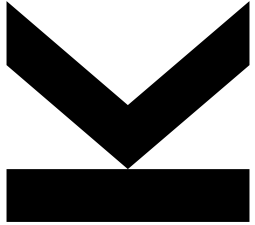


# **KOMPLEXE STICHPROBEN MERKMALE – KONSEQUENZEN – KORREKTE ANALYSE**



Johann Bacher

# ÜBERSICHT

1. Merkmale und Beispiele aus der Sozialforschung
2. Designeffekt und effektive Stichprobengröße
3. Konsequenzen für die Datenanalyse
4. Statistische Verfahren für komplexe Stichproben im Überblick
5. Beispiel MZ
6. SPSS-Makro für Ganninger/Häder/Gabler-Ansatz
7. Anmerkung zum R-Modul Survey
8. Fazit

# MERKMALE KOMPLEXER STICHPROBEN

## Merkmale nach Strugis (2004):

- Mehrere Stufen, z.B. Schulen → Klassen → SchülerInnen
  - Schichtung auf der ersten Stufe (→ Genauigkeitsgewinn, Kostengründe, Mindeststichprobe für kleine Schichten); z.B. PISA Schichtung nach Schulform und Schulgröße, geringere Auswahlwahrscheinlichkeit für kleine Schulen)
  - Gewichtung, da ungleiche Auswahlwahrscheinlichkeiten
- 

- Weitere Merkmale bei Wolters (1985): Schätzmethode, Merkmalsraum, Stichprobengröße

Bacher (2009)

# BEISPIEL PISA

- Schichtung nach Schulform (→ Genauigkeitsgewinn) und Schulgröße (→ Kosten, kann Genauigkeitsgewinn wieder reduzieren)
- größenproportionale Auswahl der Schulen ab bestimmter Größe, bei kleinen und sehr kleinen Schulen geringere Auswahlwahrscheinlichkeit
- Vollerhebung innerhalb einer ausgewählten Schule des Zieljahrgangs, falls Zieljahrgang kleiner / gleich 35 SchülerInnen
- Zufallsauswahl bei mehr als 35 SchülerInnen des Zieljahrgangs (→ Klumpeneffekt sollte nicht zu groß werden)

# BEISPIEL MIKROZENSUS (MZ)

- Schichtung von Privatadressen (=Privathaushalt) nach Bundesland
- Disproportionale Auswahl nach Bundesland (→Gewährleistung einer Mindeststichprobe für Burgenland und Vorarlberg)
- Zufallsauswahl von Privathaushalten
- Vollerhebung aller Mitglieder eines Privathaushalts (direkte Befragung aller Mitglieder ab 15 Jahren)
- Verbleib des Haushalts im MZ für fünf Quartale
- Nachgewichtung der Daten (2014 neues Gewichtungsverfahren)

# DESIGNEFFEKT UND EFFEKTIVE STICHPROBE

... misst den Genauigkeitsgewinn / -verlust einer komplexen Stichprobe im Vergleich zu einer einfachen Zufallsauswahl

$$\text{DEFF}(T) = \frac{\sigma(T)_{\text{komplex}}^2}{\sigma(T)_{\text{SRS}}^2}$$

Bsp.:  $\text{DEFF}(T)=4$  ... 4mal größere Stichprobe erforderlich

$$\text{DEFFSQRT}(T) = \sqrt{\text{DEFF}(T)} = \frac{\sigma(T)_{\text{komplex}}}{\sigma(T)_{\text{SRS}}}$$

Bsp.:  $\text{DEFF}(T)=2$  ... 2mal größeres Vertrauensintervall

$$\text{NEFF}(T) = \frac{n_{\text{komplex}}}{\text{DEFF}(T)}$$

Bsp.:  $\text{NEFF}(T)=1000$  .... Genauigkeit der komplexen Stichprobe entspricht jener einer einfachen Zufallsauswahl von  $n=1000$

Bacher (2009)

# SCHÄTZUNG DES DESIGNEFFEKTS

Designeffekt nach Ganninger/Häder/Gabler (2007; vgl. Bacher 2009):

Durch ungleiche Auswahlwahrscheinlichkeiten  
und Gewichtung bedingter Designeffekt

$$\text{DEFF}(P) = n \frac{\sum w_i^2}{(\sum w_i)^2}$$

$$\text{DEFF}(T) = \text{DEFF}(P) \cdot \text{DEFF}(C)$$

Durch Klumpung (Ähnlichkeit innerhalb einer Primäreinheit)  
bedingter Designeffekt

$$\text{DEFF}(C) = 1 + (b^* - 1) \rho$$

Klumpengröße

Intra-Klassenkorrelation

# EIN EINFACHES RECHENBEISPIEL

HH	Person	X(i)	SST	SSW
1	1	1	0,2500	0,0000
1	2	1	0,2500	0,0000
2	1	0	0,2500	0,0000
2	2	0	0,2500	0,0000
3	1	0	0,2500	0,0000
3	2	0	0,2500	0,0000
4	1	1	0,2500	0,0000
4	2	1	0,2500	0,0000
		0,500	2,0000	0,0000
	n		8	
	M		4	
	rho		1,000	
	b		2,000	
	DEFF		2,000	

Maximale Ähnlichkeit in den Klumpen (=Haushalten)

→ Keine Varianz innerhalb der Klumpen →  $SSW=0$

→ Rho ist daher gleich 1

→  $DEFF = 1 + (2-1)*1 = 2,00$

→  $NEFF = 8/2 = 4$



# KONSEQUENZEN FÜR DIE DATENANALYSE

- Vernachlässigung der Gewichtung → verzerrte Parameterschätzungen
- Vernachlässigung des Designeffekts → verzerrte Schätzungen der Standardfehler der Parameter, i.d.R. Unterschätzung der Standardfehler → Überschätzung der Signifikanz → Fehlschlüsse (H1 wird irrtümlich häufiger akzeptiert)
- Beispiel PISA2003 (Bacher 2009):
  - Mittelwert Österreich:  $\bar{x} = 506$
  - Hypothese: Österreich hat besser abgeschnitten als der OECD-Durchschnitt:  $\mu = 500$
  - Standardfehler:  $\hat{\sigma}_{\text{SRS}} = 1,37$ ;  $\hat{\sigma}_{\text{komplex}} = 3,23$

# KONSEQUENZEN FÜR DIE DATENANALYSE

$$t_{\text{SRS}} = \frac{506 - 500}{1,37} = 4,38 \quad (p < 0,01);$$

$$t_{\text{komplex}} = \frac{506 - 500}{3,23} = 1,86 \quad (p > 0,01);$$

# ANALYSEVERFAHREN

Zugang	Eigenschaften	Software
Designbasierte Verfahren	<ul style="list-style-type: none"> <li>• nur Stichprobendesign fließt in die Schätzung ein</li> <li>• keine weiteren Annahmen erforderlich</li> <li>• erforderlich sind Parameter des Stichprobendesigns (Auswahlwahrscheinlichkeiten usw.)</li> <li>• Parameter oft nicht verfügbar → Schätzungen / Vereinfachungen → modellbasierte Schätzung</li> </ul>	<p>Module "SURVEY" in SPSS, STATA, SAS und R</p> <p>In SPSS für Zusatzmodul CSSAMPLE hohe Lizenzgebühren</p>
Modellbasierte Verfahren	<ul style="list-style-type: none"> <li>• Stichprobendesign geht nur partiell durch Definition der Ebenen und Gewichte ein</li> <li>• weitere Annahmen erforderlich (z.B. Varianzhomogenität in den Klumpen)</li> </ul>	<p>Mehrebenenmodelle in SPSS, STATA, SAS und R</p> <p>SPSS-Makro zu Ganninger/Häder/Gabler für raschen Einstieg</p>

# BEISPIEL MZ (NEET-RATE 2006-2013)

Ungewichtet und ohne Berücksichtigung des Designeffekts

Deskriptive Statistik

	N	Mittelwert	
	Statistik	Statistik	Standardfehler
neet_15_24	177818	,0689	,00060
Gültige Werte (Listenweise)	177818		

Gewichtet, aber ohne Berücksichtigung des Designeffekts

Deskriptive Statistik

	N	Mittelwert	
	Statistik	Statistik	Standardfehler
neet_15_24	177818	,0743	,00062
Gültige Werte (Listenweise)	177818		

Gewichtet und Berücksichtigung des Designeffekts  
Ganninger/Häder/Gabler-Ansatz

Bericht

neet\_15\_24

gg	Mittelwert	Standardfehler des Mittelwertes
1,00	,0743	,00126
Insgesamt	,0743	,00126

$$\begin{aligned} \text{DEFFSQRT} &= 0,00126 / 0,00062 \\ &= 2,032 \\ \text{DEFF} &= 4,109 \\ \text{Rho} &= 0,3598 \\ b^* &= 6,8545 \\ \text{DEFF(C)} &= 3,1066 \\ \text{DEFF(P)} &= 1,3227 \end{aligned}$$

**BEACHTE: Designeffekt für jede Variable unterschiedlich!!!**

# BEISPIEL MZ (NEET-RATE 2006-2013)

Gewichtet und Berücksichtigung des Designeffekts – Ganninger/Häder/Gabler

SRS=0,00062!!!!

Bericht

neet\_15\_24

gg	Mittelwert	Standardfehler des Mittelwertes
1,00	,0743	,00126
Insgesamt	,0743	,00126

DEFFSQRT=2.03

ohne DEFF(P)

DEFFSQRT=1.76

Gewichtet, HLM-Schätzung, lme4 in R

```
Random effects:
Groups   Name             Variance Std.Dev.
asbhh    (Intercept) 0.02562 0.1601
Residual                   0.04160 0.2040
Number of obs: 177818, groups:  asbhh, 35385

Fixed effects:
              Estimate Std. Error t value
(Intercept) 0.073762    0.001018  72.44
```

DEFFSQRT=1.64

Random effects:

```
Groups   Name             Variance Std.Dev.
asbhh    (Intercept) 0.02451 0.1565
Residual                   0.04233 0.2058
Number of obs: 177818, groups:  asbhh, 35385
```

Fixed effects:

```
              Estimate Std. Error t value
(Intercept) 0.072240    0.001036  69.71
xweibl      0.015237    0.001432  10.64
xalter      0.032051    0.001469  21.82
xgeb       -0.039977    0.003720  -10.75
xstaatsb   -0.095679    0.003998  -23.93
```

DEFFSQRT=1.67

# BEISPIEL MZ (NEET-RATE 2006-2013)

## ■ Designbasiertes Vorgehen:

- Stufe 1: Haushalte = PSU, Schichtung nach Bundesland, Zufallsauswahl
- Stufe 2: Personen in Haushalten = SSU, keine Schichtung, Vollerhebung

## ■ STATA14

```
svyset asbhh [pweight=hgew8],  
strata(xnuts2) fpc(xhhnuts2)  
vce(linearized)  
singleunit(certainty) || asbper,  
fpc(xhhszize)
```

```
. svydescribe  
  
Survey: Describing stage 1 sampling units  
  
pweight: hgew8  
VCE: linearized  
Single unit: certainty  
Strata 1: xnuts2  
SU 1: asbhh  
FPC 1: xhhnuts2  
Strata 2: <one>  
SU 2: asbper  
FPC 2: xhhszize
```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
11	2,518	12,654	1	5.0	21
12	4,068	21,093	1	5.2	21
13	4,344	19,654	1	4.5	27
21	3,924	18,865	1	4.8	20
22	4,060	20,084	1	4.9	20
31	4,430	23,162	1	5.2	25
32	3,942	20,017	1	5.1	20
33	4,146	21,725	1	5.2	20
34	3,953	20,564	1	5.2	25
9	35,385	177,818	1	5.0	27

# BEISPIEL MZ (NEET-RATE 2006-2013)

```
. svy linearized : mean neet_15_24
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =      9      Number of obs =   177,818
Number of PSUs  =  35,385      Population size = 1,012,481
                                   Design df      =    35,376
```

	Linearized			
	Mean	Std. Err.	[95% Conf. Interval]	
neet_15_24	.0741982	.0012329	.0717816	.0766148

Note: Strata with single sampling unit treated as certainty units.

Gute Übereinstimmung mit SPSS-  
MAKRO für Ganninger/Häder/Gabler-Ansatz

SE=0,00126

# SPSS-MAKRO FÜR GANNINGER/HÄDER/GABLER

```
compute gg=1.
compute hgew=gewjahr/8.
means tab=hgew by jahr/cells=mean.
include file="c:\bacher\daten\koblbauer\spss_makros\deff_berechnen_v2.sps".
deff_berechnen_v2 var=neet_15_24/gew=wfinal/cluster=asbhh/grp=gg/hochgew=hgew.

deff_berechnen_v2
var=neet_15_24/gew=wfinal/cluster=asbhh/grp=bundesland/hochgew=hgew.
```



### Bericht

neet\_15\_24

gg	Mittelwert	Standardfehler des Mittelwertes
1,00	,0743	,00126
Insgesamt	,0743	,00126

### Mittelwerte

[DatenSet1] C:\Bacher\Daten\koblbauer\daten\_neu\neet\_finall.sav

### Verarbeitete Fälle

	Fälle					
	Eingeschlossen		Ausgeschlossen		Insgesamt	
	N	Prozent	N	Prozent	N	Prozent
neet_15_24 * gg	1012481	100,0%	0	0,0%	1012481	100,0%

### Bericht

Summe

gg	neet_15_24
1,00	75124,29
Insgesamt	75124,29

**Bericht**

neet\_15\_24

bundesland	Mittelwert	Standardfehler des Mittelwertes
1,00 B	,0683	,00617
2,00 K	,0655	,00413
3,00 NÖ	,0724	,00268
4,00 OÖ	,0598	,00237
5,00 S	,0591	,00379
6,00 STM	,0644	,00286
7,00 T	,0617	,00333
8,00 V	,0799	,00536
9,00 W	,1095	,00317
Insgesamt	,0734	,00113

**Bericht**

neet\_15\_24

ajahr Referenzjahr	Mittelwert	Standardfehler des Mittelwertes
2006	,0765	,00292
2007	,0737	,00290
2008	,0744	,00291
2009	,0821	,00317
2010	,0740	,00303
2011	,0715	,00319
2012	,0686	,00304
2013	,0728	,00310
Insgesamt	,0743	,00107

# WEITERES VORGEHEN

- Gewichtung der Daten mit “weight wdeff=w\_alt/deff.” (Wert für deff eingeben!)
  - Compute deff=4.109.
  - Compute wdeff=w/deff.
  - \*w ist bereits vorhandenes Gewicht.
  - Weight by wdeff.
  
- DEFF hängt von der Variablen ab → Wert der Zielvariablen verwenden oder bei mehreren Variablen Median
  - Compute deff1=4.109.
  - Compute deff2=3.890.
  - Compute deff3=4.600.
  - Compute deff4=2.600.
  - Compute deff=median(deff1 to deff4).
  - \*Möglich auch “mean”.

# ANMERKUNG ZUM R-MODUL SURVEY

## R-Modul SURVEY (Lumley 2010)

- Analyse der MZ-Daten durch einstufiges Verfahren mit Zurücklegen oder als zweistufiges Verfahren
- Gute Approximation durch einstufiges Verfahren bei Beispiel von Lumley (2010: 44-45)
- Einstufiges Verfahren → Rechenzeit akzeptabel
  - `onestage<-svydesign(id=~asbhh, strata=~xnuts2, weight=~hgew, data=mzdat, nest=TRUE)`
- Zweistufiges Verfahren → lange Rechenzeit
  - `options(survey.lonely.psu = "remove")`
  - `twostage<-svydesign(id=~Rasbhh+Rasbper, fpc=~xhhnuts2+xhhszize, strata=~xnuts2, weight=~hgew, data=mzdat, nest=TRUE)`

# ANMERKUNG ZUM R-MODUL SURVEY

ONESTAGE: > svymean(~neet\_15\_24,onestage)

	mean	SE
neet_15_24	0.074198	0.0012

TWOSTAGE: > svymean(~neet\_15\_24,twostage)

	mean	SE
neet_15_24	0.074198	0.0012

SRS: > svymean(~neet100+neet\_15\_24,srs\_design)

	mean	SE
neet100	7.428653	0.0674
neet_15_24	0.074287	0.0007

# FAZIT

- Komplexe Stichproben sollten nicht mit Standardstatistikverfahren analysiert werden.
  - Verzerrte Parameterschätzung (wenn keine Gewichtung)
  - Verzerrte Schätzung der Standardfehler (i.d.R. Unterschätzung)
  - Inkorrekte Inferenz (i.d.R. Überschätzung der Signifikanz)
- Komplexe Stichproben können modell- und/oder designbasiert analysiert werden.
- Für modell- und designbasierte Analysen steht Software zur Verfügung.
- Einarbeitsaufwand ist gering.
- Rascher Einstieg durch Makro für SPSS

# LITERATUR

- Bacher, J., 2009: Analyse komplexer Stichproben. In: Weichbold, M., Bacher, J., Wolf, Chr. (Hg.): Umfrageforschung. Herausforderungen und Grenzen. Wiesbaden, 253-273.
- Lee, E. S., Fortofer, R. N., 2006: Analyzing Complex Survey Data. Thousand Oaks.
- Lumley, T., 2010: Complex Surveys. A Guide to Analysis Using R. New Jersey.
- Quatember, A., 2015: Datenqualität in Stichprobenerhebungen. 2. Auflage. Berlin-Heidelberg.
- Sturgis, P., 2004: Analysing Complex Survey Data: Clustering, Stratification and Weights. Social Research Update, Issue 43, University of Surrey.
- Wolter, K. M., 1985: Introduction to Variance Estimation. New York.