Korrekte Analyse komplexer Daten?

Johann Bacher

Johannes Kepler Universität Linz

Linz 2008

IHS Wien, Monday, November 03, 2008, 16:00-18:00 Uhr, SRS

- Was sind komplexe Stichproben?
- Warum sind Standardverfahren zur Analyse komplexer Stichproben nicht geeignet?
- Welche Verfahren sind geeignet?
- Welche Alternativen gibt es zur statistischen Signifikanz?
- Wie stark müssen Zusammenhänge sein? (Oder wie schwach dürfen Zusammenhänge sein?)

1. Komplexe Stichproben

= mehrstufige Auswahlverfahren, für welche die i.i.d.-Annahme (i.i.d. = independent identical distributed) 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)

Merkmale (Sturgis 2004)

- mehrstufige Auswahl mit Schichtung und Klumpung
- Gewichtung (ungleiche Auswahlsätze, Antwortausfälle (unit-nonresponse))

Merkmale (Wolter 1985) → Anhang A1

- 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

Abbildung 1: Beispiel komplexer Stichprobenplan

SchülerInnen	Eltern	
Übergang in die Sekundarstufe I		
nicht befragt	15x VS 4. Klasse 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)

2. Analyse komplexer Stichproben mit Standardprogrammen

- falsche Schätzung des Standardfehlers von Parametern (z.B. Mittelwert, Standardabweichung, Regressionskoeffizient)
- 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

$$\mathrm{DEFF(T)} = \left[1 + (\underbrace{n / n_B}_{n_W} - 1) \cdot \rho - (\underbrace{n / n_B}_{n_W}) \cdot \rho_S \right]$$

→ Anhang A2

n_W = durchschnittliche Klumpengröße

n_B = Zahl der Primäreinheiten (Klumpen)

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

ρS = Intraschichtenkorrelation (Schichtungseffekt, „Homogenität" der Schichten)

Abbildung 2: Beispiel für inkorrekte Behandlung

Mittelwert in Mathe PISA2003	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

Beispiel entnommen aus Bacher (2006)

weitere Beispiele für Fehlinterpretationen

- Regionale Vergleiche („Wien verursacht schlechtes Abschneiden von Österreich in PISA“) → nicht haltbar, wenn komplexes Stichprobendesign berücksichtigt wird
- Zweite Generation erzielt bei PISA2006 schlechtere Testleistungen als die erste Generation → nicht haltbar, wenn komplexes Stichprobendesign berücksichtigt wird

3. Genauigkeitsgewinn und –verlust bei komplexen Stichproben

Abbildung 3: Kennzahlen für komplexe Stichproben

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

Berechnung des Standardfehlers von Parameterschätzungen T

- Explizit oder mit Linearisierung nach Taylor (→ SPSS, STATA …)
- BRR-Methode (→ PISA, OECD 2005a) Anhang A3
- Jackknife-Verfahren (→ PIRLS, IEA2008) Anhang A4

→ Lee/Forthofer (2008)

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

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

$$T = \widehat{Y}$$

U_h = Stichprobenfehler 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:

$$V(T) = V_3(T) = V_1(T) + V_2(T) + V_3(T)$$

Formel gelten auch für Mittelwert, wenn mit

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

gerechnet wird.

4. Korrekte Analyse

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)

Vergleich STATA SPSS → Anhang A5

Mehrebenenmodelle, z.B. HLM (Raudenbush u.a 2004)

5. Beispiel für korrekte Analyse

1	3000 Schüler/innen 150 Schulen	größenproportionale Auswahl von 2 Schulen	zufällige Auswahl von 5 Schüler/innen in jeder Schule
2	500 Schüler/innen 25 Schulen	größenproportionale Auswahl von 2 Schulen	zufällige Auswahl von 5 Schüler/innen in jeder Schule
3	100 Schüler/innen 10 Schulen	größenproportionale Auswahl von 2 Schulen	zufällige Auswahl von 5 Schüler/innen in jeder Schule
	3600 Schüler/innen 185 Schulen	Stichprobe =	30 Schüler/innen 6 Schulen

	idnr	Schule	Schicht	Schüler Schicht GG	Schulen Schicht GG	Schulen Stichprobe	Schüler Schule GG	Schüler Stichprobe	Testscore
1	1	1	1	3000	150	2	25	5	451
2	2	1	1	3000	150	2	25	5	450
3	3	1	1	3000	150	2	25	5	429
4	4	1	1	3000	150	2	25	5	458
5	5	1	1	3000	150	2	25	5	445
6	6	2	1	3000	150	2	15	5	553
7	7	2	1	3000	150	2	15	5	559
8	8	2	1	3000	150	2	15	5	571
9	9	2	1	3000	150	2	15	5	570
10	10	2	1	3000	150	2	15	5	539
11	11	3	2	500	25	2	18	5	127
12	12	3	2	500	25	2	18	5	106

```
GET
  FILE='D:\texte\ihs\beispiel1.sav'.
DATASET NAME DatenSet1 WINDOW=FRONT.

weight off.

fre var=schicht.
des var=testscore/stat=mean stddev semean.
```

	N	Mittelwert		Standardab-weichung
		Statistik	Standardfehler	
Testscore	30	497,71	53,175	291,252

```
*Berechnung der Gewichte und Auswahlwahrscheinlichkeiten.
*Wahrscheinlichkeit P1 für die Auswahl einer Schule
*Formel bei größenproportionaler Auswahl in der Schicht:
*    P 1= Zahl der getesten Schulen * Zahl der Schüler in der Stichprobe
*                      / Schüler Gesamt in der Schicht
*.
compute p1=SchulenStichprobe*SchülerSchuleGG/SchülerSchichtGG.
compute w1=1/p1.

*Wahrscheinlichkeit für die Auswahl eines Schülers/einer Schülerin
*in der Schule i.
compute p2=SchülerStichprobe/SchülerSchuleGG.
compute w2=1/p2.

*Berechnung des Hochrechnungsgewichts.
*Hochrechnungsgewicht sollte in CSSAMPLE verwendet werden.
compute wtot=w1*w2.
```

```
weight by wtot.

fre var=schicht.
des var=testscore/stat = mean stddev semean.
compute ww=wtot * (30/3600).

weight by ww.

fre var=schicht.
des var=testscore/stat=mean stddev semean.
```

	N	Mittelwert		Standardabweichung
		Statistik	Standardfehler	
ungewichtet	30	497,71	53,175	291,252
Testscore (Hochrechnung auf GG)	3600	462,71	2,471	148,280
gewichtet auf Stichprobe	30	462,71	27,531	150,794

```
CSPLAN ANALYSIS
 /PLAN FILE='D:\texte\ihs\strat.csaplan'
 /PLANVARS ANALYSISWEIGHT=wtot
 /PRINT PLAN
 /DESIGN  STRATA= schicht  CLUSTER= schule
   /ESTIMATOR TYPE=EQUAL_WOR
   /INCLPROB VARIABLE= p1
 /DESIGN
   /ESTIMATOR TYPE=EQUAL_WOR
   /INCLPROB VARIABLE= p2.

CSDESCRIPTIVES
 /PLAN FILE = 'D:\texte\ihs\strat.csaplan'
 /SUMMARY VARIABLES =testscore
 /MEAN
 /STATISTICS SE DEFF DEFFSQRT CIN (95)
 /MISSING SCOPE = ANALYSIS CLASSMISSING = EXCLUDE.
```

		Schätzung	Standard-fehler	95%-Konfidenzintervall		Effekt des Stichproben plans	Wurzel aus dem Effekt
				Untere Grenze	Obere Grenze		
Mittel wert	Testscore	462,71	46,656	314,23	611,19	2,896	1,702

gewichtet auf Stichprobe	30	462,71	27,531	150,794

Wurzel aus Effekt = 46,656 / 27,531 = 1,6946

	Schätzung	Standard-fehler	95%-Konfidenzintervall		Effekt des Stichproben plans	Wurzel aus dem Effekt
			Untere Grenze	Obere Grenze		
zweistufig mit Schichtung	462,71	46,656	314,23	611,19	2,896	1,702
zweistufig ohne Schichtung	462,71	56,059	318,61	606,82	4,181	2,045
einfache Zufallsauswahl mit Gewichtung	462,71	23,560	414,53	510,90	,738	,859

6. Weiteres Beispiel

Abbildung 5: Ergebnisse aus CSDESCRIPTIVES

	Schätzung	Standard-fehler	95%-Konfidenzintervall		Effekt des Stichprobenplans	Wurzel aus dem Effekt
			Untere Grenze	Obere Grenze		
lehre	-2,9274	,03436	-2,9973	-2,8575	2,420	1,556
rela	-1,7419	,01803	-1,7786	-1,7053	1,590	1,261
anfor	-2,2527	,03442	-2,3228	-2,1827	2,230	1,493
alter	-2,4316	,03626	-2,5053	-2,3578	2,058	1,434
deutsch	2,0655	,04010	1,9839	2,1471	3,008	1,734
mathe	2,0098	,03600	1,9366	2,0831	2,261	1,504
….						
ahs	,3123	,02206	,2674	,3572	3,520	1,876

Bei multivariaten Verfahren wird der Designeffekt i.d.R. geringer!

Abbildung 6: 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

BUB = Geschlecht des Kindes, WALLEIN = weiblicher Alleinerzieherhaushalt, VERANT = väterliche (Mit-)Verantwortung, SCHICHT = soziale Schicht der Eltern, AHS_Nähe = AHS in Wohnortnähe, MATURA = Bildungsaspiration der Eltern (1=Matura oder höher, 0=sonst), SCHLEIST = schulischen Leistungen in der 4. Klasse VS,

7. Verbleibende große Fallzahlen

komplexe Stichprobe → effektive Stichprobengröße = kleiner als die gezogene Stichprobe → günstig für Signifikanztestung (nicht alle Zusammenhänge sind signifikant) → dennoch häufig großes „n“, so dass alle Zusammenhänge signifikant sind

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)
- Verzicht auf NHSST-Paradigma (NHSST = null hypothesis statistical significance tests) und Verwendung von Effektstärken

Effektstärken

→ Psychologie (z.B. Thompson 2007), Medizin (Singh 2006) und andere Disziplinen üblich

Gründe

- Generalisierung nicht angestrebt (z.B. Evaluierung)
- kleine Fallzahl, bedeutsame Effekte, aber nicht signifikant
- bessere Interpretierbarkeit

sinnvolle Ergänzung zum NHSST-Paradigma (NHSST = null hypothesis statistical significance tests), aber kein Ersatz

Typen von Effektstärken (z.B. Thompson 2007)

- d-Maße = Differenzenmaße (Cohens d ..)
- r-Maße = Zusammenhangsmaße (Korrelation r usw.)

Abbildung 7: Vergleich Cohens d und t-Test

Cohens d:	t-Test Mittelwertdifferenz
$d = \dfrac{\bar{x}_A - \bar{x}_B}{s_{X,pooled}}$ $s_{X,pooled} = \sqrt{\dfrac{(n_A - 1)\cdot s_{XA}^2 + (n_B - 1)\cdot s_{XB}^2}{(n_A + n_B - 2)}}$	$t = \dfrac{\bar{x}_A - \bar{x}_B}{s_{(\bar{x}_A - \bar{x}_B)}}$ $s_{(\bar{x}_A - \bar{x}_B)} = \sqrt{\dfrac{(n_A - 1)\cdot s_A^2 + (n_B - 1)\cdot s_B^2}{(n_A + n_B - 2)}} \times \sqrt{\dfrac{1}{n_A} - \dfrac{1}{n_B}}$

Abbildung 8: Beispiel

	MW	Standardabweichung	n	2*n	n/2	n-200	n+500
Mädchen	1,96	0,83	721	1442	361	521	1221
Buben	2,11	0,85	756	1512	378	556	1256
		MW-Differenz	0,15	0,15	0,15	0,15	0,15
		S(pool)	0,84	0,84	0,84	0,84	0,84
		Wurzel(1/n1+1/n2)	0,052	0,037	0,074	0,061	0,04
		Effektstärke d	0,179	0,179	0,179	0,178	0,179
		t-test	3,429	4,85	2,426	2,927	4,442
		U2	0,536	0,536	0,536	0,536	0,536
		U1	0,133	0,133	0,133	0,133	0,133

8. Schwellenwerte für Effekt- und Zusammenhangsmaße

Trotz der expliziten Warnung von Cohen (1977: 12-13) werden vielfach seine Schwellenwerte zitiert.

Abbildung 9: Schwellenwerte von Cohen

	Cohens d	Pearson r ($r = \frac{d}{\sqrt{d^2 + (1/pq)}}$)
small:	d = 0,20	r = 0,10 (=$\frac{0,1}{\sqrt{0,1^2 + (1/0,25)}}$)
medium:	d = 0,50	r = 0,30
large:	d = 0,80	r = 0,50

(a) Cohen (1977: 40), (b) Cohen (1977: 79-81)

Schwellenwerte sind für soziologische Analysen nicht geeignet, insbesondere für die soziale Ungleichheitsforschung, wenn der Zusammenhang sozialstruktureller Variablen mit bestimmten Outputvariablen untersucht wird.

Brauchbare Schwellenwerte:

- Empirische Ableitung (Vergleich mit anderen Studien)
- Theoretische Ableitung aus Variablenmodell → Korrelation von 0,10 sind bedeutsam → Anhang A6

9. Fazit

- komplexe Stichproben → i.i.d.-Annahme verletzt
- mehrstufige Auswahlverfahren → Genauigkeitsverlust → Standardstatistikmodule führen zu Fehlschlüssen → **Spezialsoftware oder Zusatzmodule verwenden!!!**
- **Genauigkeitsverlust bei komplexem Design sollte bei Planung berücksichtigt werden!!!**
- **Effektstärken sinnvolle ergänzende Informationen!!!**

- Schwellenwerte von Cohen für Sozialstrukturanalyse nicht brauchbar, Sozialstrukturanalyse **geringere Schwellenwerte** (→ einheitliche Standards bei Publikationen, Vernetzung und Austausch, mehr Selbstbewusstsein!)

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.

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 Peter, Lachmayr Norbert (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.

Anhang A1: Kriterien von Wolter

Kriterien nach Wolter	einfach	komplex
degree of complexity of sample design	telefonische Befragung von n=500 Jugendlichen auf der Basis einer Zufalls-stichprobe	Befragung und Testung von Schüler/innen mit Hilfe mehrstufigen Verfahrens (z.B. PISA)
degree of complexity of sample estimator	Anteil der Jugendlichen mit starkem politischen Interesse	Verhältnis SpitzenschülerInnen zu RisikoschülerInnen
multiple characteristics of variables of interest	nur wenige Variable, z.B. polit. Interesse, Parteienpräferenz	Kompetenzen plus indiv. und schulische Kontextmerkmale
descriptive and analytical uses of the survey data	nur deskriptive Aussagen erwünscht	Einfluss der sozio-ökonomischen Merkmale auf Kompetenzen
the scale or size of survey	geringer Umfang von n=500	insgesamt über 250.000 Schüler/innen in ca. 40 Ländern

Anhang A2: Desingeffekt DEFFSQRT für unterschiedliche Konstellationen von nb, ρ und ρ_s

n_b	25	50	100	200	400	800
n	1000	1000	1000	1000	1000	1000
n_w	40	20	10	5	2,5	1,25
	maximaler Stichprobenfehler bei einfacher Zufallsauswahl aus großer Grundgesamtheit					
$\sigma(p)_{einfach}$	±1,6%	±1,6%	±1,6%	±1,6%	±1,6%	±1,6%
	Effekte der Intraklassenkorrelation ρ und unterschiedlicher Clustergröße n_w, keine Schichtung der Primäreinheiten $\rho_s = 0$					
ρ = 0,0	1,00	1,00	1,00	1,00	1,00	1,00
0,2	2,97	2,19	1,67	1,34	1,14	1,02
0,5	4,53	3,24	2,35	1,73	1,32	1,06
0,8	5,67	4,02	2,86	2,05	1,48	1,10
1,0	6,32	4,47	3,16	2,24	1,58	1,12
	Effekte bei Schichtung für unterschiedliche Werte von ρ_s und unterschiedliche Clustergrößen n_w, feste Intraklassenkorrelation von $\rho = 0,20$					
ρ_s = 0,00	2,97	2,19	1,67	1,34	1,14	1,02
0,05	2,61	1,95	1,52	1,24	1,08	0,99
0,10	2,19	1,67	1,34	1,14	1,02	0,96
0,15	1,67	1,34	1,14	1,02	0,96	0,93
0,20	0,89	0,89	0,89	0,89	0,89	0,89

Anhang A3: 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)

Anhang A4: Jackknife-Methode

Table 3.11 ■ The Jackknife replicates for stratified two-stage sample designs

Pseudo-stratum	School	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	1	2	1	1	1	1	1	1	1	1	1
1	2	0	1	1	1	1	1	1	1	1	1
2	3	1	0	1	1	1	1	1	1	1	1
2	4	1	2	1	1	1	1	1	1	1	1
3	5	1	1	2	1	1	1	1	1	1	1
3	6	1	1	0	1	1	1	1	1	1	1
4	7	1	1	1	0	1	1	1	1	1	1
4	8	1	1	1	2	1	1	1	1	1	1
5	9	1	1	1	1	2	1	1	1	1	1

$$\sigma^2_{(\hat{\theta})} = \sum_{i=1}^{G} (\hat{\theta}_{(i)} - \hat{\theta})^2$$

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

Anhang A5: Vergleich von COMPLEX SAMPLE in SPSS und SURVEY METHODS in STATA

	STATA9	SPSS14
Stichprobenziehung:	nein	ja
Varianzschätzmethoden:		
• Taylor-Linearisierung	ja	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.

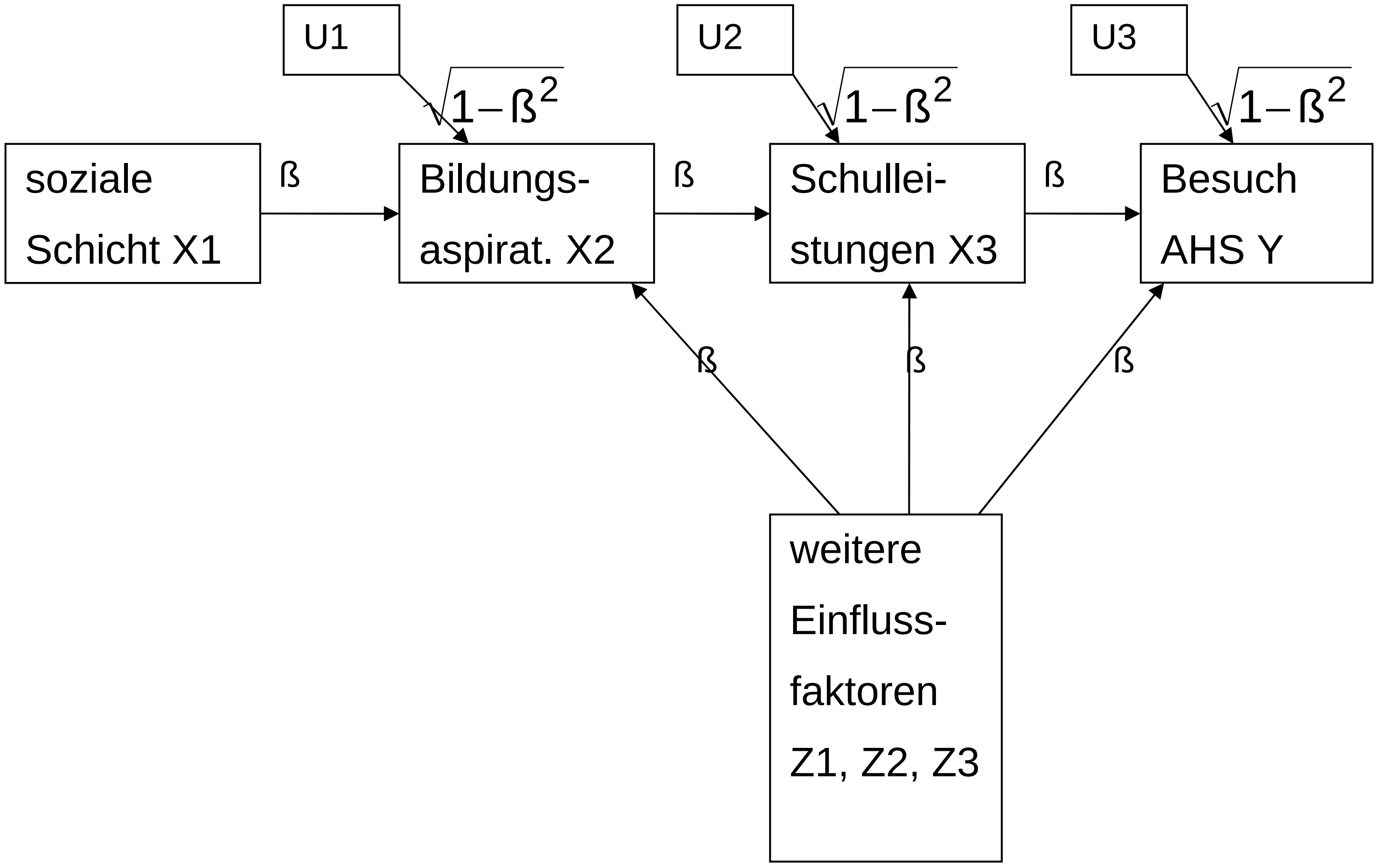
(c) kann einfach geprüft werden, indem eine Variabel $d = y_1 - y_2$ mit y_1 = erste Messung und y_2 = zweite Messung gebildet und getestet, ob diese von Null verschieden ist.

(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)

Anhang A6: Ableitung von theoretischen Korrelationen



$$V(X1) = V(U1) = V(U2) = V(U3) = V(Z1) = V(Z2) = V(Z3) = 1$$

$$V(X2) = ß^2 V(X1) + (\sqrt{1-ß^2})^2 V(U1) + ß^2 V(Z1) + ß^2 V(Z2) + ß^2 V(Z3) = 1 + 3 \cdot ß^2$$

$$V(X3) = ß^2 V(X2) + (\sqrt{1-ß^2})^2 V(U2) + ß^2 V(Z1) + ß^2 V(Z2) + ß^2 V(Z3) =$$
$$ß^2(1+3\cdot ß^2) + (1-ß^2) + 3\cdot ß^2 = 1 + 3\cdot(ß^2 + ß^4)$$

$$V(Y) = ß^2 V(X3) + (\sqrt{1-ß^2})^2 V(U3) + ß^2 V(Z1) + ß^2 V(Z2) + ß^2 V(Z3) =$$
$$ß^2(1+3\cdot ß^2 + 3\cdot ß^4) + (1-ß^2) + 3\cdot ß^2 = 1 + 3\cdot(ß^2 + ß^4 + ß^6)$$

$$C(X1, Y) = ß \cdot ß \cdot ß = ß^3$$

$$Cor(X1,Y) = \frac{C(X1,Y)}{\sqrt{V(X1)\cdot V(Y)}} = \frac{ß^3}{\sqrt{1\cdot(1+3\cdot(ß^2+ß^4+ß^6))}} = \frac{0{,}5^3}{\sqrt{1\cdot(1+3\cdot(0{,}5^2+0{,}5^4+0{,}5^6))}} =$$
$$\frac{0{,}125}{\sqrt{1\cdot(1+3\cdot(0{,}328))}} = \frac{0{,}125}{\sqrt{1{,}984375}} = 0{,}089$$

allgemeine Formel: $Cor(X1, Y) = \dfrac{ß^{p+1}}{\sqrt{1 + q \cdot \sum_{i=1}^{p+1} ß^{2i}}}$

p = Zahl der intervenierenden Variablen

q = Zahl der zusätzlichen Einflussfaktoren

ß = direkter Einfluss zwischen den Variablen

Annahme: X1, Zi und Ui voneinander unabhängig