Statistik und Computergestützte Datenanalyse II: Gliederung

- 1 Sortieren von Fällen
- 2 Aggregieren von Daten
- 3 Zusammenfügen von SPSS-Datendateien
 - 3.1 Zusammenfügen von Fällen (vertikal)
 - 3.2 Zusammenfügen von Variablen (horizontal)

- Die Reihenfolge der Anordnung der Fälle im Daten-Editor von SPSS ist im Allgemeinen ohne Bedeutung. Bestimmte Prozeduren (z.B. das Zusammenfügen von Datendateien) erfordern aber eine bestimmte Reihenfolge (Sortierung) der Fälle. Außerdem ermöglicht eine Sortierung manchmal auch eine leichtere visuelle Inspektion der Daten.
- SPSS kann Daten nach den Werten in einer Variablen sortieren, die z.B. vom Typ String oder numerisch sein kann. Die Daten können aufsteigend oder absteigend nach den Werten der anzugebenden Sortiervariablen sortiert werden. Aufsteigend heißt bei numerischen Variablen entsprechend der Größe der Werte (missing = kleinster Wert) und bei String-Variablen alphabetisch (ä wird "nahezu" wie a behandelt, etc.; Groß=Kleinschreibung).
- Weisen mehrere Fällen den gleichen Wert in der Sortiervariablen auf (Tieblock), so ist deren Reihenfolge in diesem Block "zufällig". Durch die Angabe einer (oder mehrerer weiterer) Sortiervariablen kann man auch die Sortierreihenfolge innerhalb der Tieblocks bestimmen.

	vpnr	name	alter			vpnr	name	alter
1	1	Muster	44		1	2	Maier	<mark>6</mark> 5
2	2	Maier	65		2	3	Mühe	19
3	3	Mühe	19	Aufstaizand contiant	3	4	Muhe	<mark>6</mark> 5
4	4	Muhe	65	Auisteigena sortiert	4	1	Muster	44
5	5	Muster	32		5	5	Muster	32
C				IIallie	C			

Unter Daten/Fälle_sortieren... kann die Datendatei nach einer oder mehreren Stringoder/und numerischen Variablen sortiert werden.



ta *uebung0910.sav [DataSet1] - IBM SPSS Statistics Dateneditor										
Datei B <u>e</u> arbei	t Ansich <u>D</u> a	ter <u>T</u> ra	ansform	it <u>A</u> nalysier	Direkt <u>m</u> a	arke <u>G</u> rafił E	xtras Fenste	<u>H</u> ilfe		
🛅 🗟 🖨	- 🕠 🗠	-24		L 🗐 🔢	M	🐮 🔜 🗄	44 🎛	1∯	0	
1 : gebjahr		1930			:	Sichtbar: 36	von 36 Varia	blen		
	phantas	6	sex	gebtag	gebmon	gebjahr	gewicht			
1	Loreley		2	17	11	1930	48,00	*		
2	Dornroesch	nen	2	2	1	1954	70,00			
3	lockes		2	9	7	1955 '	63,00			
4	Waltraud		2	8	5	i 1955	56,00			
5	katharina		2	28	8	1956	65,00			
6	Shaik		2	14	10	1958	55,00			
7	Kama		2	9	5	1959	69,00			
8	Schleumel		2	26	6	1959	65,00			
9	Mandela		1	15	11	1961	100,00			
10	bachus		2	29	11	1962	-			
11	Karlchen		2	5	6	1963	95,00		1	
12	ginger		2	28	11	1963	58,00			
13	sevilla		2	23	5	1964	75,00			
14	Pinocina		2	20	10	1966 1966	68,00		-	
15	cinderel		2	16	11	1966	58,00			
16	palomino		2	29	11	1966	89,00			
17	orange		2	12	8	1966	58,00			
18	uhrh		2	17	1	1967	56,00			
19	Sonnenblu	me	2	20	5	1967	66,00			
20	tiddel		2	12	12	1967	78,00	-		
	1									
Datenansich	t Variable	nansi	cht							
	IBM SF	PSS S	tatistics	-Prozessor	ist bereil	Unicod	le:ON			

Hier sind die Personen (=Fälle) jetzt aufsteigend nach dem Geburtsjahr (Variable gebjahr) sortiert.

Personen mit gleichem Geburtsjahr werden in ihrer Reihenfolge "zufällig" angeordnet.

Um innerhalb der Tieblöcke zu sortieren, kann man auch mehrere Sortiervariablen angeben, im Beispiel gebjahr, gebmon und gebtag.

16								_
		phantas	sex	gebtag	gebmon	gebjahr	gewicht	
	1	Loreley	2	17	11	1930	48,00	
	2	Dornroeschen	2	2	1	1954	70,00	
	3	Waltraud	2	8	5	1955	56,00	
	4	lockes	2	9	7	1955	63,00	
	5	katharina	2	28	8	1956	65,00	
	6	Shaik	2	14	10	1958	55,00	
	7	Kama	2	9	5	1959	69,00	
	8	Schleumel	2	26	6	1959	65,00	
	9	Mandela	1	15	11	1961	100,00	
	10	bachus	2	29	11	1962		
	11	Karlchen	2	5	6	1963	95,00	
	12	ginger	2	28	11	1963	58,00	
	13	sevilla	2	23	5	1964	75,00	
	14	orange	2	12	8	1966	58,00	
	15	Pinocina	2	20	10	1966	68,00	
	16	cinderel	2	16	11	1966	58,00	
	17	palomino	2	29	11	1966	89,00	
	18	uhrh	2	17	1	1967	56,00	
	19	Sonnenblume	2	20	5	1967	66,00	
	20	tiddel	2	12	12	1967	78,00	-
		4)	•
	Datenansic	t Variablenans	icht					
		IBM SPSS S	Statistics	-Prozessor	ist bereil	Unicod	e:ON	



In diesem Fall wird innerhalb der Fälle, die in der ersten Sortiervariablen (gebjahr) den gleichen Wert haben aufsteigend nach der zweiten Variablen (gebmon) sortiert und dann in den Fällen, die in beiden ersten Sortiervariablen den gleichen Wert haben hinsichtlich der dritten Variablen (gebtag) sortiert. Die Reihenfolge der Sortiervariablen ist also bedeutsam!

Statistik und Computergestützte Datenanalyse II: Gliederung

- 1 Sortieren von Fällen
- 2 Aggregieren von Daten
- 3 Zusammenfügen von SPSS-Datendateien
 - 3.1 Zusammenfügen von Fällen (vertikal)
 - 3.2 Zusammenfügen von Variablen (horizontal)

- Manchmal ist man daran interessiert, statistische Analysen nicht auf der Ebene der Personen (Fälle) sondern auf der Ebene von Gruppen von Personen, die in einer oder mehreren Variablen übereinstimmen, zu berechnen. Dazu kann man eine neue Datendatei erzeugen (aggregieren), die als Fälle nicht mehr die Personen, sondern die Mittelwerte einer Gruppe von Personen (oder andere Statistiken) enthält. Mit dieser neuen Datendatei kann man dann wieder wie üblich alle möglichen Analysen durchführen.
- Beispiel: Im Rahmen der Lehrvaluation beurteilen Studierende die Qualität von Seminaren. Geben wir diese Daten in SPSS ein, so ist jeder Fragebogen (Person) eine Zeile und durch eine Variable (hier vcode) wird eindeutig die beurteilte Veranstaltung identifiziert.



vcode	pd1	us2	ir3	n_k
102003	3	4	4	
102003	4	,	3	
102003	4	3	4	
102014	4	3	2	
				\rangle 26
102014	2	3	3	
102019	4	4	4	-

- ➤ Ingesamt enthält unsere Datendatei die Urteile von n = 4406 Studierenden, die sich auf k = 240 Veranstaltungen beziehen. Die Zahl der Studierenden pro Seminar schwankt zwischen n₁₀₂ = 3 und n₂₂ = 84 (durchschnittlich \overline{n}_{\bullet} = 18.4).
- Wollten wir nun z.B. wissen, wie die mittlere Bewertung der Frage (des Likert-Items) pd1= "Das Seminar verläuft nach eine klaren Gliederung", gemittelt über die Veranstaltungen ausfällt, so müssten wir den Mittelwert über die mittleren Bewertungen der Veranstaltung bestimmen.
- Frage: Resultiert nicht der selbe Mittelwert, wenn wir gleich über alle Antworten mitteln? Formal: Wenn x_{ij} die Antwort der Person i in der Veranstaltung j auf die obige Frage ist, gilt dann

$$\frac{\sum_{i=1}^{n_j} \sum_{j=1}^k x_{ij}}{n} \stackrel{?}{=} \frac{\sum_{j=1}^k \overline{x}_j}{k} \quad \text{wenn } \overline{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

der Mittelwert der Veranstaltung j(mit n_j Teilnehmern) ist?

Die Antwort lautet im Allgemeinen: Nein.



- > Der Mittelwert der Mittelwerte \overline{x}_j entspricht nur dann dem Mittelwert über alle Daten x_{ij} , wenn alle Gruppen j die gleiche Größe aufweisen und keine missing data existieren, also gilt: $n_1 = n_2 = ... = n_j = ... = n_k$ (die Zahl aller gültigen Werte in allen Gruppen gleich sind).
- Allgemein gilt: Statistiken auf der Ebene der Rohdaten (Personen) entsprechen nicht (notwendig) den Statistiken auf einer höheren Ebene (hier: Veranstaltungen). Dies gilt auch, z.B. wenn man die Varianzen der Rohwerte und der Varianzen der gemittelten Werte berechnet oder die Korrelationen zwischen zwei Variablen auf der Ebene der Personen und der Veranstaltungen. (Die Varianzen und Korrelationen sind noch nicht einmal gleich, wenn die Stichprobengrößen gleich sind.)
- Man muss also immer überlegen, auf welcher Ebene man die Analysen durchführen will und dies dann auch berichten.
- Im Beispiel würde man, wenn man den Gesamtmittelwert über die Mittelwerte der Veranstaltungen statt über alle Personen berechnen würde, alle Veranstaltungen gleich gewichten. Ein Seminar, das von 3 Studierenden bewertet wurde, würde den Gesamtwittelwert also genauso stark beeinflussen, wie eines, das von 60 Studierenden bewertet wurde. Beim Mittelwert über alle Personen würde das letztere Seminar andernfalls 30 mal stärker eingehen.

Um Gruppen von Fällen (Personen) zu einem Fall zusammenfassen und die in den Gruppen aggregierten Werte (z.B. Mittelwerte) in eine neue SPSS-Datendatei zu schreiben, muss man die Gruppen durch die Werte in einer (oder mehreren) kategorialen Break-Variablen definieren.

vcode	pd1	us2	ir3							
102003	3	4	4)		vcode	pd1mean	us2mean	•••	nbreak
102003	4	,	3			102003	3.38	3.00		8
					↗	102014	3.72	3.46		26
102003	4	3	4		1	102019	3.78	3.50		10
102014	4	3	2							
							1 Fall = 1 Ve	eranstaltung		
102014	2	3	3		/ 		1			
102019	4	4	4		break-v	variable:				
					VC	ode				
16	1 Fall = 1 Person									



Hier werden eine oder mehrere kategoriale Break-Variablen angegeben. Jede Kombination von Break-Variablenwerten definiert eine Gruppe.

Hier werden die Variablen angegeben, über die aggregiert werden soll sowie deren Aggregierungsfunktion (Standard (Funktion): mean). Nur diese Variablen werden in die neue Datei geschrieben.

Auswertungsstatis	tik Bestimmte Werte	Anzahl der Fälle
Mittelwert	O Erste	Gewichtet
© Media <u>n</u>	© <u>L</u> etzte	O Gewichtet fehlend
© <u>S</u> umme	⊘ Minim <u>u</u> m	O Ungewichtet
© Standa <u>r</u> dabwe	eichung 🔘 Ma <u>x</u> imum	O Ungewic <u>h</u> tet fehlend
© <u>P</u> rozentwerte	© An <u>t</u> eile	gkeiten
	Wert:	
Ounter		



Hier kann als zusätzliche Variable die Zahl der aggregierten Fälle pro Gruppe erzeugt werden (sinnvoll!). Der Name kann abweichend vom Weglasswert N_BREAK spezifiziert werden.

Hier wird unter (Datei ...) der Name der . sav-Datei angegeben werden, in der die aggregierten Daten gespeichert werden sollen. Diese Datei wird dann durch die Prozedur erstmal nur gespeichert und nicht gleich geöffnet.

(Möglich ist z.B. alternativ auch das Hinzufügen der Daten in die aktuelle Datei oder ein neues Datenblatt.)

SORT CASES BY voode						
AGGREGATE		Nebenstehend die Syntax des				
/OUTFILE='d:\stat2spss\fesem	_aggr.sav'	eingefügten Aggregierungsbefehls.				
/ PRESORTED						
/BREAK=vcode						
/pd1_mean=MEAN(pd1)						
/us2_mean=MEAN(us2)	SORT CASES	BY vcode.				
/ir3_mean=MEAN(ir3)	AGGREGATE /OUTFILE='d:\stat2spss\fesem_aggr.sav' /PRESORTED /BREAK=vcode					
/us4_mean=MEAN(us4)						
/int_mean=MEAN(int)						
/N_BREAK=N.	/pd1=MEAN(pd1)					
	/us2=MEAN	(us2)				
Sinnvoll ist meistens (wie rechts) vor	/ir3=MEAN	(ir3)				
der Ausführung des Befehls in der	/us4=MEAN	(us4)				
Syntax die Namen für die Variablen						
in der neuen aggregierten Datei von	/int=MEAN(int)					
dem Zusatz _mean zu betreien.	/N_BREAK=N.					

fesem.sav	[DataSet2] - IB	SM SPSS St	tatistics Dat	teneditor	-	-	(manager	-		×						1	
Datei B <u>e</u> arl	beiten Ansicht	<u>D</u> aten 1	<u>T</u> ransformi	erei <u>A</u> nalysi	ieren Dire	ekt <u>m</u> arket	tin <u>i G</u> rafik	Extra	is Fenster <u>I</u>	Hilfe	R	oh	datendat	ei fese	m.sav		
) 🖪 🖛	1	S 🛃 🗐	P H		2 ==	4 🗄	14	🗞 🍋 👎	b						J	
							Sichtb	ar: 4	1 von 41 Varia	ablen							
	vcode	pd1	us2	ir3	us4	ir5	pd6	us7	/ pd8								
1																	
2	102003	4		3	3	4	3	t	fesem_ago	gr.sav [DataSe	t3] - IBM SPS	S Stat	tistics Dateneditor				x
3	102003	3	3	3	4	3	4		Datei B <u>e</u> ar	beiten Ansi	cht <u>D</u> aten	Tran	sformieren <u>A</u> nalys	sieren Direkt <u>m</u> ar	keting <u>G</u> rafik Extr	as Fenster <u>H</u> i	ilfe
4	102003	3	3	4	3	4	4				• • • • • • •	+		🐮 🖂 📰 🖚	📰 🎝 🚱 🌑		
5	102003	4	4	4	4	3	3			- Annual -					Sichtha	r: 42 yon 42 Varial	blen
6	102003	3	3	3	2	3	2							÷0		:	
- /	102003	3	1	3	3	ქ ე	2		1	VCODE 102002	pd1_mea	an 2.20	usz_mean	Ir3_mean	us4_mean	rs_mean	
0	102003	4	3 2	4	2	3	4			102003		2,30	3,00	3,50	3,25	3,13	
10	102014	4	3	2	3	4	4		2	102014		3,1Z	3,40	3,22	3,11	3,29	- 1
11	102014	4	3	4	4	3	4		3	102019		3,10	3,50	3,50	4,00	3,11	- 1
12	102014	3	2	3	4	3	3		4	103001		3,19	3,13	2,94	3,00	3,13	- 1
13	102014	4	3	5	4	3	3		5	103004		3,11	3,03	2,04	3,73	2,71	- 1
14	102014	4	4	4	4	3	4		7	103005		3,96	3,38	3,30	3,40	3,06	- 1
15	102014	4	4	4	4	4	3		1	103000		4,00	3,75	3,50	3,50	2,91	- 0
16	102014	1	1	3	Л	3	3		0	103007		4,00	4,00	3,29	4,00	2,00	
	1					_		╡╠	9	102010		4,00	3,01	3,04	3,90	2,00	-
Datenansio	ht Variablen	ansicht							10	103010		3,90	3,37	3,75	3,95	2,62	-
			IBM SPS	SS Statistics	s -Prozess	sor ist be	reit U	Ini	10	102011		3,91	3,30	3,18	3,04	1,50	
			1						12	102012		3,90	3,90	3,70	4,00	3,13	
								F	14	102015		3,02	3,52	3,52	3,90	3,17	
									14	102015		3,10	3,00	3,70	3,70	2,07	
		-						X	15	102010		3,41	3,24	3,06	3,82	2,00	
Die ar	nschlie	ßend	d zu ö	öffnei	nde	aggi	re- 🖊		10	4		1/1	5.09	1.64	4 00	100	
gierte	Datei	fes	sem_	_agg	r.s	av	ſ		Datenansic	<mark>ht</mark> Variable	nansicht						
	IBM SPSS Statistics -Prozessor ist bereit Unicode:ON																

Mittelwerte etc. und Interkorrelationen der drei Variablen pd1, us2 und ir3.

Rohdatendatei fesem.sav

Deskriptive Statistik

	Ν	Minimum	Maximum	Mittelwert	Standardab weichung
pd1	4331	1	4	3,66	,615
us2	4303	1	4	3,46	,750
ir3	4276	1	4	3,26	,773
Gültige Werte (Listenweise)	4128				

Korrelationen

		pd1	us2	ir3
pd1	Korrelation nach Pearson	1	,152**	,258**
	Signifikanz (2-seitig)		,000	,000
	Ν	4331	4236	4212
us2	Korrelation nach Pearson	,152**	1	,386**
	Signifikanz (2-seitig)	,000		,000
	Ν	4236	4303	4189
ir3	Korrelation nach Pearson	,258**	,386**	1
	Signifikanz (2-seitig)	,000	,000	
	Ν	4212	4189	4276

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Aggregierte Datei fesem_aggr.sav

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardab weichung
pd1_mean	240	1,88	4,00	3,6651	,30727
us2_mean	240	2,47	4,00	3,5322	,33149
ir3_mean	240	2,08	4,00	3,3476	,37150
Gültige Werte (Listenweise)	240				

Korrelationen

		pd1_mean	us2_mean	ir3_mean
pd1_mean	Korrelation nach Pearson	1	,192**	,233**
	Signifikanz (2-seitig)		,003	,000
	Ν	240	240	240
us2_mean	Korrelation nach Pearson	,192**	1	,603**
	Signifikanz (2-seitig)	,003		,000
	Ν	240	240	240
ir3_mean	Korrelation nach Pearson	,233**	,603**	1
	Signifikanz (2-seitig)	,000	,000	
	Ν	240	240	240

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Statistik und Computergestützte Datenanalyse II: Gliederung

- 1 Sortieren von Fällen
- 2 Aggregieren von Daten
- 3 Zusammenfügen von SPSS-Datendateien
 - 3.1 Zusammenfügen von Fällen (vertikal)
 - 3.2 Zusammenfügen von Variablen (horizontal)

Man kann in SPSS zwei Datendateien zusammenfügen. Zu unterscheiden ist der Fall, dass Fälle hinzugefügt werden sollen und der Fall, dass Variablen hinzugefügt werden sollen.



Man kann in SPSS zwei Datendateien zusammenfügen. Zu unterscheiden ist der Fall, dass Fälle hinzugefügt werden sollen oder Variablen hinzugefügt werden sollen.



Fälle werden hinzugefügt unter Daten/Datei en_zusammenfügen/Fälle_hinzufügen...

Von beiden zusammenzufügenden Dateien wird eine (hier: **d1.sav**) in SPSS geöffnet. Die Datei, die hinzugefügt werden soll, kann entweder gleichzeitig ebenfalls in SPSS geöffnet sein und muss dann durch Anklicken nach obigem Aufruf ausgewählt werden ...

Fälle hinzufügen zu d1.sav[DataSet1]	×
Wählen Sie aus der Liste der geöffneten Datasets oder aus einer Datei das Dataset aus, das mit dem aktiven Dataset zusammengefügt werden soll.	
Ein offenes Dataset	
d2.sav[DataSet2]	
C Eine externe SPSS Statistics-Datendatei	
	Durchsuchen
Datendateien, die nicht im SPSS Statistics-Format vorliegen, müssen erst in SPSS Statistics geöffnet werden, bevor sie beim Zusammenfügen verwe	ndet werden können.
Weiter Abbrechen Hilfe	

... oder sie liegt extern (in SPSS nicht geöffnet) vor und muss unter dem unteren Menüpunkt angewählt werden.

Fälle hinzufügen zu d1.sav[DataSet1]		×
Wählen Sie aus der Liste der geöffneten Datasets ode © Ein offenes Dataset d2.sav[DataSet2] © Eine externe SPSS Statistics-Datendatei Datendateien, die nicht im SPSS Statistics-Format von	er aus einer Datei das Dataset aus, das mit dem aktiven Dataset zu Fälle hinzufügen: Datei lesen Suchen in: stat2spss Image: Stat2spss Image: Stat2spss Image: Stat2spss	Esammengefügt werden soll.
	Dateiname: d2.sav	Öffnen Abbrechen
	Datei aus <u>R</u> epository abrufen	Hilfe



Alle Variablen, die nur in der ersten Datei (d1) existieren, werden mit (*) markiert, alle die nur in der zweiten Datei existieren (d2) mit (+). Sollen diese Variablen in die neue Datei aufgenommen werden, so müssen sie mittels vusätzlich nach rechts kopiert werden. Aktiviert man diese Option, so wird eine neue Variable mit dem (default) Namen source01 erzeugt, die die Information enthält, aus welcher Datei ein Fall stammt (sinnvoll!).

ta *d1.sav [l Datei Bea

🖹 🗟 i

8

9

111

Datenansicht Variablenansicht

Fälle hinzufüger	n aus C:\Users\tstaufe	n\Documents\st	at2spss\d2.sav	×
Nicht paarige Var	iablen:	Variablen i vpnr i1 i2 sex(*) i3(+)	in neuem aktive	em Dataset:
Umb <u>e</u> ne	nnen	source	01	
(*) = Aktives Data (+)=C:\Users\tsta	set ufen\Documents\stat <u>E</u> infügen <u>Z</u> ur	t2spss\d2.sav ücksetzen A	bbrechen	Hilfe
	vpnr	i1	i2	i3
	101	3	3	2
10	102	4	2	1
۵2	105	5	1	3
	111	2	4	4

		vp	nr	se	x	i	1	i	2		
			1		0		2		1		
			3		1		2		4		d
			4		0		3		2		
			5		1		,		4		
Ja	taSet1] - IBM SP	55 Statistics Da	steneditor								
rb	eiten Ansicht	<u>D</u> aten <u>T</u> ra	nsformieren	Analys	ieren (Direkt <u>m</u>	arketing	<u>G</u> rafik	Extr	as	Fenster
3		× 🖺 🛓		M 🖱			III 1		•	-	
									Sich	tbar:	6 von 6 \
	vpnr	sex	i1		i2		i3	SO	urce01	1	var
	1	0		2		1				0	
	3	1		2	4	4				0	
	4	0		3		2				0	
	5	0			4	4		-		0	
	101			3		3	2	2		1	
	102	-		4	:	2	1	1		1	
	105			5		1		3		1	

2

Δ

IBM SPSS Statistics -Prozessor ist bereit

Δ

Unicode:ON

Die zusammengefügte Datei behält den Namen der ersten Datei und sollte via Datei / Spei chern_unter... mit einem neuen Namen versehen werden (z.B. d1d2.sav) Hilfe

/ariablen

Statistik und Computergestützte Datenanalyse II: Gliederung

- 1 Sortieren von Fällen
- 2 Aggregieren von Daten
- 3 Zusammenfügen von SPSS-Datendateien
 - 3.1 Zusammenfügen von Fällen (vertikal)
 - 3.2 Zusammenfügen von Variablen (horizontal)

Der Dialog zum Hinzufügen von Variablen findet sich unter Daten/Datei en_zusammenfügen/ Vari abl en_hi nzufügen.... Von beiden Dateien wird eine vorab (hier: d1.sav) geöffnet und dann die zweite auf die Abfrage wie oben hinzugefügt (hier: d3.sav).





Alle Variablen, die in beiden Dateien existieren, werden hier aufgelistet. Die doppelten Variablen aus der zweiten Datei werden standardmäßig nicht mit aufgenommen. Unter Schlüsselvariablen sind die Variablen anzugeben, aufgrund derer die eindeutige Zuordnung der Fälle in beiden Dateien erfolgt.

🙀 Variablen hinzufügen aus DataSet3		×
Ausg <u>e</u> schlossene Variablen:		Neues aktives Dataset:
sex(+) vpnr(+)	•	vpnr(*) sex(*) i1(*) i2(*) f1(+) f2(+)
Umbenennen Fälle nach Schlüsselvariablen abgleichen Fälle werden in der Reihenfolge der Schlüsselvariablen in beiden Datasets sortierf Arteres Dataset ist Schlüsseltabelle Aktives Dataset ist Schlüsseltabelle Beide Dateien liefern Fälle	*	Schlüssel <u>v</u> ariablen:
Date <u>i</u> indikator als Variable: source01 (*) = Aktives Dataset (+)=DataSet3		
K Einfügen Zurücksetzen Abbrechen	Hilfe	

Bevor die Schlüsselvariable(n) angegeben werden können, müssen zunächst die Optionen "Fälle mittels Schlüsselvariablen verbinden" und "Fälle werden in der Reihenfolge …" aktiviert werden. (Die Voreinstellung "Beide Dateien liefern Fälle" behält man bei.)

Variablen hinzufügen aus DataSet3	x	
Ausg <u>e</u> schlossene Variablen: sex(+)	Neues aktives Dataset:	
Umbenennen Fälle nach Schlüsselvariablen abgleichen Fälle werden in der Reihenfolge der Schlüsselvariablen in beiden Datasets sortiert Anderes Dataset ist Schlüsseltabelle Aktives Dataset ist Schlüsseltabelle	f1(+) f2(+) Schlüsselvariablen:	Dann können eine oder, wenn nötig, mehrere Variablen aus den ausgeschlossenen Variabler in die Liste "Schlüsselvariablen" kopiert werden.
Beide Dateien liefern Falle Dateiindikator als Variable: source01 (*) = Aktives Dataset (+)=DataSet3 OK Einfügen Zurücksetzen Abbrechen Hilf	e	
IBM SPSS Statistics 22	×	Es erfolgt standardmäßig die
Warnung: Der Abgleich nach Schlüsselvariablen schlägt fehl, wincht in aufsteigender Reihenfolge nach den Schlüsselvariabler	enn die Daten n sortiert sind.	nebenstehende Warnung.

d1

vpnr	sex	i1	i2
1	0	2	1
3	1	2	4
4	0	3	2
5	0	,	4

d3

vpnr	sex	f1	f2
5	1	4	4
1	0	3	2
3	1	2	1

In diesem Fall ist die Datei d3 nicht bezüglich der Schlüsselvariablen vpnr aufsteigend sortiert und das Zusammenfügen schlägt fehl. Die Datei **d3** ist daher <u>vor-</u> <u>her</u> nach der Schlüsselvariablen vpnr zu sortieren:

vpnr	sex	f1	f2
1	0	3	2
3	1	2	1
5	1	4	4
L	/		1

🖹 🕂 👌		¥ 🔓 🚣 🗉		🕺 🔟 🗄	i 🕸 🎛 👔	4 💊 🍬 👎	5
7:					Si	chtbar: 6 von 6 \	/ariable
	vpnr	sex	i1	i2	f1	f2	٧
1	1	0	2	1	3	2	4
2	3	1	2	4	2	1	
3	4	0	3	2			
4	5	0		4	4	4	
5							
	4						

Das Ergebnis sieht dann wie folgt aus: