

Applied Sampling

Seymour Sudman (1983), S. 145-186

In: Peter Rossi, James Wright, Andy Anderson (Hrsg.): *Handbook of Survey Research, New York u.a. (1983)*

ABSTRACT

1) Neben einleitenden Worten vier Datenpools: CPS (Current Population Survey), GALLUP (continuing U.S. polls), DAS (Detroit Area Study), UDDR (Unfounded Doctoral Dissertation Research).

2) Die Populationswahl ist mit der Forschungsfrage verbunden. Zunächst ist der ‚unit‘ zu wählen wie Individuen, Haushalte etc und dann innerhalb des Units noch durch ‚characteristics‘ die Wahl einzugrenzen. Eine zu enge Definition ist dabei hinderlich und operationalisierbar sollte sie auch sein. So empfiehlt es sich nicht alle gebärfähigen sondern alle Frauen zwischen 12 und 50 als Population zu definieren.

3) Es gibt viele Arbeiten mit kleinem ‚n‘ aber viele werden deswegen auch schon nicht veröffentlicht. Zur Bewertung des Samples von Arbeiten (Population + Samplewahl → Generalisierung) kann durch eine Credibility-Scale auf den Dimensionen i) Generalisierbarkeit [Geographie, Limits-Diskussion, Wahl der Population], ii) Größe, iii) Ausführung und iv) Ressourcen-Nutzung ein Punktwert (x/35) gebildet werden. Beispiele werden aufgelistet und sikutiert.

4) Es gibt randomisierte Sample aus den Populationen oder systematisch gewählte. Random muss Gleichverteilt sein und sollte nach der Wahl gegen-getestet werden bzw. durch erwerbhbare Zufallstabellen garantiert sein. Systematisch wäre jedes i-te Element einer Seite oder Spalte und dann eine zeilenweise Auslesung. Bei Menschen sind beide Verfahren als gleichwertig einzustufen da keine Periodizitäten erwartet werden. Dagegen bei nicht-humanen Samples sollte besser auf die Randomisierung zurück gegriffen werden, weil die Periodizität mit der Systematik zusammen fallen könnte.

5) Es gibt Listen oder Zufallswahlverfahren. Bei Listen ist besonders problematisch DIE Liste zu finden (öffentlich mit Fehlern, kommerziell teuer). Zudem sind selbst wenn Listen vorliegen drei Fehler wahrscheinlich 1. Leere / Unbrauchbare Felder, 2. Duplikate und 3) Ausschlüsse durch die Liste selbst. Lösungen sind zu 1) diese zu ignorieren und den Selektionsabstand zu verkleinern um die Ausfallquote zu decken. Bei 2) sollten beide gewichtet übernommen werden, denn die Interviews sind mit großer Wahrscheinlichkeit nicht identisch und zu 3) sind vorhandene Listen durch weitere Quellen zu stützen. Zufallswahlverfahren inkludieren viele unvergebene Nummern und so ist entweder eine Zufallswahl + Randomisierung der letzten drei Ziffern oder eine Clusterbildung und aufsteigende Änderung der letzten Ziffern zu verwenden. Sollen nur kleine Samples aus großen Listen gewählt werden mit speziellen daraus nicht ableitbaren Kriterien, dann empfiehlt es sich entweder das Sample aus der Gesamtliste um den vermuteten Anteil der kleinen Population in der Liste zu vergrößern oder ein Network-Sampling verfahren, welches z.Z. noch getestet wird.

6) Neben der ungefähren Richtgröße für nationale Studien von 1000+ differieren die Sample-Größen bei regionalen Studien zwischen 100 und 1000 je nach Feld der Fragestellung. Neben den formalen Kriterien bestimmen oft die Geldgeber über die Größe. Bei der Stratifikation als Alternativmethode werden die Zusammensetzungen von Samples über deren Proportionen nachgestellt.

1) Einleitung

INTRODUCTION

- Basisideen über “select and evaluate ... samples” (S. 145).
- 1. Beispiele über gutes und schlechtes Sampling werden gegeben
- 2. Populationsdefinition durch theoretische und operationale Selektion
- 3. Random vs systematic vs short-cut
- 4. Problems of lists, blanks, duplications and omission
- 5. Größe des Samples
- 6. Stratifikation

BASIS

- “no uniform standard of quality that must always be reached by every sample” (S. 146).
- Explorative Daten (kleine n’s) für die “generating hypotheses for later study” (S. 146) vs “large-scale continuing studies used to supply the input for major policy decisions” (S. 146).
- Beispiele
 - o CPS (Current Population Survey)
 - o GALLUP (continuing U.S. polls)
 - o DAS (Detroit Area Study)
 - o UDDR (Unfounded Doctoral Dissertation Research)
 - o “lowest quality samples generally consist of 20-50 respondents” (S. 148) eigentlich nur für “early stages of a research design” (S. 148).

INAPPROPRIATE DESIGN

- Abhängigkeit von der Verfügbarkeit entscheidet über Bewertung des Samples: “depends on how it is to be used and what resources are available” (S. 149).
- PRO-QUAL “Allmost all researchers will agree that a small study well-designed and executed is superior to a large study that has been messed-up” (S. 149) KONTRA-QUAL “to do more than the lowest quality exploratory research “(S. 149) nämlich “Pilot Tests, Exploratory Research, Motivational Research” (S. 148, war alles groß geschrieben).
- “many academics never consider the use of broader and more representative samples, even if resources are available” (S. 149).

BIASED SAMPLES

- “very small and poorly chosen samples may be justified as the first stage” (S. 150) sonst aber nicht.

2) Populationswahl

DEFINING POPULATION

- “sample is ... defined as a subset from a larger population” (S. 150) daher ist es notwendig erst über die Population zu sprechen
- Dabei wichtig: „make explicit the universe they wish to study“ (S. 150) und zwar: “The first step in defining the population is to decide whether it is a population of individuals, households, institutions, transactions, or whatever” (S. 151) und das sind “units” S. 151)
- Manchmal ist es nicht klar, wie zwischen z.B. Individuen und Haushalten abgegrenzt werden kann, weil es eine „fuzzy are in the middle“ (S. 151) gibt. Ausschlaggebend ist die „purpose of the study“ (S. 151).

- Nach der Entscheidung über das Universum der Units folgt die Wahl „what units to exclude“ (S. 151), also z.B. nach Kriterien wie „Geography ... Age of individuals ... Other demographic variables ... Other individual variables ... Household variables“ (S. 151) und das sind die „characteristics“ (S. 152). Bei Organisationen gibt es z.B. minimale Größen u.ä.

OVERDEFINING POPULATIONS

- „avoid overdefining a population if it is not critically necessary“ (S. 152)

OPERATIONAL DEFINITIONS

- Auch sollte darauf geachtet werden (wenn z.B. nach gebärfähigen Frauen gesucht wird) lieber das Alterskriterium heran gezogen werden sollte, denn die Gebärfähigkeit herauszubekommen wäre schon eine Studie an sich – die Population wäre nicht operational!
- Biases bei der Selektion und Bestimmung sind zu notieren und limitieren die Aussagefähigkeit

3) Bewertung von empirischen Arbeiten

SMALL-SCALE

- Die Limitierungen durch kleine n's werden oft nicht mit notiert und z.T. auch gar nicht zur Publikation zugelassen.

CREDIBILITY SCALE

- Dies ist ein Bewertungsbogen für Studien mit kleinen n's anhand der diese bewertet werden können. Faktoren sind 1) Generalisierungsgrad, dazu gehören geographische Verteilung des Samples (0-10 Punkte), Diskussion der Einschränkungen durch das Sample (0-5 Punkte) und Wahl besonderer Populationen (-5/+5 Punkte); 2) Größe: Sample size (0-5); 3) Sample Execution als Response Rate (0-5); 4) Resource use von schlecht bis fair (0-5); Maximale Punktzahl sind somit 35 (vgl. S. 155 Tabelle, 154 ff. Beschreibung). Berechnung erfolgt 1. Durch Verhältnisbildung e.g. $31/53 = .89$ Creditability Score und 2. durch Bewertung der Summe der vier Teilbereiche.

EXAMPLES

- Diskutiert werden ideale Beispiele .89 (31/35) und schlechte Varianten .20 (7/35) und anhand der oben aufgestellten vier Bereiche diskutiert. Dazu sollten wir also in der Diss. auch Stellung nehmen und selbstkritisch einordnen.

4) Von der Population zum Sample

RANDOM SAMPLING

- „procedures are easy and inexpensive“ (S. 163) aber manchmal „not appropriate“ (S. 163). Und zwar dann wenn „either because there is no way to select a simple random sample or because other more complex methods are also much more efficient“ (S. 164).
- Definition: „Simple random Samples ... are those for which (a) the probabilities of selection are equal for all elements“ (S. 164) also Gleichverteilung. Und weiter „(b) sampling is done in one stage with elements of the sample selected independently of one another in contrast to more complex samples“ (S. 164).

RANDOM NUMBERS

- Es folgt ein Geleier darüber, was nun Zufallszahlen eigentlich sind. Nämlich Zahlen ausgewählt nach einem „probability mechanism“ (S. 164). Und zwar sollte man daher

auf gedruckte Zufallszahlen (z.B. der RAND Corporation, 1955) zurückgreifen (vgl. 165) und nicht auf ungleich-verteilte Roulettspiele oder Würfel (vgl. 164).

- Tests zur Überprüfbarkeit von Zufall ist die (Gleich-)Verteilung mit $1/n$ als Wert ohne signifikante Schwankungen oder der Pokertest, bei dem die Zahlen für z.B. imaginäre Wettdurchläufe verwendet werden (vgl. 165).
- Durch die Anzahl der Elemente des Samples wird die Tabelle jeweils anders verwendet und so können zufällige Personen (HH, o.Ä.) ausgewählt werden. Das ist mitunter harte Arbeit bei 1mil Zahlen (vgl. 166).
- „A useful property of simple random samples is that simple random subsamples of a simple random sample is also a simple random sample“ (S. 167).

SYSTEMATIC SAMPLING

- „taking every *ith* case is widely used by professionals“ (S. 167)
- Werden aber z.B. Telefonbuchseiten verwendet, dann könnte mit der Länge der verschiedenen Spalten gearbeitet werden (vgl. S. 167).

SYSTEMATIC VS RANDOM

- „systematic samples are really complex samples with unknown properties ... sometimes better, sometimes worse, than simple random samples“ (S. 169).
- Aber eigentlich sind sie doch gleich: „will be about the same except in very unusual situations of periodicities“ (S. 169) das ist aber wiederum „in the real world it is very unlikely that there would be exact periodicities corresponding to the sampling interval“ (S. 169).
- Es ist darauf zu achten, dass bei „nonhuman populations such as businesses, financial accounts, temperature or climate, or lumber or crop yields there may not only be periodicities in the data but also some linear trends“ (S. 170). Hier ist also das simple random sampling besser als das systematische.

5) Quellen für das Sampling

LISITS: USES AND LIMITATIONS

- Problematischer Punkt ist nun „finding the appropriate list“ (S. 170) um daraus das Sample zu ziehen. Denn diese Liste muss irgendwie auf die Population bezogen werden können.
- In Amerika gibt es ein Zensusbüro welches Daten erstellt aber die Wahl der Grundlage selbst trifft.
- Anfälligkeit öffentlich zugänglicher Daten „greater error over time“ (S. 171) wg. Aktualisierung
- Kommerzialiserte Daten sind aktueller (aber kosten z.T.) (S. 171).

LISTS: BLANKS AND INELIGIBLES

- „three common problems in the use of lists“ (S. 172) und davon ist die erste „may contain blank or ineligible units“ (S. 172). Die anderen sind unten folgend Duplikate und Omissions.
- Die erste Variante trifft z.B. beim systematischen Sampling zu, wenn man aus mehrspaltigen Listen zeilenweise auswählt: „may list name of a business firm or professional office, or in some listings a school, park ... These are ineligible listings“ (S. 172) obwohl eigentlich Individuen gesucht werden. Leer-Räume sind natürlich genau so ungünstig.
- Es gibt naive Lösungen, die aber hohe Biases haben: 1) „to take the next name“ (S. 172) oder 2) „to count down a fixed number of eligible names from the top of the page“ (S. 172). Diese sind beide nicht zu empfehlen „incorrect procedure“ (S. 172).

- Richtige Lösung ist es "They should be ignored and not included" (S. 172) darauf hin muss aber der Sampling-Abstand von Anfang an verkleinert werden!!! Es ist daher mit einer Ausfallquote zu rechnen: „one knows that p% ... is eligible ... the sampling intervall is $i = Np/n$... N is the total size“ (S. 172).

LISTS: DUPLICATIONS

- „often these biases will be very small, they may sometimes be serious“ (S. 174)
- "it is wasteful to discard interviews that have already been collected" (S. 175) und daher sollte die Auswertung gewichtet werden "to weight the results by the inverse of the probabilities of selection" (S. 175).

LISTS: OMISSIONS

- "most serious problem in using lists is the omission of an important fraction of the total population" (S. 175), die Liste bildet also nur einen Teil der Gesamtpopulation ab.
- "usually be ignored, such as typographical and clerical errors" (S. 175) aber einige Listen "exclude important segments of the total population" (S. 176). Z.B. werden Personen ohne Telefon aus Telefonbüchern nicht zu entnehmen sein und auch nicht, die nicht gelistet werden wollen.
- „three alternatives are possible for handling omissions in lists: (a) discarding the list; (b) using the list and ignore omissions; (c) using the list with supplementation" (S. 176) wobei jede Variante ihre vor und Nachteile hat. Vor der Entscheidung sollte daher "the purpose of the study and the quality of the list" (S. 176) getestet werden.
- Die Qualität könnte durch ein "relist a sample of 20 block. If the new listing showed that less than 2 or 3% of the old list was wrong, it could be used" (S. 176).
- Bester Weg ist nach ihnen c) in dem man vorhandene Listen nutzt und diese ergänzt (vgl. 176).

DIRECTORIES vs. RANDOM DIGIT DAILING

- Randomisierte Zufallswählung bei gleichverteilten Zahlen ist zwar tatsächlich randomisiert aber "very large majority of numbers that are not in use at all" (S. 177).
- Vorschlag ist die Wahl von wenigen Zufallshaushalten zu kombinieren mit einer Randomisierung der letzten drei Ziffern. Das hat angeblich große Wirkung gezeigt.
- Weiterer Vorschlag ist die Clusterbildung, so dass eine Nummer als Start gewählt wird und dann aufwärts telefoniert wird, bis eine fixe Zahl gelisteter Nummern des Samples erreicht wurde (vgl. S. 177). So erreicht man auch nicht gelistete.

SCREENING FOR SPECIAL POPULATIONS

- "studies of special populations" (S. 179) ist das Ziel, aber es bestehen nur generelle Listen. Ist der Anteil „p of the total population ... available" (S. 179) dann kann einfach "select a sample $1/p$ times larger than would have been required" (S. 179).
- Geht diese einfache Variante nicht, dann ist z.B. ein Screening von Interesse, bei dem alle Daten kurz auf Relevanz geprüft werden – das ist bei kleinen Populationen aus großen Listen oft sehr schwer (vgl. 179).
- Bei guter geographischer Verteilung oder sehr großen Clustern in Listen gibt es Varianten dies zu lösen, bei sehr kleinen empfiehlt sich Network-Sampling, was derzeit aber noch getestet wird.

- "empirical approach that seeks what sizes have been used" (S. 180) als Daumenregel sozusagen oder der rein formale Weg

SIZES USED

- "national studies, regardless of subject matter, typically have samples of 1000 or more" (S. 180).
- "Regional studies vary" (S. 180) und zwar zwischen 100 (Finanz, Labor), 700 (Behaviour, Attitudes) und 1000 (Medizinisch).
- Pilottests oder kleinere Pretests weisen im Durchschnitt Zahlen von 1500-2500 (National) / 500-1000 (Regional) bei Individuen und Haushalten, sowie 500-100 (National) / 200-500 (Regional) bei Institutionen auf.

STRATIFICATION

- "stratified sampling" (S. 182) für die optimale Größe des 'strata-samples'. Dieses birgt „smallest sampling error and hence the most information“ (S. 182).
- "if group A is 60% of the total population, group B 30% and group C is 10%, then the sampling rate for group C is six times the sampling rate for group A, and three times the sampling rate for group B" (S. 185)