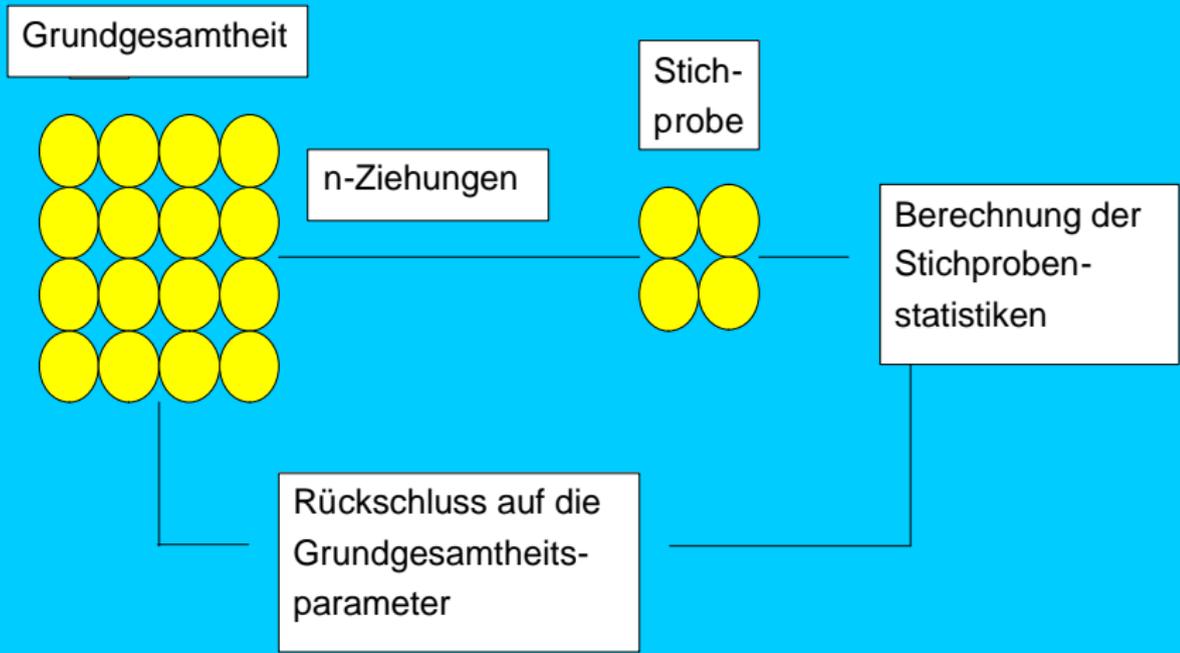
Vor- und Nachteile von Voll- und Teilerhebungen

Vor- und Nachteile der Vollerhebungen

- ▶ Alle Informationen liegen unverfälscht vor
- ▶ Kosten steigen überproportional mit dem Umfang der Population
- ▶ Bei destruktiven Tests unsinnig

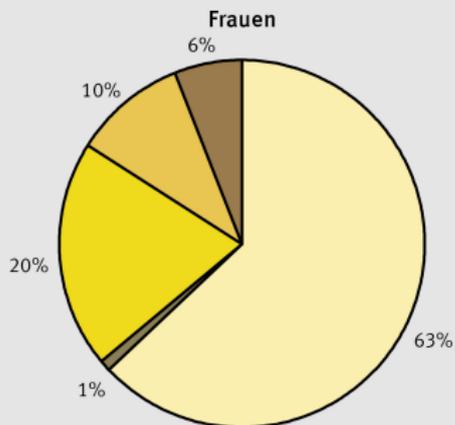
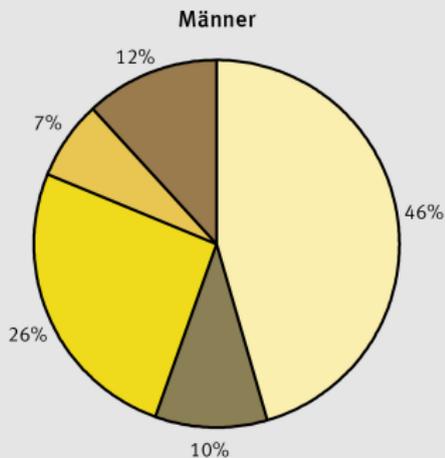
Vor- und Nachteile der Teilerhebungen

- ▶ Relativ kostengünstig
- ▶ Stichprobenergebnisse können genauer sein
- ▶ Fehlerhafter Induktionsschluss beim Ableiten der Populationsparameter aus den Stichprobenstatistiken



Unbezahlte Arbeit nach Arbeitsbereichen

2001/2002



Statistisches Bundesamt 2003 - 02 - 0408

Abbildung: Juhu

Angestrebte Grundgesamtheit, Auswahlgesamtheit und Inferenzpopulation

Angestrebte Grundgesamtheit

Umfasst alle Elemente, für die entsprechende Theorie gelten soll

Auswahlgesamtheit

Umfasst alle Elemente, die eine prinzipielle Chance haben, in eine Stichprobe zu gelangen

Inferenzpopulation

Ist die Grundgesamtheit, über die auf der Basis der vorliegenden Stichprobe tatsächlich Aussagen gemacht werden können bzw. ist die Grundgesamtheit, aus welcher die vorliegende Stichprobe tatsächlich eine Zufallsstichprobe darstellt

Angestrebte Grundgesamtheit und Inferenzpopulation sollten möglichst genau übereinstimmen!

Coverage der zugrunde liegenden Auswahlliste

Undercoverage

Untersuchungseinheiten, die für die Untersuchung relevant sind, fallen aus der Auswahlgrundlage heraus und bleiben somit unberücksichtigt

Overcoverage

Objekte, die gar nicht untersucht werden sollten, haben eine Auswahlwahrscheinlichkeit

Dopplungen

Sind Elemente mehrfach enthalten, erhöht dies ihre Chance in die Stichprobe zu gelangen

Zufallsstichproben I

Einfache Zufallsstichprobe

Jede mögliche Stichprobe und auch jedes Element besitzen dieselbe Chance ausgewählt zu werden

Geschichtete Zufallsstichprobe

Die Elemente der Grundgesamtheit werden so in Gruppen (Schichten, strata) eingeteilt, dass jedes Element der Grundgesamtheit zu einer – und nur zu einer – Schicht gehört. Danach werden einfache Zufallsstichproben aus jeder Schicht gezogen.

Zufallsstichproben II

Klumpenstichprobe

Eine einfache Zufallsauswahl, bei der die Auswahlregeln nicht auf die Elemente der Grundgesamtheit, sondern auf zusammengefasste Elemente (Klumpen, Cluster) angewendet werden und dann jeweils die Daten aller Elemente des ausgewählten Clusters erhoben werden

Mehrstufige Auswahlverfahren

Kombination mehrerer Auswahlverfahren; die Grundgesamtheit wird zuerst in Primäreinheiten eingeteilt, welche die erste Auswahlgrundlage liefern. Aus diesen Primäreinheiten wird dann eine Zufallsstichprobe der Sekundäreinheiten gezogen, die dann bereits die Erhebungseinheiten bilden können. Oft folgen noch weitere Auswahlstufen, welche auf den Sekundäreinheiten (bzw. Tertiäreinheiten usw.) basieren

Willkürliche und bewusste Auswahlen

Willkürliche Auswahlen (Auswahlen aufs Geratewohl)

Die Entscheidung über die Aufnahme eines Elementes der Grundgesamtheit in die Stichprobe erfolgt unkontrolliert und liegt nur im Ermessen des Auswählenden

Bewusste Auswahlen (Auswahlen nach Gutdünken)

Erfolgen zwar nach einem Auswahlplan, d.h. die diesem Plan zugrunde liegenden Kriterien sind sogar meist angebbbar und überprüfbar, dennoch sind inferenzstatistische Techniken nicht anwendbar. Es gibt viele verschiedene Arten bewusster Auswahlen

Bewusste Auswahlen I

Auswahl extremer Fälle

Selektion derjenigen Fälle (z. B. Leistungssportler), die in Bezug auf ein bestimmtes Merkmal eine extreme Ausprägung besitzen

Auswahl typischer Fälle

„Charakteristisch“ für die Grundgesamtheit angesehene Objekte; das Problem liegt in der Angabe der Kriterien, was ein typischer Fall ist

Konzentrationsprinzip

Auswahl derjenigen Fälle, bei denen ein interessierendes Merkmal so stark ausgeprägt ist, dass diese Fälle nahezu die gesamte Verteilung in der Grundgesamtheit bestimmen

Bewusste Auswahlen II

Schneeball-Verfahren

- ▶ Beginnt mit einer Person
- ▶ Diese benennt weitere und zwar aufgrund eines vom Forscher bestimmten Kriteriums, wie z. B. „Wen würden sie in beruflichen Angelegenheiten um Hilfe bitten?“
- ▶ Diese benannten Personen müssen jetzt ihrerseits angeben wer ihnen aufgrund des Kriteriums in den Sinn kommt

Quotaverfahren

Quotaverfahren

Basieren auf der Auswahl von Personen in der Art, dass bestimmte Merkmale in der Stichprobe exakt in derselben Häufigkeit vorkommen wie in der Grundgesamtheit; die endgültige Auswahl bleibt dem Interviewer vorbehalten.

Lassen sich in solche mit unabhängiger oder kombinierter Quote unterscheiden

Kritikpunkte am Quotaverfahren

Stellen keine Zufallsauswahlen dar; dadurch ist die Verallgemeinerung erschwert, wenn nicht sogar unmöglich

Das Verfahren bedingt eine erhöhte Auswahlwahrscheinlichkeit für Personen, die häufig an ihrem Wohnsitz anzutreffen sind

Verzerrungsgefahr durch die Interviewer, da diese vor allem Personen auswählen werden, die sich kooperativ verhalten

Das Quotaverfahren stellt darum keine Lösung des Problems der Ausfälle dar: Das Problem wird nur verdeckt

Nonresponse

Unit-Nonresponse

Die zu befragende Person antwortet gar nicht

Item-Nonresponse

die Person antwortet nur auf einzelne (u. U. heikle) Fragen nicht

Problem von Nonresponse

- ▶ Unterscheiden sich die Nichtantwortenden in irgendeiner Weise systematisch in den interessierenden Variablen von denen die antworten?
- ▶ Wenn ein Unterschied vorläge, würden die angestrebte und die Inferenzpopulation schon deswegen nicht mehr deckungsgleich sein
- ▶ Dieses Problem besteht nicht, wenn die Ausfälle stichprobenneutral sind

Ausschöpfungsquoten

Je stärker die tatsächlich realisierte (die Anzahl der tatsächlich gezogenen Elemente) von der zu realisierende Stichprobe (die Anzahl der Angedachten Elemente) abweicht, desto größer können die Fehler sein. Je größer der Unit- Nonresponse, desto geringer ist die Ausschöpfungsquote

Ursachen und Folgen von Unit-Nonresponse

Verweigerer

Personen, welche eine Befragung explizit verweigern

Nicht-Befragbare

Personen, welche aufgrund körperlicher Gebrechen oder schwerer akuter Erkrankung aus der Stichprobe ausscheiden

Schwer-Erreichbare

- ▶ Trotz mehrfacher Kontaktversuche nicht an ihrem Wohnsitz angetroffene Personen
- ▶ Offensichtlich hängt das-nicht-erreichbar-sein mit bestimmten anderen Merkmalen des Personenkreises zusammen