

# Der Berner Stichprobenplan

Ein Vorschlag für eine effiziente Klumpenstichprobe

Ben Jann

ETH Zürich  
jann@soz.gess.ethz.ch

Wissenschaftliche Jahrestagung der ASI und der  
Methodensektion der DGS  
Berlin, 14.-15. Oktober 2005

# Gliederung

- 1 Ausgangslage
- 2 Ergebnisse
  - Der Stichprobenplan
  - Simulation
- 3 Zusammenfassung

# Kostenintensität einfacher Zufallsstichproben

- Einfache Zufallsstichproben weisen eine hohe geographische Granularität auf.
- Dies kann hohe Kosten verursachen. Beispiele:
  - Bei einer Face-to-Face-Befragung müssen die Interviewer in viele Gemeinden fahren, in denen vielleicht nur gerade ein Interview durchzuführen ist.
  - Bei der Erstellung einer Registerstichprobe müssen Registerdaten von sehr vielen Gemeinden beschafft werden.

## Traditioneller Lösungsansatz

- Mit Klumpenstichproben kann eine günstigere Kostenstruktur erreicht werden.
- Es wird dabei z.B. eine Anzahl von Gemeinden mit dem SPPS-Verfahren gezogen (sampling with probabilities proportional to size).  
Sodann werden in jeder ausgewählten Gemeinde  $k$  Zielpersonen für die Befragung ausgewählt ( $k =$  Klumpengrösse)
- **Nachteil:** Eine Klumpenstichprobe hat eine höhere Varianz als ein SRS (simple random sample)  $\Rightarrow$  es werden mehr Fälle benötigt.

## Idee (Fritschi et al. 1976)

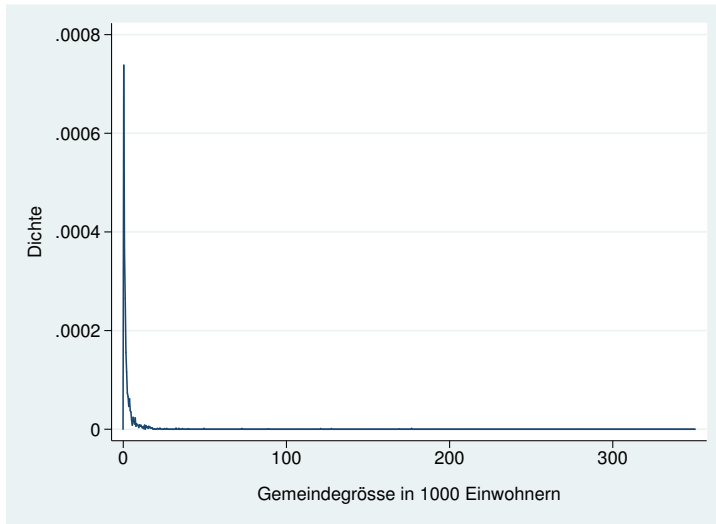
- Wie lassen sich die Eigenschaften einer Klumpenstichprobe verbessern, ohne die Kostenvorteile preiszugeben?
- Lösungsansatz: Eine Klumpung sollte nur da vorgenommen werden, wo sie wirklich notwendig ist.
- Konkret: Die Stichprobe wird zweigeteilt
  - in eine einfache Wahrscheinlichkeitsauswahl bei grossen Gemeinden
  - und eine Klumpenstichprobe bei kleinen Gemeinden.

# Vorgehen

- *Schritt 1:* Generierung einer (hypothetischen) Stichprobe aus der Gesamtpopulation. In Gemeinden mit  $k$  oder mehr Treffern, wird eine der Anzahl Treffern entsprechende Zahl von Zielpersonen ausgewählt.
- *Schritt 2:* Ziehung einer Klumpenstichprobe aus den Gemeinden mit weniger als  $k$  Treffern (SPPS mit reduzierter Auswahlwahrscheinlichkeit).

# Simulation

- Ausgangsdaten:
  - Gemeindeverzeichnis der Schweiz 2002 mit allen politischen Gemeinden und ihren Einwohnerzahlen (2876 Gemeinden, mittlere Einwohnerzahl 2518, Schiefekoeffizient 22.7).
  - Generierte Population mit drei Merkmalen, die in unterschiedlicher Weise mit der Gemeindegrösse korreliert sind ( $X_1: r = 1$ ;  $X_2: r = 0.48$ ;  $X_3: r = 0$ ).
- Ziehung von
  - einfacher Zufallsstichprobe (SRS)
  - Klumpenstichprobe (SPPS)
  - Stichprobe nach dem Berner Stichprobenplan (BSP)
- $n = 1000$ ,  $k = 10$ ; 10000 Replikationen





## Mittelwerte der Simulation

	$\mu$	SRS	SPPS	BSP
X1	0.9758	0.9766 (0.0912)	0.9928 <b>(0.2140)</b>	0.9915 <b>(0.1208)</b>
X2	0.3104	0.3106 (0.0166)	0.3123 <b>(0.0284)</b>	0.3132 <b>(0.0216)</b>
X3	0.4999	0.4998 (0.0159)	0.4999 (0.0158)	0.5001 (0.0159)
Gemeinden		<b>568.7</b>	<b>85.7</b>	<b>86.4</b>
Fälle		999.8	1003.6	999.8

# Design-Effekte

	X1	X2	X3
SPPS vs. SRS	2.3469	1.7069	0.9903
BSP vs. SRS	1.3254	1.2973	0.9996
BSP vs. SPPS	0.5647	0.7600	1.0094

# Zusammenfassung

- Das Verfahren scheint recht erfolgreich zu sein: Die Varianzen werden im Vergleich zur Klumpenstichprobe deutlich verringert, ohne dass die Anzahl Gemeinden in der Stichprobe erhöht würde.
- Aber: Der Erfolg ist abhängig von
  - Verteilung der Gemeindegrößen
  - Zusammenhangsstruktur
- Ausblick
  - Wäre eine Anwendung in Deutschland denkbar?
  - Alternativer Ansatz: Geschichtete Klumpenstichprobe?

# Literatur

- ▶ Fritschi, P., R. Meyer und W. Schweizer (1976) Ein neuer Stichprobenplan für ein gesamtschweizerisches Sample  
*Schweizerische Zeitschrift für Soziologie* 2: 149–158.