

- 1 Auswahl von Fällen
- 2 Schichtung von Analysen

- Häufiger will man bestimmte Analysen nur für Fälle (Personen) durchführen, die eine bestimmte **Bedingung** erfüllen (=Selektion),
  - z.B. will man nicht das mittlere Gewicht aller Personen, sondern nur für eine Teilmenge der Stichprobe, z.B. die Frauen, bestimmen.
  - Die Bedingung kann sich dabei auch auf die Verknüpfung mehrerer Variablen beziehen (z.B. "Geschlecht männlich" und "Alter größer als 18 Jahre").
  - Die Auswahl von Fällen geschieht bei SPSS unter Daten/Fälle auswählen...
- Häufiger will man Analysen auch für alle Ausprägungen einer Variable getrennt ausgeführt haben (=Schichtung),
  - z.B. für Männer und Frauen, alle Religionszugehörigkeiten getrennt usw.
  - Dies kann bei SPSS unter Daten/Aufgeteilt Datei ... angefordert werden.
  - Dabei wird dann keine Bedingung angegeben, sondern nur die Variable(n). Im Folgenden werden die dann angeforderten Analysen automatisch für jede (Kombination von) Ausprägung(en) dieser Variablen getrennt ausgeführt.
  - Der Befehl ersetzt also eine Aufeinanderfolge (einfacher) Selektionen.
  - Vorsicht bei Schichtungsvariablen mit vielen Werten und Kombinationen von Schichtungsvariablen: Diese können zu einer Flut von Analysen führen!

# SPSS: Fallauswahl

- Eine **<Bedingung>** ist ein logischer Ausdruck in dem Variablen und Konstanten (hier immer mit Dezimalpunkt schreiben!) mit **Operatoren** verknüpft werden

Vergleichsoperatoren	
=	gleich
>	größer als
<	kleiner als
>=	größer oder gleich
<=	kleiner oder gleich
<>	ungleich

Logische Operatoren	
NOT	nicht
AND	und
OR	oder (logisches)

Klammern ( )

Arithmetische Operatoren	
+	Addition
-	Subtraktion
*	Multiplikation
/	Division
**	Exponentiation: 2**3=8

Funktionen (nur eine sehr kleine Auswahl)	
ABS ( <var> )	der Betrag der Variable <var>
SQRT ( <var> )	die Wurzel aus der Variable <var>
RND ( <var> )	der ganzzahlig gerundete Wert von <var>
MEAN ( <var <sub>1</sub> > , ... , <var <sub>k</sub> > )	der Mittelwert aus den Variablen <var <sub>1</sub> > bis <var <sub>k</sub> >
MISSING ( <var> )	wahr, falls der Wert von <var> ein missing value ist

- Beispiele für **<Bedingung>**en (anhand von Variablen aus der Übungs-Datei):

<code>sex = 2</code>	alle Frauen
<code>sfach = "Psychologie"</code>	alle, die Psychologie studieren (bei String-Variablen müssen die Werte in Hochkommata eingeschlossen werden.)
<code>sex = 1 AND gewicht &gt; 60</code>	alle Männer, die mehr als 60 kg wiegen
<code>alter &gt;= 18 AND alter &lt;= 65</code>	alle, die zwischen 18 und 65 Jahre (inklusive) alt sind
<code>schulm &gt; schulv</code>	alle, bei denen die Schulbildung der Mutter höher ist, als die des Vaters
<code>MISSING(alter) OR MISSING(sex)</code>	alle Personen, die entweder im Alter oder im Geschlecht (oder beiden Variablen) einen fehlenden Wert aufweisen
<code>MEAN(pers1, pers2, pers3) &gt; 4.5</code>	alle, bei denen der Mittelwert aus den Ratings der drei ersten Persönlichkeitsfragen größer als 4.5 ist

# SPSS: Fallauswahl

- Bei den Operatoren muss es eine Priorität (d.h. eine Standardreihenfolge der Abarbeitung) geben damit z.B. klar ist,
  - ob  $2+3*4$  dem Wert  $(2+3)*4$  (also 20) oder  $2+(3*4)$  (also 14) entspricht
  - ob `NOT sfach = "Psychologie" AND alter = 18` dem Ausdruck `(NOT sfach = "Psychologie") AND alter = 18` entspricht (dies wären alle 18-jährigen, die nicht Psychologie studieren) oder dem Ausdruck `NOT (sfach = "Psychologie" AND alter = 18)` entspricht (dies wären alle, die keine 18-jährigen Psychologiestudierenden sind. Dies wiederum wären alle, die entweder nicht Psychologie studieren oder nicht 18 Jahre alt sind oder beides nicht.)

- Diese Priorität ist bei SPSS wie folgt:

Klammern ()	Funktionen	**	* /	+ -	> < <> = >= <=	NOT	AND	OR
-------------	------------	----	--------	--------	-------------------	-----	-----	----

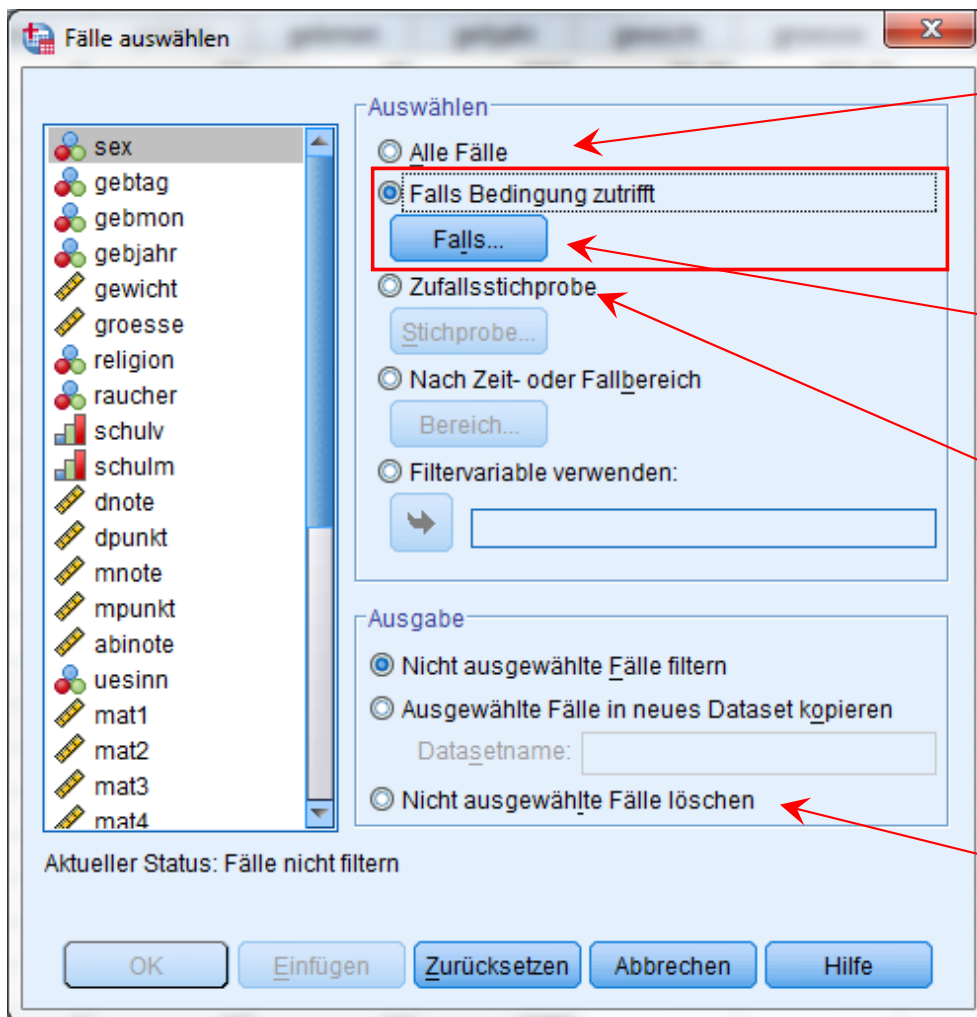
← höchste Priorität

niedrigste Priorität →

- Wenn man sich nicht sicher ist, helfen also immer Klammern (weil sie die höchste Priorität haben).

# SPSS: Fallauswahl

Die Auswahl von Fällen erfolgt unter Daten/Fälle auswählen... (Hinweis: In älteren SPSS-Versionen stehen die Menüs Daten und Transformieren nur aus dem Dateneditor heraus zur Verfügung.)



Zur Aufhebung der Auswahl die Option "Alle Fälle" aktivieren.

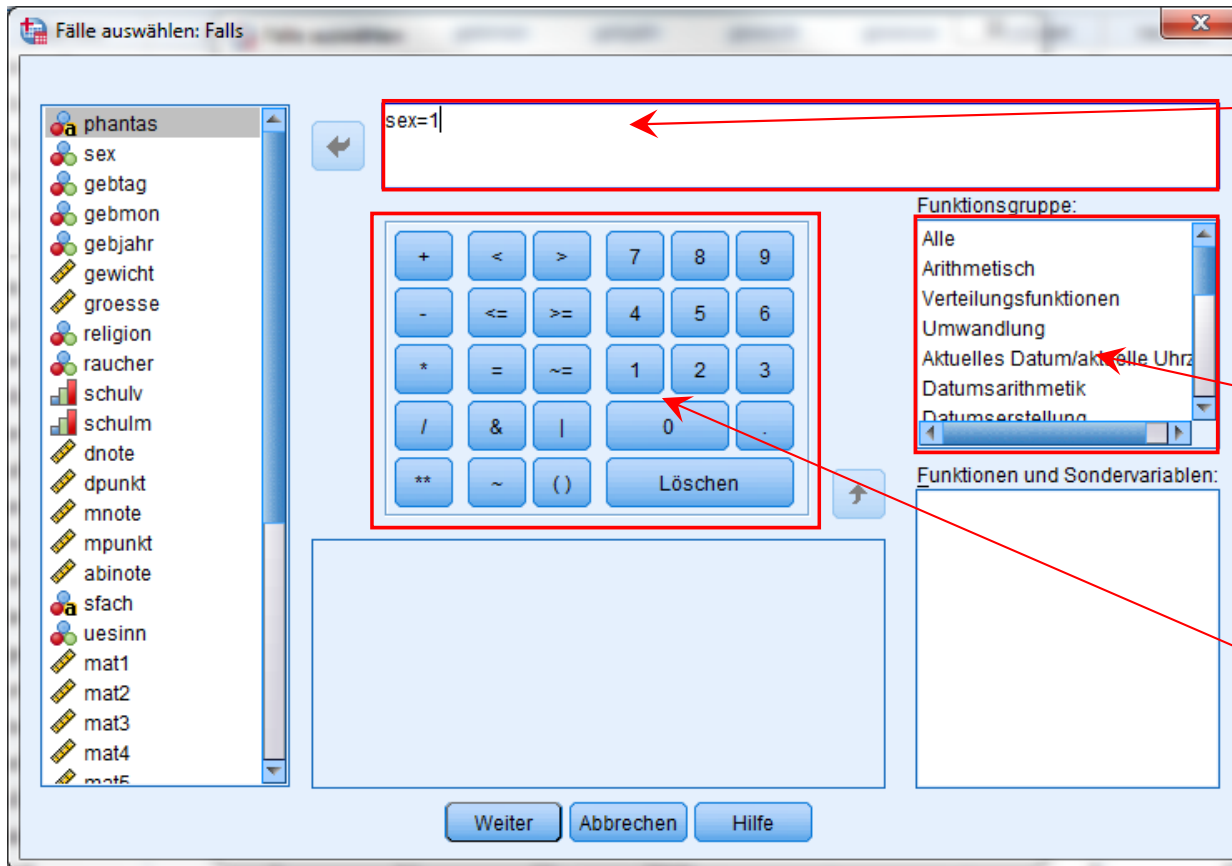
Zur Auswahl einer Teilstichprobe von Fällen hier die Einstellung "Falls Bedingung zutrifft" anklicken und dann auf den Button (Falls...) klicken.

Hier kann auch eine zufällige Auswahl von Fällen angefordert werden (dazu muss der dann ungefähr realisierte Prozentsatz der auszuwählenden Fälle angegeben werden)

Hier können Fälle nicht nur temporär selektiert, sondern permanent aus der Datei gelöscht werden (selten sinnvoll).

# SPSS: Fallauswahl

Nach dem Drücken des Buttons (Fal | s...) im vorangegangenen Dialog erscheint:



In dieses Fenster wird die Bedingung eingetragen, die die selektierte Gruppe erfüllen muss

Hier findet sich eine Liste der in SPSS zur Verfügung stehenden Funktionen.

Statt die Zahlen und Operatoren direkt einzutippen, kann dieses Bedienfeld verwendet werden (meist zu umständlich)

**Wichtig:** Die Fallauswahl gilt anschließend so lange bis entweder (a) eine andere Fallauswahl getroffen wird, (b) eine neue Datendatei geöffnet wird oder (c) die Auswahl unter Daten/Fälle auswählen... explizit wieder aufgehoben wird.

# SPSS: Fallauswahl

Im Dateneditor kann man ganz links an der Fallnummer erkennen, ob ein Fall ausgewählt ist oder nicht (in diesem Fall ist er durchgestrichen).

Bsp. rechts: die  
Bedingung ist `sex=1`

gültige Fälle

The screenshot shows the IBM SPSS Statistics Dateneditor interface. The main window displays a data table with columns: phantas, sex, gebtag, gebmon, gebjahr, gewicht, and groesse. The rows are numbered 1 through 19. Rows 1 through 6, 8 through 16, and 18 through 19 are crossed out, indicating they are filtered out. Rows 7 and 17 are not crossed out, indicating they are the 'gültige Fälle' (valid cases) that match the filter condition 'sex=1'. The status bar at the bottom right of the window shows 'Filter aktiv', indicating that a filter is currently applied.

	phantas	sex	gebtag	gebmon	gebjahr	gewicht	groesse
<del>1</del>	adelesofi	2	24	10	1983	56,00	160,00
<del>2</del>	aida	2	19	11	1976	68,00	168,00
<del>3</del>	Angela	2	14	1	1983	50,00	163,00
<del>4</del>	Apfel	2	8	8	1974	60,00	168,00
<del>5</del>	Apoldina	2	27	6	1978	50,00	153,00
<del>6</del>	atim	2	11	11	1982	65,00	171,00
7	Bacardi	1	20	7	1982	62,00	177,00
<del>8</del>	bachus	2	29				
<del>9</del>	Banumda	2	7				
<del>10</del>	bellatalk	2	1				
<del>11</del>	bienemaja	2	7				
<del>12</del>	Budnikovina	2	3				
<del>13</del>	candis	2	25				
<del>14</del>	Christmas	2	19				
<del>15</del>	cookies	2	22				
<del>16</del>	Elamea	2	1	11	1983	54,00	166,00
17	elgen	1	16	3	1977	74,00	189,00
<del>18</del>	Eumex	2	25	8	1983	60,00	177,00
<del>19</del>	FlowerPower	2	24	6	1983	67,00	181,00

Dass aktuell eine Fallauswahl aktiv ist, erkennt man hier an der Anzeige "Filter aktiv"



# SPSS: Fallauswahl

SPSS realisiert die Fallselektion intern so, dass eine neue Variable mit dem Namen `filter_$` an die Datendatei angefügt wird, die den Wert 1 erhält, wenn der Fall (die Person) selektiert ist und eine 0, falls nicht.

Bsp. rechts: die Bedingung ist `sex=1`

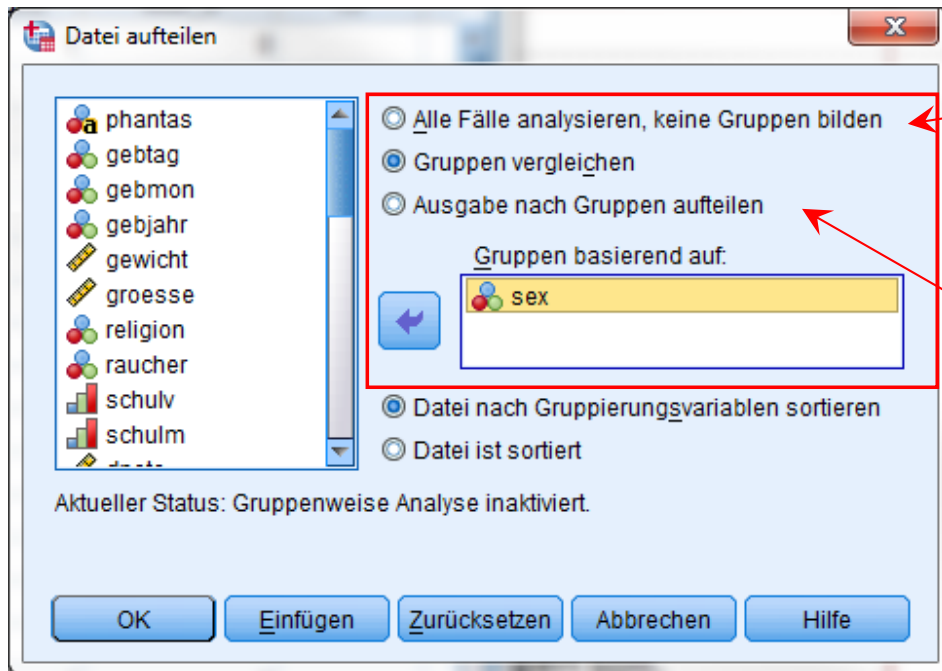
gültige Fälle

The screenshot shows the IBM SPSS Statistics Dateneditor interface. The main window displays a data table with columns for variables: pers5, pers6, pers7, pers8, kohorte, filter \$, and var. The 'filter \$' column contains binary values (0 or 1) indicating case selection. A red box highlights the 'filter \$' column header, and another red box highlights the values 1 in the 'filter \$' column for rows 7 and 17. A red arrow points from the text 'Bsp. rechts: die Bedingung ist sex=1' to the 'filter \$' column header. Another red arrow points from the text 'gültige Fälle' to the rows 7 and 17. The status bar at the bottom indicates 'Filter aktiv'.

	pers5	pers6	pers7	pers8	kohorte	filter \$	var
1	6	5	6	5	1	0	
2	4	4	5	4	1	0	
3	4	4	2	3	1	0	
4	3	3	3	4	1	0	
5	2	2	6	4	1	0	
6	4	3	7	5	1	0	
7	3	3	4	4	1	1	
8	.	.	.	.	1	0	
9	6	5	5	4	1	0	
10	2	3	6	1	1	0	
11	3	4	3	7	1	0	
12	4	3	4	4	1	0	
13	6	6	6	5	1	0	
14	5	4	5	5	1	0	
15	2	2	4	2	1	0	
16	5	5	6	6	1	0	
17	5	4	4	3	1	1	
18	6	5	5	4	1	0	
19	3	3	5	4	1	0	

# SPSS: Schichtung

Die Schichtung von Analysen erfolgt unter Daten/Aufgeteilt Datei :



Zur Aufhebung der Schichtung die Option "Alle Fälle analysieren..." aktivieren.

Zur Schichtung entweder die Option "Gruppen vergleichen" oder "Ausgabe nach Gruppen aufteilen" aktivieren sowie die Variable(n) angeben, nach denen die Schichtung erfolgen soll.

**Wichtig:** Analog zur Fallauswahl (und unabhängig davon!) gilt: Die Schichtung gilt so lange, bis (a) eine andere Schichtung festgelegt wird, (b) eine neue Datendatei geöffnet wird oder (c) diese wieder aufgehoben wird. Eine Schichtung wird in der Fußzeile des Dateneditors ganz rechts angezeigt, z.B. wie hier: Filter aktiv Aufteilen nach sex

Der Unterschied zwischen beiden Optionen besteht darin, dass die untere die Ausgabe der Ergebnisse der nachfolgenden Analysen in getrennten Tabellen und die obere in einer Tabelle bewirkt (siehe folgendes Beispiel)

# SPSS: Schichtung

Ausgabe deskriptiver Statistiken der Variable `gewicht` bei Schichtung nach der Variable `religion` (ohne zusätzliche Selektion) mittels der Option "Gruppen vergleichen"

## Deskriptive Statistik

religion		N	Minimum	Maximum	Mittelwert	Standardabweichung
.	gewicht	1	58,00	58,00	58,0000	.
	Gültige Werte (Listenweise)	1				
1 evangelisch	gewicht	185	40,00	173,00	64,6541	14,17895
	Gültige Werte (Listenweise)	185				
2 katholisch	gewicht	175	41,00	99,00	63,8914	10,09750
	Gültige Werte (Listenweise)	175				
3 keine	gewicht	93	48,00	100,00	66,5484	10,84728
	Gültige Werte (Listenweise)	93				
4 sonstige	gewicht	25	45,00	92,00	62,5200	11,59784
	Gültige Werte (Listenweise)	25				

Ausgabe deskriptiver Statistiken der Variable `gewicht` bei Schichtung nach der Variable `religion` (ohne Selektion) mittels der Option „Ausgabe nach Gruppen aufteilen“

**religion Religionszugehörigkeit = .**

Deskriptive Statistik<sup>a</sup>

	N	Minimum	Maximum	Mittelwert	Standardabweichung
gewicht	1	58,00	58,00	58,0000	.
Gültige Werte (Listenweise)	1				

a. religion = .

**religion Religionszugehörigkeit = 1 evangelisch**

Deskriptive Statistik<sup>a</sup>

	N	Minimum	Maximum	Mittelwert	Standardabweichung
gewicht	185	40,00	173,00	64,6541	14,17895
Gültige Werte (Listenweise)	185				

a. religion = 1 evangelisch

... usw. ...

**religion Religionszugehörigkeit = 4 sonstige**

Deskriptive Statistik<sup>a</sup>

	N	Minimum	Maximum	Mittelwert	Standardabweichung
gewicht	25	45,00	92,00	62,5200	11,59784
Gültige Werte (Listenweise)	25				

a. religion = 4 sonstige