ÖSG

Korrekte Analyse komplexer Stichproben

Wien, 22. Oktober 2009

Johann Bacher

Johannes Kepler Universität Linz

Linz 2009

- Was sind komplexe Stichproben?
- Warum sind Standardverfahren zur Analyse komplexer Stichproben nicht geeignet?
- Welche Verfahren sind geeignet?
- Welche Alternativen gibt es zur statistischen Signifikanz?

1. Komplexe Stichproben

= mehrstufige i.d.R. geschichtete Auswahlverfahren, für welche die Annahme einer einfachen Zufallsauswahl aus einer großen Grundgesamtheit nicht gilt

Beispiele

- Mikrozensus (Haslinger/Kytir 2005; Stadler 2005)
- PISA (Programme for International Student Assessment, OECD 2005b; Schreiner u.a. 2007)
- PIRLS (Progress in International Reading Literacy Study; Mullis et al. 2007; Suchan u.a. 2007)
- Öibf-Bildungsstudie (Schlögl/Lachmayr 2004; Bacher/Beham/Lachmayr 2008)

Abbildung 1: Beispiel eines komplexen Stichprobenplans

SchülerInnen	Eltern	
Übergang in die Sekundarstufe I		
	15x VS 4. Klasse	
nicht befragt	15x HS 1. Klasse	
	15x AHS 1. Klasse	15x HS 1. Klasse Nahbereich
Übergang in die Sekundarstufe II		
15x HS 4. Klasse	15x HS 4. Klasse	
15x AHS 4. Klasse	15x AHS 4. Klasse	
15x AHS 5. Klasse	15x AHS 5. Klasse	
15x BMS 1. Klasse	15x BMS 1. Klasse	
15x BS/PT 1. Klasse	15x BPS/PT 1. Kl.	
15x BHS 1. Klasse	15x BHS 1. Klasse	
Übergang in den Tertiärbereich		
15x BHS 5. Klasse	15x BHS 5. Klasse	
15x AHS 8. Klasse	15x AHS 8. Klasse	

Quelle: Eigendarstellung öibf, entnommen aus Bacher/Beham/Lachmayr (2008: 69)

Merkmale (Sturgis 2004)

- mehrstufige Auswahl mit Schichtung und Klumpung
- Gewichtung (Ursachen: ungleiche Auswahlsätze und Antwortausfälle/unit-nonresponse)

Merkmale (Wolter 1985)

- degree of complexity of sample design
- degree of complexity of sample estimator
- multiple characteristics of variables of interest
- descriptive and analytical uses of the survey data
- the scale or size of survey

2. Analyse komplexer Stichproben mit Standardprogrammen?

- falsche Schätzung von Parametern (z.B. Mittelwert, Standardabweichung, Regressionskoeffizient), wenn keine Gewichtung erfolgt
- falsche Schätzung der Standardfehler von Parametern
- mehrstufiges Verfahren i.d.R. Unterschätzung des Standardfehlers = Genauigkeitsverlust (Ursachen: Klumpeneffekt stärker als Schichtungseffekt, relativ große Klumpen bzw. relative wenige Primäreinheiten auf Stufe 1)
- Unterschätzung des Standardfehlers → Überschätzung der Signifikanz → inhaltliche Fehlschlüsse

Tabelle 1: Konsequenzen inkorrekter Behandlung

	einfache Zufallsauswahl	komplexe Stichprobe
Mittelwert in Mathe ohne Gewichtung	512	512
Mittelwert in Mathe mit Gewichtung	506	506
Standardfehler	1,37	3,23
t-Teststatistik für 500	$t = \frac{506 - 500}{1{,}37} = 4{,}38$	$t = \frac{506 - 500}{3{,}23} = 1{,}86$
p(einseitig)	0,000	0,0316
p(zweiseitig)	0,000	0,0632

vgl. Bacher (2008)

3. Messung des Genauigkeitsverlusts bzw. –gewinns

Abbildung 2: Kennzahlen für komplexe Stichproben

Designeffekt: $\mathrm{DEFF}(T) = \dfrac{\sigma(T)^2_{\text{komplex}}}{\sigma(T)^2_{\text{einfach}}}$	$\mathrm{DEFF}(T) = \dfrac{3{,}23^2}{1{,}37^2} = 5{,}56$
$\mathrm{DEFFSQRT}(T) = \sqrt{\dfrac{\sigma(T)^2_{\text{komplex}}}{\sigma(T)^2_{\text{einfach}}}} = \dfrac{\sigma(T)_{\text{komplex}}}{\sigma(T)_{\text{einfach}}}$	$\mathrm{DEFFSQRT}(T) = \dfrac{3{,}23}{1{,}37} = 2{,}36$
effektive Stichprobengröße $\mathrm{NEFF}(T) = \dfrac{n(T)_{\text{komplex}}}{\mathrm{DEFF}(T)}$	$\mathrm{NEFF}(T) = \dfrac{4597}{5{,}56} = 827$

$$\sigma(T)^2_{SRS} = \frac{\sigma(T)^2}{n} = \frac{\sigma(T)^2_{bPSU} + \sigma(T)^2_{wPSU}}{n}$$

$$\sigma(T)^2_{Complex} = \frac{\sigma(T)^2_{bPSU}}{n_B} + \frac{\sigma(T)^2_{wPSU}}{n = n_B \cdot n_w} = \frac{\sigma(T)^2_{bStrat} + \sigma(T)^2_{wStrat}}{n_B} + \frac{\sigma(T)^2_{wPSU}}{n = n_B \cdot n_w}$$

bei Schichtung

$$\sigma(T)^2_{Complex} = \frac{\sigma(T)^2_{bPSU}}{n_B} + \frac{\sigma(T)^2_{wPSU}}{n = n_B \cdot n_w} = \frac{\sigma(T)^2_{bStrat} + \sigma(T)^2_{wStrat}}{n_B} + \frac{\sigma(T)^2_{wPSU}}{n = n_B \cdot n_w} - \frac{\sigma(T)^2_{bStrat}}{n_B}$$

$$DEFF(T)=\frac{\dfrac{\sigma(T)^2_{bStrat}+\sigma(T)^2_{wStrat}}{n_B}+\dfrac{\sigma(T)^2_{wPSU}}{n=n_B\cdot n_w}-\dfrac{\sigma(T)^2_{bStrat}}{n_B}}{\dfrac{\sigma(T)^2_{bPSU}+\sigma(T)^2_{bPSU}}{n}}$$

$$=\frac{n_w\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}-n_w\frac{\sigma(T)^2_{bStrat}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}$$

$$=\frac{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}+n_w\sigma(T)^2_{bPSU}-\sigma(T)^2_{bPSU}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}-n_w\frac{\sigma(T)^2_{bStrat}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}$$

$$=\frac{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}+\frac{(n_w-1)\sigma(T)^2_{bPSU}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}-n_w\frac{\sigma(T)^2_{bStrat}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}$$

$$=1+(n_w-1)\underbrace{\frac{\sigma(T)^2_{bPSU}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}}_{\rho}-n_w\underbrace{\frac{\sigma(T)^2_{bStrat}}{\sigma(T)^2_{bPSU}+\sigma(T)^2_{wPSU}}}_{\rho_s}=1+(n_w-1)\rho-n_w\rho_s$$

$$DEFF(T) = \frac{\sigma(T)^2_{Complex}}{\sigma(T)^2_{SRS}} = -\left[1 + (\underbrace{n/n_B}_{n_w} - 1)\cdot\rho - (\underbrace{n/n_B}_{n_w})\cdot\rho_S\right] = 1 + (n_w - 1)\cdot\rho - n_w\cdot\rho_s$$

n_w = durchschnittliche Klumpengröße

n_B = Zahl der Primäreinheiten (Klumpen)

ρ = Intraklassenkorrelation (Klumpeneffekt, „Homogenität“ innerhalb der Klumpen, z.B. Schulen)

ρ_s = Intraschichtenkorrelation (Schichtungseffekt, „Homogenität“ der Schichten)

Einige Sonderfälle:

- Keine Schichtung ρ_s=0; kein Klumpeneffekt ρ=0 =>
  $DEFF(T) = 1 + (n_w - 1) \cdot \rho - n_w \cdot \rho_s = 1$ = kein Genauigkeitsverlust oder –gewinn
- Keine Schichtung ρ_s=0; starker Klumpeneffekt ρ = 1 = > $DEFF(T) = 1 + (n_w - 1) \cdot \rho$ = starker Verlust, insbes. wenn n_w sehr groß ist, nur ein Klumpen =>
  $DEFF(T) = 1 + (n - 1) \cdot \rho$; im Extremfall ist $DEFF(T) = 1 + (n - 1) \cdot \rho = 1 + (n - 1) \cdot 1 = n$
- Schichtung ρ_s>0; Klumpeneffekt gleich Schichtungseffekt ρ=ρ_s =>
  $DEFF(T) = 1 + (n_w - 1) \cdot \rho - n_w \cdot \rho_s = 1 - \rho$ = Genauigkeitsgewinn

4. Korrekte Gewichtung

Zweistufiges Auswahlverfahren: $w_{hij} = \frac{1}{p_{hi}} \cdot \frac{1}{p_{j/hi}}$

Zusätzliche Adjsutierung: $w_{hij} = \frac{1}{f_{hi} \cdot p_{hi}} \cdot \frac{1}{f_{j/hi} \cdot p_{j/hi}}$

gelegentlich Reskalierung auf n: $w^*_{hij} = \frac{n}{N} \cdot w_{hij}$

Abbildung 3: Beispiel für Berechnung der Gewichte

Strata	School	Sample frame	p1	Sample frame within school	p2	weights	“wrong” weights (a)
1	1	3 of 150	0,02	5 of 20	0,25	200,0000	200,0000
1	2	3 of 150	0,02	5 of 25	0,20	250,0000	200,0000
1	3	3 of 150	0,02	5 of 15	0,33	150,0000	200,0000
2	1	2 of 25	0,08	5 of 18	0,277	45,0000	50,0000
2	2	2 of 25	0,08	5 of 22	0,227	55,0000	50,0000

(a) Gewichtung zur Beseitigung der Disproportionalität:

	Grundgesamtheit	Stichprobe	Gewicht
Schicht 1	3.000 (85,7%)	15 (60%)	200
Schicht 2	500 (14,3%)	10 (40%)	50
Gesamt	3.500	25	

5. Berechnung von Standardfehlern

- Explizit oder mit Linearisierung nach Taylor (→ SPSS, STATA …)
- BRR-Methode (→ PISA, OECD 2005a; STATA)
- Jackknife-Verfahren (→ PIRLS, IEA2008; STATA)

→ Lee/Forthofer (2008)

Für zweistufige Auswahl lässt sich V(T) darstellen als (siehe SPSS-Algorithms):

$$V(T) = V_2(T) = \underbrace{\sum_{h=1}^{H} U_h}_{V_1(T)} + \sum_{h=1}^{H} \sum_{i=1}^{n_h} \pi_{hi} \sum_{k=1}^{K_{hi}} U_{hik}$$

$T = \hat{Y}$ = Merkmalssumme

U_h = Stichprobenvarianz von Y der Primäreinheiten (z.B. Schulen) innerhalb der Schicht h auf Stufe 1

π_{hi} =Auswahlwahrscheinlichkeit der Primäreinheit i der Schicht h

U_{hik} = Stichprobenfehler der Sekundäreinheit k (z.B. Schüler/innen) der Primäreinheit i der Schicht h

Für dreistufige Auswahl gilt:

$$V(T) = V_3(T) = V_2(T) + \sum_{h=1}^{H}\sum_{i=1}^{n_h}\pi_{hi}\sum_{k=1}^{K_{hi}} f_{hik}\sum_{j=1}^{n_{hik}}\sum_{l=1}^{L_{hikj}} U_{hikjl}$$

Formel gelten auch für Mittelwert, wenn mit

$$z_{hij} = w_{hij}\left(y_{hij} - \hat{\bar{Y}}\right) / \hat{N}$$

gerechnet wird.

Verwendung von Spezialsoftware

- WesVar (WESTAT 2008)
- SUDAAN (RTI-International 2008; Lee/Forthofer 2006).

Zusatzmodule von Standardstatistikprogrammen

- ComplexSample (SPSS Inc. 2008)
- PROC SURVEY (SAS Institute 2008)
- SURVEY METHODS (StataCorp 2008)
- SVY in R (Lumley 2003)

Spezialprogramme, z.B. Mehrebenenmodelle (HLM), LatentGold, …

Abbildung 4: Vergleich von COMPLEX SAMPLE in SPSS und SURVEY METHODS in STATA

	STATA9	SPSS14
Stichprobenziehung:	nein	ja
Varianzschätzmethoden:		
• Taylor-Linearisierung	ja	ja
(ungleiche Auswahlwahrscheinlichkeiten auf Stufe I)	nein	ja
• BBR-Verfahren	ja	nein
• Jackknife-Verfahren	ja	nein
Statistische Verfahren		
• (relative) Häufigkeiten	ja	ja
• Mittelwert	ja	ja
• Verhältniszahlen	ja	ja
• Tabellenanalyse inkl. Unabhängigkeitstest	ja	ja

	STATA9	SPSS14
• t-Test für theoretischen Mittelwert	nein (a)	nein (a)
• t-Test für unabhängige Stichproben	nein (b)	nein (b)
• t-Test für abhängige Stichproben	nein (c)	nein (c)
• Korrelationskoeffizienten	nein (d)	nein (d)
• Allgemeines lineares Modell	nein	ja
• lineare Regression	ja	nein (e)
• logistische Regression	ja	ja
• Regressionsmodell für ordinale Variablen	ja	ja
• weitere Spezialverfahren zur Regressionsanalyse	ja (f)	nein

(a) kann einfach geprüft werden, indem eine neue Variable y* mit y* = y - μ mit μ = theoretischer Mittelwert gebildet und getestet wird, ob diese von Null verschieden ist.

(b) kann mittels einfacher Regression geprüft werden. Die Gruppierungsvariable (z.B. Geschlecht) ist die unabhängige Variable, die Untersuchungsvariable (z.B. Punkte bei einem Test) die abhängige. Die

Mittelwertdifferenz ist dann gleich dem nicht standardisierten Regressionskoeffizienten, die Signifikanz der Mittelwertdifferenz gleich der Signifikanz des Regressionskoeffizienten.

(c) kann einfach geprüft werden, indem eine Variabel $d = y_1 – y_2$ mit y_1 = erste Messung und y_2 = zweite Messung gebildet und getestet, ob diese von Null verschieden ist.

(d) Signifikanzen von Korrelationen können mittels der einfachen Regression für folgende Korrelationskoeffizienten berechnet werden: Pearsonscher Korrelationskoeffizient r, Phi, punktbiserale Korrelation.

(e) Submodell des allgemeinen linearen Modells

(f) z.B. Probit-Regression, Intervall-Regression, Poisson-Regression usw. (StataCorp. 2005)

6. COMPLEX SAMPLE - gleiche Auswahlwahrscheinlichkeiten innerhalb der Schichten

```
get file="d:\tarnai\beispiel1roh.sav".

*Komplexes Design.

compute p1=-9.
if (schicht = 1) p1=3/150.
if (schicht = 2) p1=2/25.
fre var=p1.

compute p2=-9.
if (schule = 1) p2=5/20.
if (schule = 2) p2=5/25.
if (schule = 3) p2=5/15.
if (schule = 4) p2=5/18.
```

```
if (schule = 5) p2=5/22.
fre var=p2.

compute wtot=(1/p1)*(1/p2).
weight by wtot.

des var=weiblich testscore/stat=mean stddev se.


* Analysevorbereitungsassistent.
CSPLAN ANALYSIS
 /PLAN FILE='d:\tarnai\beispiel1roh'
 /PLANVARS ANALYSISWEIGHT=wtot
 /PRINT PLAN
 /DESIGN  STRATA= Schicht  CLUSTER= Schule
   /ESTIMATOR TYPE=EQUAL_WOR
   /INCLPROB VARIABLE= p1
 /DESIGN CLUSTER= idnr
   /ESTIMATOR TYPE=EQUAL_WOR
   /INCLPROB VARIABLE= p2.
```

```
* Deskriptive Statistiken für komplexe Stichproben.
CSDESCRIPTIVES
 /PLAN FILE = 'beispiel1roh'
 /SUMMARY VARIABLES =weiblich Testscore
 /MEAN
 /STATISTICS SE count popsize cin
 /MISSING SCOPE = ANALYSIS CLASSMISSING = EXCLUDE.
```

		Schätzung	Standardfehler	95%-Konfidenzintervall		Umfang der Grundgesamtheit	Ungewichtete Anzahl
		Untere Grenze	Obere Grenze	Untere Grenze	Obere Grenze	Untere Grenze	Obere Grenze
Mittelwert	weiblich	,5429	,05933	,3541	,7317	3500,000	25
	Testscore	458,98	22,443	387,55	530,40	3500,000	25

Tabelle 2: Ergebnisse aus multivariater Analyse

	b (unstand.)	ß (stand.)	t (einfach)	p (einfach)	t (komplex)	p (komplex)
(Konstante)	0,31		29,75	0,000	15,98	0,000
BUB	0,00	0,00	0,20	0,841	0,19	0,849
WALLEIN	-0,01	-0,01	-0,23	0,816	-0,16	0,871
VERANT	0,03	0,04	1,70	0,089	1,43	0,163
SCHICHT	*0,02*	*0,10*	*3,94*	*0,000*	*2,93*	*0,006*
AHS_Nähe	*0,18*	*0,19*	*8,32*	*0,000*	*3,06*	*0,004*
MATURA	*0,30*	*0,32*	*8,68*	*0,000*	*6,62*	*0,000*
SCHLEIST	*0,19*	*0,35*	*9,37*	*0,000*	*7,68*	*0,000*
Interaktionen						
BUB*WALLEIN	0,06	0,02	1,00	0,316	1,00	0,326
BUB*VERANT	0,01	0,01	0,30	0,763	0,29	0,777
BUB*SCHICHT	0,01	0,04	1,47	0,142	*1,92*	*0,064*
BUB*AHS_Nähe	0,06	0,03	1,34	0,182	1,26	0,216
BUB*MATURA	-0,07	-0,05	-1,35	0,177	-1,62	0,116
BUB*SCHLEIST	-0,03	-0,04	-1,02	0,310	-1,09	0,283

7. Analyse von PISA – Größenproportionale Auswahl innerhalb der Schichten (PPS)

Berechnungsformeln müssen modifiziert werden: An Stelle von

$$U_h^2 = (1 - f_h) \cdot \frac{n_h}{n_h - 1} \cdot \sum_{i=}^{n_h} (Y_{hi} - \bar{Y}_h)^2$$

wird die Stichprobenvarianz je Schicht h wie folgt berechnet:

$$U_h^2 = \sum_{i=1}^{n_h} \sum_{j>i}^{n_h} \left(\frac{p_{hi} \cdot p_{hj}}{p_{hij}} - 1 \right) \cdot (Y_{hi} - Y_{hj})^2 \text{ ,mit}$$

$$Y_{hi} = \sum_{k=1}^{n_{hi}} \underbrace{w_{hik} \cdot y_{hik}}_{Y_{hik}} = \sum_{k=1}^{n_{hi}} Y_{hik} = \text{hochgerechnete Merkmalssumme}$$

p_{hi} = Auswahlwahrscheinlichkeit der Schule i in der Schicht h

p_{hj} = Auswahlwahrscheinlichkeit der Schule j in der Schicht h

p_{hij} = Auswahlwahrscheinlichkeit, dass Schule i und j gemeinsam ausgewählt werden.

Die Berechnung der gemeinsamen Auswahlwahrscheinlichkeiten p_{hij} ist aufwendig. SPSS-COMPLEX SAMPLE setzt voraus, dass sie bekannt sind. Wird die Stichprobe mittels SPSS-COMPLEX SAMPLE gezogen, werden die gemeinsamen Auswahlwahrscheinlichkeiten von SPSS berechnet, andernfalls müssen sie eingegeben werden.

Hinzukommt, dass die Berechnung wegen der Division mit p_{hij} instabil werden kann, wenn durch sehr kleine Zahlen dividiert wird.

Wolter (1985) empfiehlt folgende Vorgehensweise, wenn eine systematische PPS gezogen wird (zufällige Startzahl mit gleichem Auswahlabstand):

- Berechnung der gemeinsamen Auswahlwahrscheinlichkeiten mittels einer Approximation von Hartley und Rao (Wolter, 1985: 287):

$$p_{hij} = \frac{n-1}{n} p_{hi} p_{hj} + \frac{n-1}{n^2}\left(p_{hi}^2 p_{hj} + p_{hi} p_{hj}^2 \right) - \frac{n-1}{n^3} p_{hi} p_{hj} \sum_{k=1}^{N} p_{hk}^2$$

Nachteil: Die p_{hi} müssen für alle Elemente der Grundgesamtheit bekannt sein.

(Weitere Approximationen in Berger (2004)).

- Annahme einer Auswahl mit ungleichen Auswahlwahrscheinlichkeiten, aber mit Zurücklegen (= Hansen-Hurwitz-Schätzer, siehe Berger (2004: 310)),
- wie 2, nur mit zusätzlicher Endlichkeitskorrektur

OECD: BRR-Gewichte

Abbildung 5: BRR-Methode

Table 3.13 ■ The Fay replicates

Pseudo-stratum	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R 10	R 11	R 12
1	1	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5	1.5
1	2	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5	0.5
2	3	1.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5	0.5
2	4	0.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5	1.5
3	5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5	1.5
3	6	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5	0.5
4	7	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5	1.5
4	8	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5	0.5
5	9	1.5	1.5	1.5	0.5	1.5	0.5	0.5	1.5	0.5	0.5	0.5	1.5
5	10	0.5	0.5	0.5	1.5	0.5	1.5	1.5	0.5	1.5	1.5	1.5	0.5

$$\sigma^2_{(\hat{\theta})} = \frac{1}{G(1-k)^2}\sum_{i=1}^{G}(\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{80(1-0.5)^2}\sum_{i=1}^{80}(\hat{\theta}_{(i)} - \hat{\theta})^2 = \frac{1}{20}\sum_{i=1}^{80}(\hat{\theta}_{(i)} - \hat{\theta})^2$$

entnommen aus: OECD (2005a: 50-51)

- In SPSS COMPLEX SAMPLE können derzeit noch keine BBR-Gewichte verwendet werden. (In Stata können BBR-Gewichte genutzt werden. Zudem gibt es Makros für plausible Werte.)
- Verfügbar sind SPSS-Makros. Diese sind aber sehr rechenintensiv!

Tabelle 3: Vergleich BRR mit COMPLEX-SAMPLE

	Statistik	SE-BRR	SE-WR	SE-Equal	SE-SRS
Univariate Statistiken					
Weiblich	0,499	0,015634	0,01893	0,01792	0,00772
Mathematik	506	3,266190	3,60970	3,45771	1,54995
Abhängige = CULTPOSS					
AUT b0	-1,011	0,054747	0,05196822	0,050265	0,05052176
AUT HISEI	0,019	0,001114	0,00107307	0,00103822	0,00092326
AUT migra	-0,134	0,045301	0,04495233	0,04335574	0,04228967
AUT weiblich	0,193	0,036737	0,03986684	0,03815616	0,02920098
abh. Variable = math					
AUT b0	440,837	6,933528	8,24874994	7,92719199	4,85484573
AUT HISEI	1,686	0,123306	0,13474943	0,13045336	0,0879692
AUT migra	-44,496	5,629917	6,17232182	5,89585233	4,08844184
AUT weiblich	-11,829	3,890988	4,44537633	4,18574706	2,74470983

7. HLM – eine Alternative?

- HLM wird häufig zur Analyse der PISA-Daten eingesetzt. HLM eignet sich grundsätzlich auch zur Schätzung der „korrekten“ Standardfehler. Dazu wird die Schichtungsvariable als Kontextvariable eingeführt. Mit HLM können auch plausible Variable analysiert werden.
- Allerdings kann HLM zu verzerrten Parameterschätzern führen, wenn z.B. einige Schulen wenige SchülerInnen (siehe unten; Annahme der Varianzhomogenität verletzt) haben oder die Gewichte der SchülerInnen und Schulen stark streuen (Zaccarin/Donati 2008).

Mathematikleistungen in PISA2003

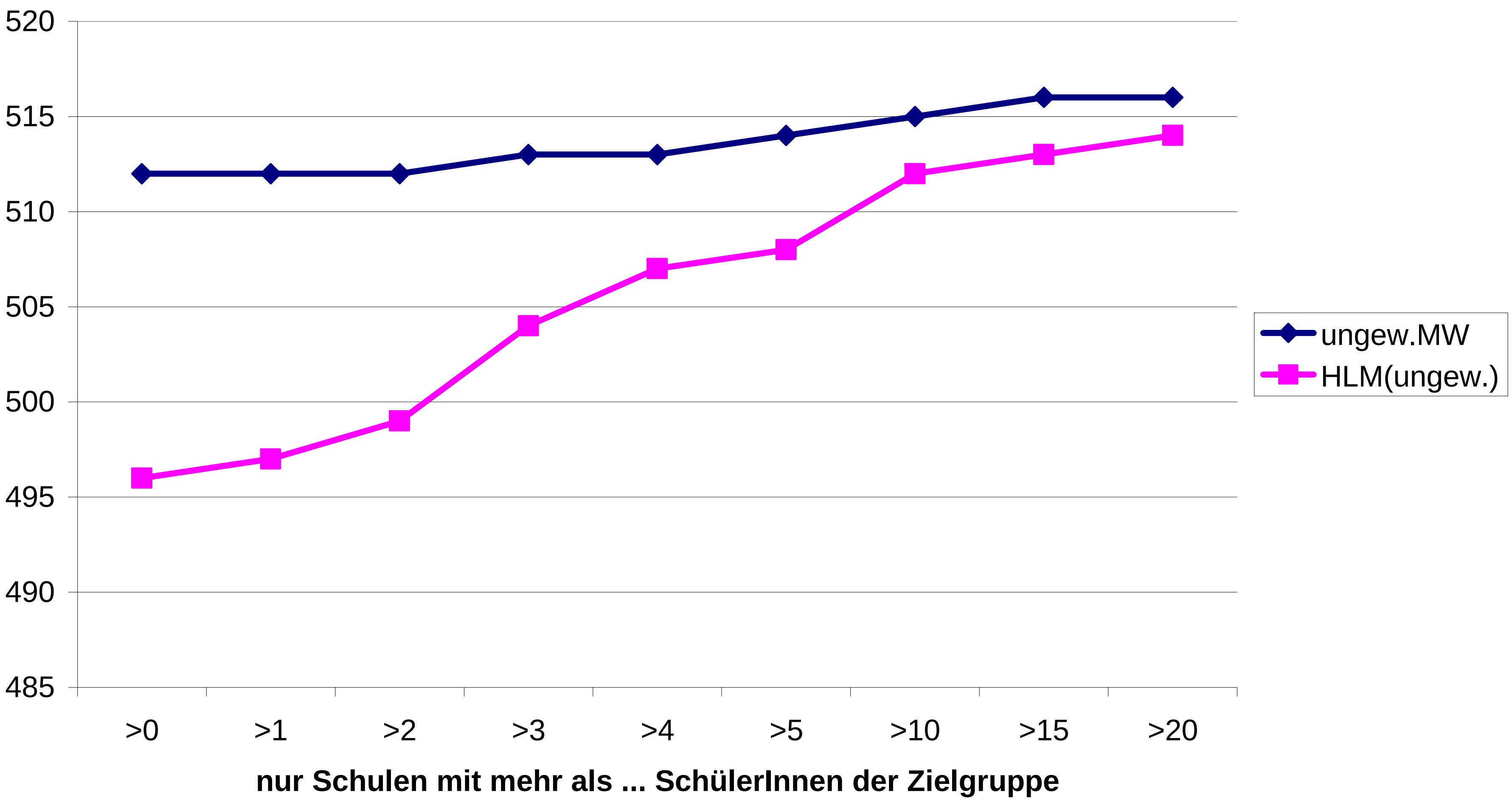

Tabelle 4: Vergleich BRR-Schätzung und HLM

	SPSS-Makro		HLM	
	Statistik	SE-BBR	Statistik	SE
Univariate Statistiken*				
Weiblich	0,505	0,016200	0,4832	0,017385
Mathematik	505,611	3,266190	500,5037	3,491675
Abhängige = CULTPOSS*				
AUT b0	-1,011	0,054747	-0,6646	0,059228
AUT HISEI	0,019	0,001114	0,0119	0,001161
AUT migra	-0,134	0,045301	-0,13478	0,04420
AUT weiblich	0,193	0,036737	0,1230	0,04256
abh. Variable = math*				
AUT b0	444,159	7,1555336	503,901	5,76599
AUT HISEI	1,644	0,127808	0,2297	0,08772
AUT migra	-44,498	5,644853	-38,167	3,70743
AUT weiblich	-12,529	3,967025	-17,004	2,96231

* unter Ausschluss von Fällen mit mind. einem fehlenden Wert in HISEI, MIGRA, WEIBLICH, CULTPOSS, MATH

8. Zusammenfassung und Ausblick

- Korrekte Analyse komplexer Stichproben ist wichtig.
- COMPLEX-Sample bietet dazu einige Möglichkeiten. Volle Ausschöpfung nur möglich, wenn die Stichprobe mit COMPLEX Sample selbst gezogen wird.
- Zur Analyse von PISA-Daten stehen Makros zur Verfügung
- STATA = derzeit bestes Standardprogramm zur Analyse von PISA und PIRLS
- Verbleibendes Problem: Auch bei Berücksichtigung des korrekten Stichprobenplans verbleiben große Stichproben (z.B. bei Verwendung mehrer Länder), so dass immer signifikante Ergebnisse vorliegen

Strategien

- Reduktion der Fallzahl auf eine bestimmte Größe, z.B. auf n=1000 (→ Strategie der OECD bei multivariaten Analysen)
- Analyse einer Substichprobe (Kalibrierungsstichprobe und Prüfstichprobe)
- Verwendung von Verfahren, die große Daten erfordern (Mischverteilungsverfahren, z.B. Analyse latenter Klassen)
- Verwendung von Effektstärken (Cohen 1977) → Psychologie (z.B. Thompson 2007), Medizin (Singh 2006) und andere Disziplinen üblich

Wünschenswert:

genaue Dokumentation des Stichprobenplans, Gewichtung und Informationen zur Berechnung von Standardfehlern

Herzlichen Dank!

Literatur

Bacher, J. 2006: Stichprobendesign, Sozialstruktur und regionale Unterschiede. In E. Neuwirth, E., I. Ponocny & W. Grossmann (Hg.): PISA 2000 und PISA 2003: Vertiefende Analysen und Beiträge zur Methodik (S. 39-51). Graz: Leykam.

Bacher, J., Beham, M., & Lachmayr, N., 2008: Geschlechterunterschiede in der Bildungswahl. Wiesbaden: VS Verlag.

Berger, Y. G., 2004: A Simple Variance Estimator for Unequal Probability Sampling without Replacement. Journal of Applied Statistics, Vol. 31, 305-315

Cohen, J., 1977: Statistical Power Analysis for Behavorial Sciences. Revised Edition. New York ua.

Haslinger, A. & Kytir, J., 2005: Stichprobendesign, Stichprobenziehung und Hochrechnung des Mikrozensus ab 2004. Statistische Nachrichten, 6, 510-518

IEA 2008: IDB-Analyzer. http://pirls.bc.edu/pirls2006/user_guide.html.

Lee, E. S. & Forthofer, R. N., 2006: Analyzing Complex Survey Data. Second Edition. New York: Sage.

Lumley, T., 2003: Analyzing Survey Data in R. R-News, 3(1), 17-20.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M. & Foy, P., 2007: PIRLS 2006 International Report. Boston: IEA (http://pirls.bc.edu/isc/publications.html#p06, 30.6.2008)

OECD (Ed.), 2005a: PISA 2003. Data Analysis Manual. SPSS® Users. Paris: OECD

OECD (Ed.), 2005b: PISA2003. Technical Report. Paris: OECD

Raudenbusch, S. u.a., 2004: HLM6. Scientific Software

RTI-International, 2008: SUDAAN. (http://www.rti.org/sudaan/index.cfm, 30.6.2008)

SAS Institute, 2008: Statistical Analysis with SAS/STAT® Software. (http://www.sas.com/technologies/analytics/statistics/stat/features.html, 30.6.2008)

Schlögl, P., Lachmayr, N., 2004b: Soziale Situation beim Bildungszugang. Motive und Hintergründe von Bildungswegentscheidungen in Österreich. Wien: Eigenverlag.

Schreiner, C., Breit, S., Schwantner, U. & Grafendorfer, A., 2007: PISA2006. Internationaler Vergleich von Schülerleistungen. Die Studie im Überblick. Graz: Leykam.

Singh, G., 2006: A shift from significance test to hypothesis test through power analysis in medical research. Journal of Postgraduate Medicine, Vol. 53, 148-150.

SPSS Inc., 2008: SPSS Complex Samples™ (http://www.spss.com/complex_samples/data_analysis.htm, 30.6.2008).

Stadler, B., 2005: Daten zum österreichischen Arbeitsmarkt. Österreichische Zeitschrift für Soziologie, 30 (3), 89-100.

StataCorp, 2008: Survey Methods. (http://www.stata.com/capabilities/svy.html, 30.6.2008)

Sturgis, P., 2004: Analysing Complex Survey Data: Clustering, Stratification and Weights. Social research UPDATE, Issue 43, University of Surrey.

Suchan B., Wallner-Paschon Chr., Stöttinger, E. & Bergmüller, S., 2007: PIRLS 2006. Internationaler Vergleich von Schülerleistungen. Graz: Leykam

Thompson, B., 2007: Effect sizes, confidence intervals, and confidence intervals for effect sizes. Psychology in the School, Vol. 444, 423-432.

WESTAT, 2008: WesVar – Software for Analysis of Data form Complex Sample. (http://www.westat.com/wesvar/index.html, 30.6.2008)

Wolter, K. M., 1985: Introduction to Variance Estimation. New York u.a.:Springer Verlag.

Zaccarin, S., Donati, Ch., 2008. The effects of sampling weights in multilevel analysis of PISA data. Universita degli studi di Trieste: working paper 119.