

## **Datentransformationen und Berechnungen**



Mit Hilfe von Transformationen werden die ursprünglichen Rohdaten verändert.

Mit Hilfe von Umcodierungen werden die Ausprägungen von Variablen der Rohdaten nach bestimmten Regeln geändert.

Mittels Berechnungen können Rechenoperationen unter Einbezug von Variablen für jedes Objekt durchgeführt werden

Umcodierungen und Berechnungen können nicht rückgängig gemacht werden!  
Es besteht nur die Möglichkeit den (nicht überschriebenen) Datensatz neu zu laden.

Im Zuge der Datenanalyse sind i.d.R. eine Reihe von Umcodierungen und Berechnungen nötig, die zumeist auch nicht unabhängig voneinander sind.

Zur Vermeidung von Fehlern und zur Sicherstellung der Übersichtlichkeit wird dringend empfohlen, Umcodierungen und Berechnungen als Syntaxbefehle durchzuführen und diese gegebenenfalls zur wiederholten Durchführung abzuspeichern.

# Umcodierungen

Beispiel:

V98

**26.) Sind Sie Raucher/in?**

- 1  nein, ich habe aufgehört zu rauchen
- 2  nein, ich war immer Nichtraucher/in
- 3  ja, ich rauche täglich ca. \_\_\_\_\_ Zigaretten pro Tag *(Bitte Anzahl angeben)*
- 4  ja, ich rauche mehrmals pro Woche
- 5  ja, ich rauche selten

V99

V98 raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 1,00	34	14,2	14,2	14,2
2,00	104	43,3	43,3	57,5
3,00	59	24,6	24,6	82,1
4,00	24	10,0	10,0	92,1
5,00	19	7,9	7,9	100,0
Gesamt	240	100,0	100,0	

**Problemstellung:** Die Variable soll in eine dichotome Variable umgewandelt werden, welche nur RaucherInnen von NichtraucherInnen unterscheidet.

Beispiel:

V98

26.) Sind Sie Raucher/in?

- 1  nein, ich habe aufgehört zu rauchen
- 2  nein, ich war immer Nichtraucher/in
- 3  ja, ich rauche täglich ca. \_\_\_\_\_ Zigaretten pro Tag (Bitte Anzahl angeben)
- 4  ja, ich rauche mehrmals pro Woche
- 5  ja, ich rauche selten

V99

V98 raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 1,00	34	14,2	14,2	14,2
2,00	104	43,3	43,3	57,5
3,00	59	24,6	24,6	82,1
4,00	24	10,0	10,0	92,1
5,00	19	7,9	7,9	100,0
Gesamt	240	100,0	100,0	

## 2 Möglichkeiten:

(1) **Umcodieren in die selbe Variable:** V98 wird überschrieben

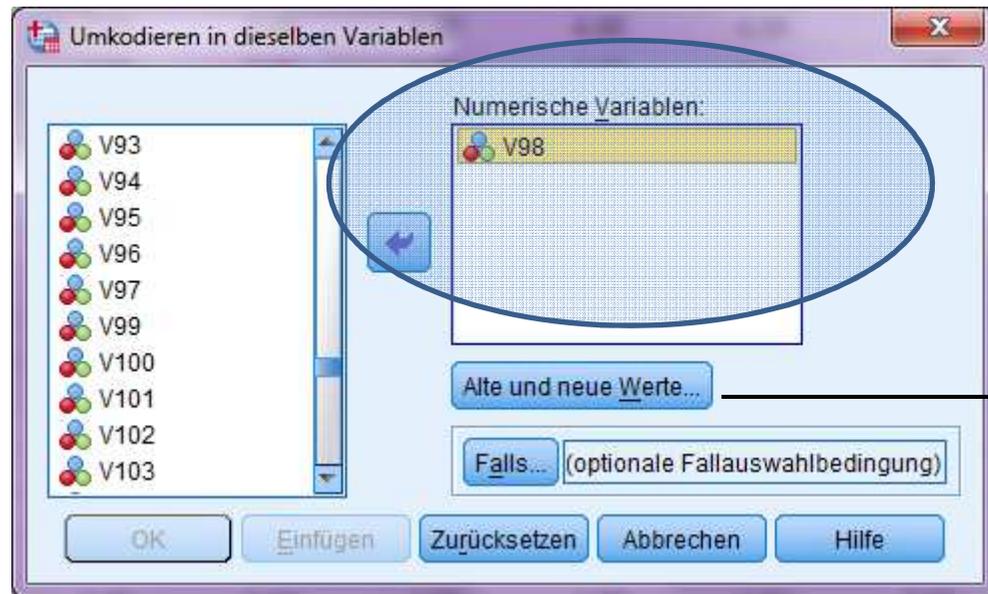
(2) **Umcodierung in eine neue Variable:** V98 bleibt erhalten; eine neue Variable (mit einem zu definierendem Namen) wird erstellt, welche die neue Codierung enthält.

V98

26.) Sind Sie Raucher/in?

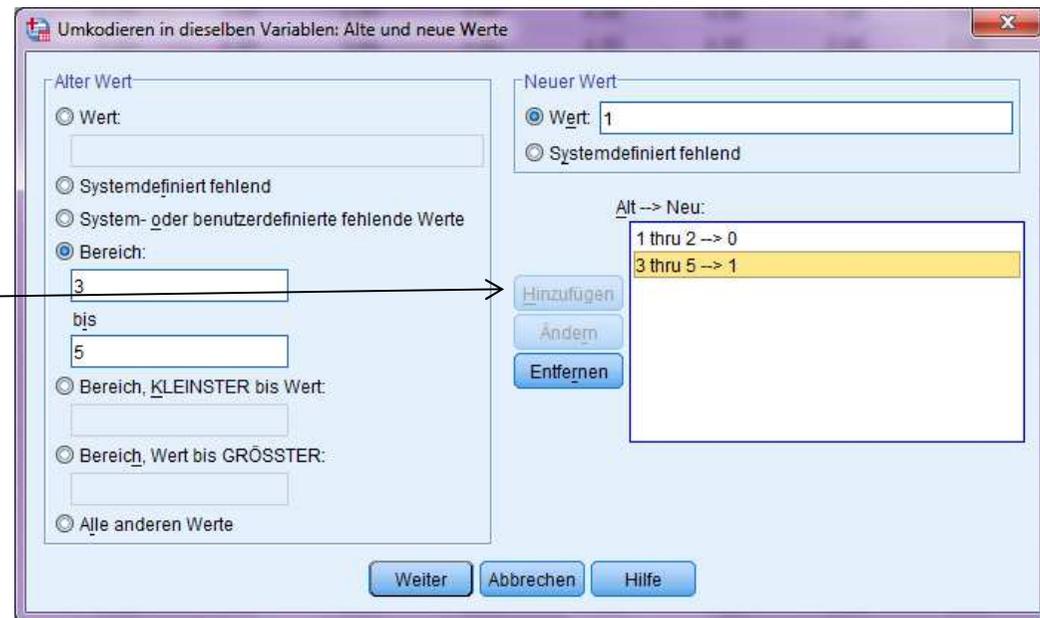
- 1  nein, ich habe aufgehört zu rauchen
- 2  nein, ich war immer Nichtraucher/in
- 3  ja, ich rauche täglich ca. \_\_\_\_\_
- 4  ja, ich rauche mehrmals pro Woche
- 5  ja, ich rauche selten

„Transformieren“->  
„Umcodieren in dieselben Variablen“



Jeweils alten und neuen Wert definieren und „Hinzufügen“ klicken

Anschließend „weiter“ und „Ok“, aber besser „Einfügen“ klicken.



Bei "Einfügen" wird folgende **Syntax erstellt:**

```
DATASET ACTIVATE DatenSet2.  
RECODE V98 (1 thru 2=0) (3 thru 5=1).  
EXECUTE.
```

nicht nötig: Wenn mehrere Datenfiles geöffnet sind, dann wird hier der entsprechende Datensatz ausgewählt.

Eigentlicher Umcodierungs-Befehl

Dies bewirkt nur, dass der Umcodierungs-Befehl sofort ausgeführt wird. Wird execute weggelassen, wird der Umcodierungsbefehl nicht sofort aber vor dem nächsten Befehl ausgeführt.

Alternativ könnte folgende Syntax geschrieben werden:

```
recode v98(1,2=0)(3,4,5=1).  
frequ v98.
```

Execute kann hier entfallen, weil gleich darauf eine Häufigkeitsverteilung zur Kontrolle der Umcodierung angefordert wird.

Nicht vergessen: „Ausführen“-> „alles“

## Resultat:

```
recode v98 (1,2=0) (3,4,5=1) .  
frequ v98.
```

### → Häufigkeiten

[DatenSet2] C:\lva\ws11-12\praktikum\daten\gesamt-corr.sav

#### Statistiken

V98 raucher

N	Gültig	240
	Fehlend	0

#### V98 raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,00	138	57,5	57,5	57,5
1,00	102	42,5	42,5	100,0
Gesamt	240	100,0	100,0	

## Weitere Möglichkeiten für den Recode-Befehl:

recode v57 to v61 (1=4)(2=3)(3=2)(4=1).

Richtung der Variablen 57 bis 61 mit jeweils 4 Ausprägungen wird umgedreht.

recode v5, v14 (10 thru highest=sysmis).

Ausprägungen von 10 bis zum höchsten aufgetretenem Wert wird bei Variable 5 und 14 auf fehlend gesetzt.

recode v5 (10 thru hi=sysmis).

hi=Abkürzung für highest. Kleinster Wert wäre lowest (lo).

recode v89 (sysmis=0).

System-Fehlende Werte werden auf 0 gesetzt. Benutzerdefiniert fehlende Werte werden nicht geändert!

recode v98(1,2=0)(else=1).

Alle Ausprägungen außer 1 und 2 werden zu 1. Vorsicht: Alle fehlende Werte sind „else“!

## Vorsicht bei Missing Values:

`recode v1 (1 thru hi=1).`

Benutzerdefinierte Missings im Bereich 1 und höher werden entfernt. Sysmis bleibt.

`recode v2 (0=0)(else=1).`

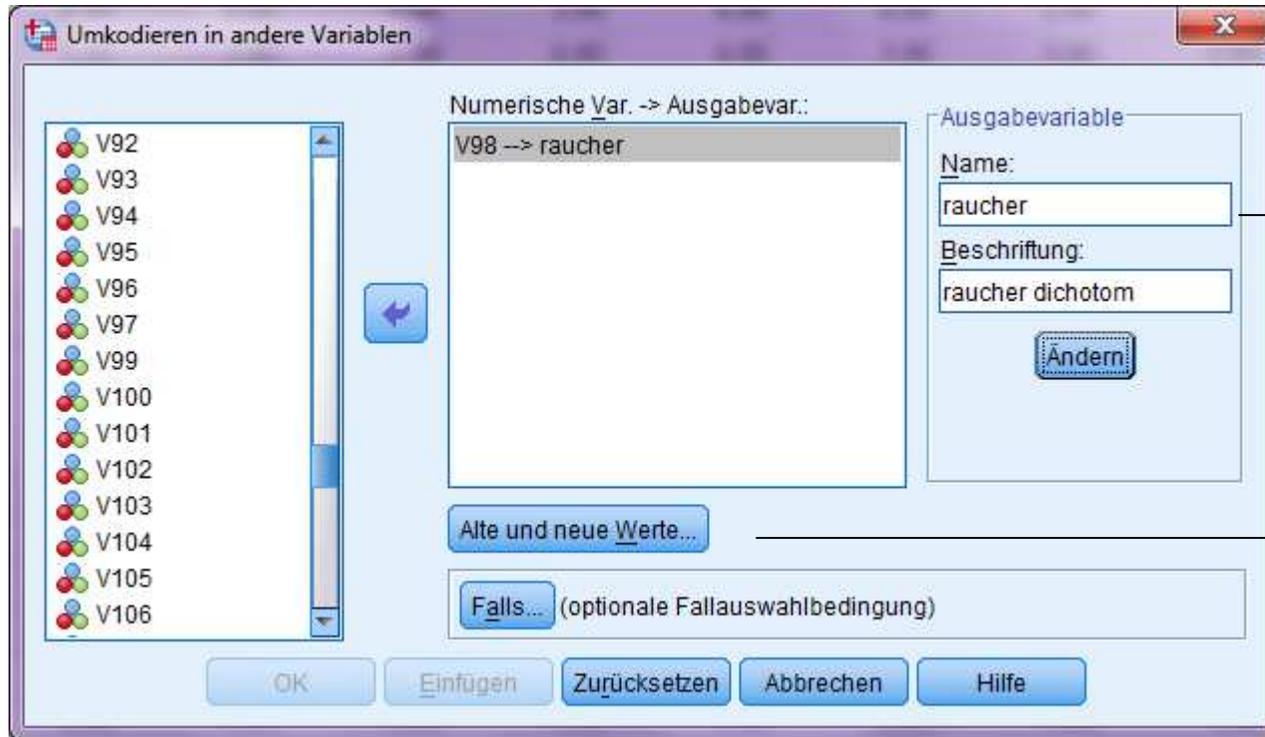
Die Variable hat nach Umcodierung keine Missings mehr, da "else" alles umfasst, was nicht 0 ist! (auch sysmis)

Empfehlung:

- Schlüsselwort „else“ vermeiden.
- Benutzerdefiniert fehlende Werte vermeiden (bringt auch in anderen Bereichen Probleme)
- Immer zur Kontrolle eine Häufigkeitsauszählung nach der Umcodierung durchführen.

## Umcodieren in eine andere Variable

„Transformieren“-> „Umcodieren in eine andere Variable“

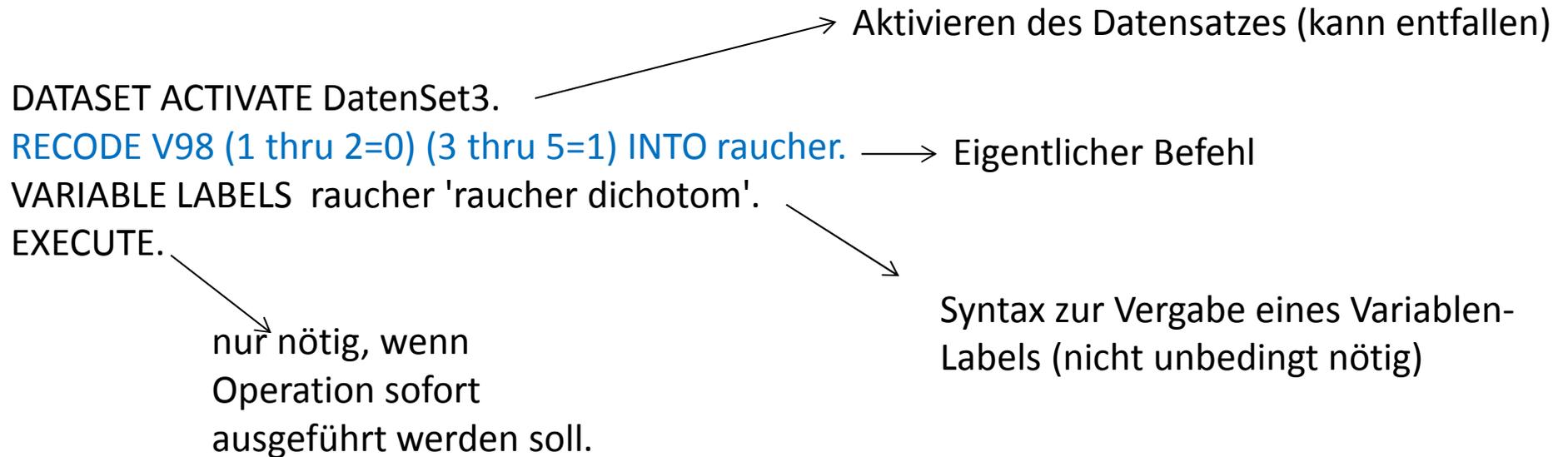


Neuen Variablennamen und Label (optional) angeben. Anschließend „ändern“ klicken.

Anschließend „Alte und neue Werte“ klicken. Hier selbes Prozedere wie bei umcodieren in dieselbe Variable.

V98 bleibt unverändert. Die neue Variable Raucher wird als letzte Variable im Datenfile hinzugefügt.

Bei "Einfügen" wird folgende **Syntax** erstellt:



---

Alternativ könnte im Syntax-Fenster geschrieben werden:

```
recode v98(1,2=0)(3,4,5=1) into raucher.  
frequ v98, raucher.
```

Ergebnis:

recode v98(1,2=0)(3,4,5=1) into raucher.  
 frequ v98, raucher.

→ Häufigkeiten

[DatenSet4] C:\lva\ws11-12\praktikum\daten\gesamt-corr.sav

**Statistiken**

		V98 raucher	raucher
N	Gültig	240	240
	Fehlend	0	0

**Häufigkeitstabelle**

**V98 raucher**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 1,00	34	14,2	14,2	14,2
2,00	104	43,3	43,3	57,5
3,00	59	24,6	24,6	82,1
4,00	24	10,0	10,0	92,1
5,00	19	7,9	7,9	100,0
Gesamt	240	100,0	100,0	

**raucher**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,00	138	57,5	57,5	57,5
1,00	102	42,5	42,5	100,0
Gesamt	240	100,0	100,0	

IBM SPSS Statistics Prozessor ist bereit | H: 0,58, W: 12,67 cm

## Unterschied der beiden Recode-Varianten:

Bei Umcodierung in neue Variable werden jene ursprünglichen Ausprägungen, **welche nicht explizit umcodiert werden auf sysmis gesetzt!**

Bei Umcodierung in dieselbe Variable bleiben jene ursprünglichen Ausprägungen, welche nicht explizit umcodiert werden erhalten.

V98 raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 1,00	34	14,2	14,2	14,2
2,00	104	43,3	43,3	57,5
3,00	59	24,6	24,6	82,1
4,00	24	10,0	10,0	92,1
5,00	19	7,9	7,9	100,0
Gesamt	240	100,0	100,0	

recode v98(1,2=0)(3,4=1) into raucher.

recode v98(1,2=0)(3,4=1) .

raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,00	138	57,5	62,4	62,4
1,00	83	34,6	37,6	100,0
Gesamt	221	92,1	100,0	
Fehlend System	19	7,9		
Gesamt	240	100,0		

V98 raucher

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig ,00	138	57,5	57,5	57,5
1,00	83	34,6	34,6	92,1
5,00	19	7,9	7,9	100,0
Gesamt	240	100,0	100,0	

# Berechnungen

Berechnungen beziehen sich auf **Variablenausprägungen von Objekten**.  
Jedem **Objekt** wird ein **Berechnungsergebnis** zugeordnet.

Häufige Anwendungsfälle:

- Anwendung von **Transformationen und arithmetischen Funktionen** (z.B. Runden, Anwendung von Grundrechnungsarten auf die Ausprägungen von Variablen).
- **Indexberechnungen**

Beispiel:

Befragte wurden mittels 5 Fragen nach Ihrer sozialen Netzwerkgröße gefragt. Diese beziehen sich auf jeweils unterschiedliche Dimensionen von sozialen Relationen:

**4. Von wie vielen Personen kannst Du Unterstützung erwarten, wenn Du Hilfe brauchst (z.B. bei Arbeiten in Deiner Wohnung / Deinem Zimmer oder wenn Du übersiedelst)?**

\_\_\_\_\_ (*bitte Anzahl angeben*)      V18

**5. Von wie vielen Personen kannst Du Dir im Bedarfsfall Geld borgen (mehr als 50€)?**

\_\_\_\_\_ (*bitte Anzahl angeben*)      V19

**6. Mit wie vielen Personen die Du kennst würdest Du über ganz persönliche Dinge sprechen wollen (z.B. wenn Du Sorgen oder Kummer hast)?**

\_\_\_\_\_ (*bitte Anzahl angeben*)      V20

**7. Wie viele Personen kennst Du, auf deren Meinung und Urteil Du bei wichtigen persönlichen Entscheidungen vertraust?**

\_\_\_\_\_ (*bitte Anzahl angeben*)      V21

**8. Denke an die vergangenen 2 Wochen: Mit wie vielen Personen hast Du gemeinsame Freizeitaktivitäten unternommen?**

\_\_\_\_\_ (*bitte Anzahl angeben*)      V22

Frage:

(1) Wie viele Netzwerkpersonen wurden von einer Befragten bei allen fünf Fragen in Summe genannt?

(2) Wie viele Netzwerkpersonen wurden von einer Befragten bei allen fünf Fragen im Durchschnitt genannt?

Zur Beantwortung muss **bei jeder Person** die **Summe der Ausprägungen** von V18 bis V22 **addiert** werden, bzw. diese Summe **durch die Anzahl der Fragen dividiert** werden.

z.B.:

	V18	V19	V20	v21	V22	
Person 1:	5	10	4	2	10	→ Summe=31
Person 2:	2	3	4	1	5	→ Summe=15

## Lösung:

(1) Zunächst wurden die **Häufigkeitsverteilungen** der fünf Variablen inspiziert. Dabei wurden potenzielle **Ausreißer** identifiziert.

In diesem Beispiel:

Bei V18 Ausprägung 50

Bei V20 Ausprägung 79

Bei V22 Ausprägung 60

Es wurde entschieden, diese Werte als ungültig, d.h. **missing** zu codieren.

```
recode v18 (50=sysmis).
```

```
recode v20 (79=sysmis).
```

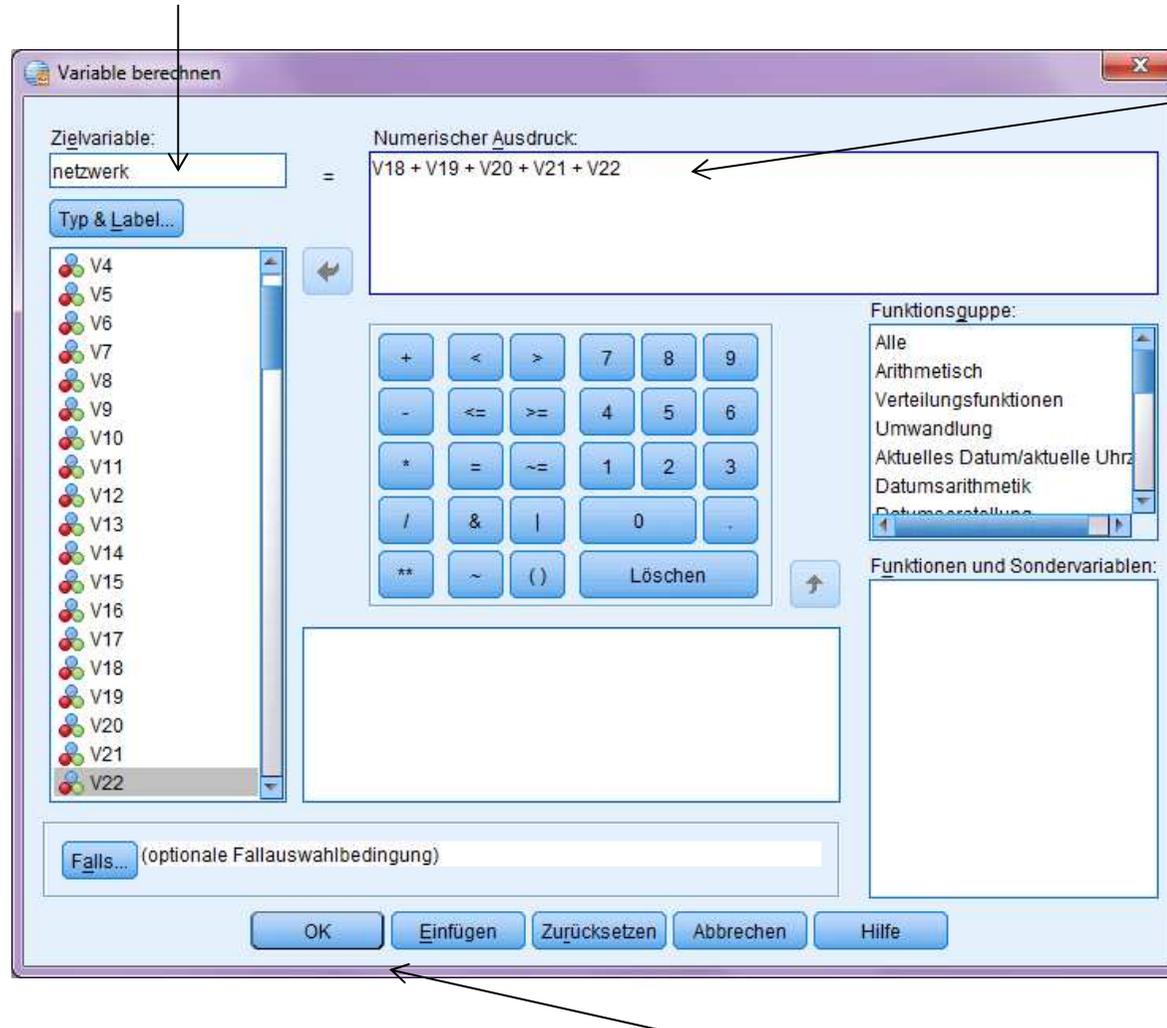
```
recode v22 (60=sysmis).
```

Deskriptive Statistik					
	N	Minimum	Maximum	Mittelwert	Standardabweichung
V18 Unterstützungspersonen	244	1,00	42,00	10,3074	6,28669
V19 Geldpersonen	245	1,00	30,00	6,7510	4,72255
V20 Persönlich_personen	244	,00	25,00	4,6967	2,96451
V21 Urteil_personen	245	,00	25,00	5,1347	3,72880
V22 Freizeit_personen	242	,00	50,00	9,8719	9,66558
Gültige Werte (Listenweise)	240				

(2) Damit nun für jede Person die Summe der fünf Variablen ermittelt werden kann, benötigt man eine **neue Variable**, in der das Ergebnis der Berechnung gespeichert werden kann.

Die Berechnung kann über das Menü mit **„Transformieren“ -> „Variable berechnen“** angefordert werden.

Eine Variable, in der das Ergebnis der Berechnung gespeichert wird muss angegeben werden. Zumeist wird dies eine neue Variable sein. Es kann aber auch eine bereits vorhandene Variable angegeben werden, welche dann überschrieben wird.

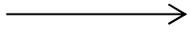


Berechnungsformel angeben.

Variablen, Funktionen, Operatoren können entweder manuell eingegeben, oder durch klicken ausgewählt werden.

Ok – oder besser – Einfügen klicken.

Bei "Einfügen" wird folgende **Syntax** erstellt:

DATASET ACTIVATE DatenSet6.  nicht nötig  
**COMPUTE netzwerk=V18 + V19 + V20 + V21 + V22.**  eigentlicher Befehl  
EXECUTE.  nicht unbedingt nötig.

---

Alternativ könnte in das Syntaxfenster geschrieben werden:

```
comp netzwerk=v18+v19+v20+v21+v22.  
frequ netzwerk.
```

Zur Berechnung des Mittelwertes der angegebenen Netzwerkgrößen bei einer Person:

```
comp netzwerkII=netzwerk/5.
```

**Statistiken**

netzwerk

N	Gültig	240
	Fehlend	6

**netzwerk**

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig 9,00	1	,4	,4	,4
10,00	1	,4	,4	,8
11,00	2	,8	,8	1,7
12,00	1	,4	,4	2,1
13,00	3	1,2	1,3	3,3
14,00	4	1,6	1,7	5,0
15,00	4	1,6	1,7	6,7
16,00	5	2,0	2,1	8,8
17,00	5	2,0	2,1	10,8
18,00	8	3,3	3,3	14,2
19,00	10	4,1	4,2	18,3
20,00	4	1,6	1,7	20,0
21,00	5	2,0	2,1	22,1
22,00	3	1,2	1,3	23,3
23,00	8	3,3	3,3	26,7
24,00	10	4,1	4,2	30,8
25,00	7	2,8	2,9	33,8
26,00	5	2,0	2,1	35,8
27,00	5	2,0	2,1	37,9
28,00	8	3,3	3,3	41,3
29,00	5	2,0	2,1	43,3
30,00	10	4,1	4,2	47,5
31,00	1	,4	,4	47,9

~

81,00	1	,4	,4	96,3
85,00	2	,8	,8	97,1
86,00	1	,4	,4	97,5
93,00	1	,4	,4	97,9
98,00	1	,4	,4	98,3
99,00	1	,4	,4	98,8
108,00	1	,4	,4	99,2
111,00	1	,4	,4	99,6
140,00	1	,4	,4	100,0
Gesamt	240	97,6	100,0	
Fehlend System	6	2,4		
Gesamt	246	100,0		

**Deskriptive Statistik**

	N	Minimum	Maximum	Mittelwert	Standardabweichung
netzwerk	240	9,00	140,00	36,3083	19,64573
Gültige Werte (Listenweise)	240				

Insgesamt bei 6 Befragten fehlt der Summenwert. Diese 6 Personen weisen bei mindestens einer der fünf Netzwerkvariablen einen fehlenden Wert auf.

Sobald einer der Ausdrücke in der Formel missing ist, ist auch das Resultat der Berechnung missing.

## Funktionen sum und mean:

Statt

```
comp netzwerk=v18+v19+v20+v21+v22.
```

könnte man auch die **sum-Funktion** verwenden:

```
comp netzwerk=sum (v18 to v22).
```

zur Berechnung des Mittelwertes die **mean-Funktion**:

```
comp netzwerkII=mean(v18 to v22).
```

Das Resultat der Summenberechnung in dieser zweiten Version liefert dennoch ein anderes Resultat.

### Statistiken

netzwerk

N	Gültig	245
	Fehlend	1

netzwerk					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	8,00	1	,4	,4	,4
	9,00	1	,4	,4	,8
	10,00	1	,4	,4	1,2
	11,00	2	,8	,8	2,0
	12,00	1	,4	,4	2,4
	13,00	3	1,2	1,2	3,7
	14,00	4	1,6	1,6	5,3
	15,00	4	1,6	1,6	6,9
	16,00	5	2,0	2,0	9,0
	17,00	5	2,0	2,0	11,0
	18,00	8	3,3	3,3	14,3
	19,00	10	4,1	4,1	18,4
	20,00	4	1,6	1,6	20,0
	21,00	5	2,0	2,0	22,0
	22,00	3	1,2	1,2	23,3
	23,00	2	,8	,8	24,1
	24,00	1	,4	,4	24,5
	25,00	1	,4	,4	24,9
	26,00	1	,4	,4	25,3
	27,00	1	,4	,4	25,7
	28,00	1	,4	,4	26,1
	29,00	1	,4	,4	26,5
	30,00	1	,4	,4	26,9
	31,00	1	,4	,4	27,3
	32,00	1	,4	,4	27,7
	33,00	1	,4	,4	28,1
	34,00	1	,4	,4	28,5
	35,00	1	,4	,4	28,9
	36,00	1	,4	,4	29,3
	37,00	1	,4	,4	29,7
	38,00	1	,4	,4	30,1
	39,00	1	,4	,4	30,5
	40,00	1	,4	,4	30,9
	41,00	1	,4	,4	31,3
	42,00	1	,4	,4	31,7
	43,00	1	,4	,4	32,1
	44,00	1	,4	,4	32,5
	45,00	1	,4	,4	32,9
	46,00	1	,4	,4	33,3
	47,00	1	,4	,4	33,7
	48,00	1	,4	,4	34,1
	49,00	1	,4	,4	34,5
	50,00	1	,4	,4	34,9
	51,00	1	,4	,4	35,3
	52,00	1	,4	,4	35,7
	53,00	1	,4	,4	36,1
	54,00	1	,4	,4	36,5
	55,00	1	,4	,4	36,9
	56,00	1	,4	,4	37,3
	57,00	1	,4	,4	37,7
	58,00	1	,4	,4	38,1
	59,00	1	,4	,4	38,5
	60,00	1	,4	,4	38,9
	61,00	1	,4	,4	39,3
	62,00	1	,4	,4	39,7
	63,00	1	,4	,4	40,1
	64,00	1	,4	,4	40,5
	65,00	1	,4	,4	40,9
	66,00	1	,4	,4	41,3
	67,00	1	,4	,4	41,7
	68,00	1	,4	,4	42,1
	69,00	1	,4	,4	42,5
	70,00	1	,4	,4	42,9
	71,00	1	,4	,4	43,3
	72,00	1	,4	,4	43,7
	73,00	1	,4	,4	44,1
	74,00	1	,4	,4	44,5
	75,00	1	,4	,4	44,9
	76,00	1	,4	,4	45,3
	77,00	1	,4	,4	45,7
	78,00	1	,4	,4	46,1
	79,00	1	,4	,4	46,5
	80,00	1	,4	,4	46,9
	81,00	1	,4	,4	47,3
	85,00	2	,8	,8	46,7
	86,00	1	,4	,4	47,1
	93,00	1	,4	,4	47,6
	94,00	1	,4	,4	48,0
	98,00	1	,4	,4	48,4
	99,00	1	,4	,4	48,8
	108,00	1	,4	,4	49,2
	111,00	1	,4	,4	49,6
	140,00	1	,4	,4	100,0
	Gesamt	245	99,6	100,0	
	Fehlend System	1	,4		
	Gesamt	246	100,0		

### Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
netzwerk	245	8,00	140,00	36,5796	19,94991
Gültige Werte (Listenweise)	245				

Bei dieser Berechnung liegt nur ein fehlender Wert vor.

Bei der Berechnung mittels „sum“ und „mean“ ist das Resultat **nur** dann missing wenn alle Variablen des Ausdruckes missing sind.

Wenn nicht alle Variablen des Ausdruckes missing sind, wird die Summe bzw. der Mittelwert für die restlichen gültigen Werte berechnet.

Durch einen **zusätzlichen Parameter** bei sum und mean, kann definiert werden, wie viele der Variablen gültige Werte aufweisen müssen, damit die Summe / der Mittelwert berechnet wird.

die Anweisung

`comp netzwerk=sum.5(v18 to v22).`

würde bewirken, dass alle Variablen (V18 bis v22) gültige Werte aufweisen müssen, damit die eine valide Summe berechnet wird.

die Anweisung

`comp netzwerkII=mean.4(v18 to v22).`

würde bewirken, dass der Durchschnitt nur dann berechnet wird, wenn 4 der 5 Variablen einen gültigen Wert aufweisen.

D.h. wenn 5 Variablen gültige Werte aufweisen, wird die Summe der 5 Variablen durch 5 geteilt. Wenn nur 4 Variablen gültige Werte aufweisen, wird die Summe der 4 Variablen durch 4 geteilt.

## Zusammenfassung:

Zur Ermittlung der gesuchten Resultate waren mehrere Schritte nötig.

Wenn wir alle Schritte zu einer gemeinsamen Syntax zusammenfassen, können wir

- die Berechnung jederzeit ohne großen Aufwand replizieren
- später nachvollziehen, was wir berechnet haben:

```
recode v18 (50=sysmis).  
recode v20(79=sysmis).  
recode v22(60=sysmis).  
comp netzwerk=sum.5 (V18 to v22).  
comp netzwerkII=mean.5 (V18 to v22).  
exe.
```

Würde am ende noch

```
comp netzwerkIII=rnd(netzwerkII).
```

eingefügt, dann wird der Mittelwert (ganzzahlig) gerundet.

## Weiteres Beispiel: Indexberechnung / Messungen 2.Ordnung

In den Sozialwissenschaften benötigen wir häufig Messungen von Dimensionen, die sehr großen Fehlerschwankungen unterliegen (Messungen mit geringer Reliabilität) .

Dies ist (unter anderem) ein Grund dafür, dass eine Dimension häufig mittels Fragebatterien gemessen wird, welche aus ähnlich formulierten Items besteht.

Die Idee dahinter ist, dass **jedes einzelne Item eine Messwiederholung darstellt** (aus praktischen Gründen macht es natürlich keinen Sinn, genau dieselbe Frage mehrmals zu stellen).

Wenn anschließend ein **Index** aus diesen einzelnen Items berechnet wird (z.B. der Mittelwert), dann geht man davon aus, dass sich **die einzelnen Messfehler zu einem gewissen Grad ausmitteln** und die Messung genauer (reliabler) wird (**klassische Testtheorie**).

## Beispiel: Skala zur Messung des Selbstwertes

15. Bitte gib an, inwieweit Deine Meinung mit den folgenden Aussagen übereinstimmt.

	trifft voll zu	trifft eher zu	trifft eher nicht zu	trifft gar nicht zu	
Ich habe eine positive Einstellung zu mir selbst.	①	②	③	④	V34
Ich glaube, ich habe eine Menge guter Eigenschaften	①	②	③	④	V35
Ich glaube, es gibt nicht viel, worauf ich stolz sein kann.	①	②	③	④	V36
Ich wünschte, ich hätte mehr Achtung vor mir selbst.	①	②	③	④	V37
Manchmal denke ich, dass ich für überhaupt nichts gut bin.	①	②	③	④	V38

Eine Messung 2.Ordnung wäre beispielsweise der **gemittelte additive Summenscore** (=Mittelwert aller Items).

Anmerkung: In der Praxis wird zuvor die Messeigenschaft der Skala (Faktorenanalyse, Reliabilitätsanalyse -> DV2) überprüft.

Damit der gemittelte additive Summenscore ein sinnvolles Resultat liefert, müssen alle **Items in dieselbe Richtung codiert** sein (in welche Richtung ist egal, sofern man die Richtung weiß)!

15. Bitte gib an, inwieweit Deine Meinung mit den folgenden Aussagen übereinstimmt.

	trifft voll zu	trifft eher zu	trifft eher nicht zu	trifft gar nicht zu	
Ich habe eine positive Einstellung zu mir selbst.	①	②	③	④	V34
Ich glaube, ich habe eine Menge guter Eigenschaften	①	②	③	④	V35
Ich glaube, es gibt nicht viel, worauf ich stolz sein kann.	①	②	③	④	V36
Ich wünschte, ich hätte mehr Achtung vor mir selbst.	①	②	③	④	V37
Manchmal denke ich, dass ich für überhaupt nichts gut bin.	①	②	③	④	V38

recode v36 to v38 (1=4)(2=3)(3=2)(4=1).  
 comp selbstwert=mean.4(v34 to v38).



## Weiteres Beispiel: Umcodierung in eine andere Variable mittels compute

Umcodieren in eine andere Variable könnte auch ohne dem (heiklen) „recode – into“ Befehl realisiert werden:

```
recode v98(1,2=0)(3,4,5=1) into raucher.
```

hier gäbe es beispielsweise Probleme, wenn die Variable "raucher" bereits existiert und nicht alle Ausprägungen explizit adressiert werden. Z.B. Variable „raucher“ besteht bereits und weist bereits durchgängig Werte auf, dann werden die Missing-Werte von V98 nicht in die Variable raucher übernommen. Schlimmer: Statt dessen bleiben die schon vorhandenen Ausprägungen!

Alternative:

```
comp raucher=v98.  
recode raucher(1,2=0)(3,4,5=1).
```

Kopiert sämtliche Inhalte von V98 nach raucher.  
Eventuell bestehende Inhalte werden überschrieben.