

Imputation fehlender Werte mit NORM 2.03

E(xpectation)-M(aximization)-
Algorithmus

Wozu?

- Optimaler Umgang mit fehlenden Werten
- geringer Varianzverlust
- vollständige Datensätze sind Voraussetzung für Strukturgleichungssoftware
- EM-Modell → 1 Datensatz
- multiple Imputation → x Datensätze

Download

- homepage B. Streicher → links
- <http://www.stat.psu.edu/~jls/misoftwa.html> → Stand-alone packages for Windows 95/98/NT

Voraussetzungen

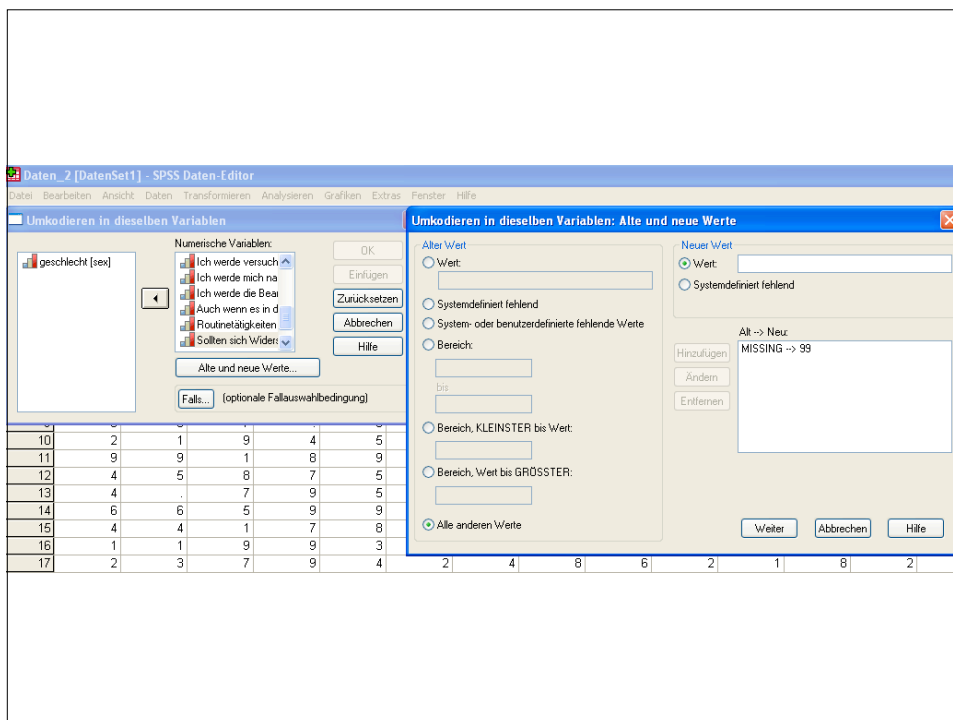
- Personen mit mehr als 30% missings ausschließen (z.B. "count" oder "nmis")
- nur numerische Variablen im Datensatz (d.h. z.B. keine string-Variablen)
- nur intervallskalierte Variablen für Imputation verwendbar
- alle Variablen müssen einen Wert haben

1. Missings im Datensatz ergänzen

- alle missings durch einen Wert (z.B. -99) ersetzen.
- Recode → into same variable
→ old: system- or user-missing
→ new: -99

RECODE

```
*alle Variablen* (MISSING=-99) .  
EXECUTE .
```

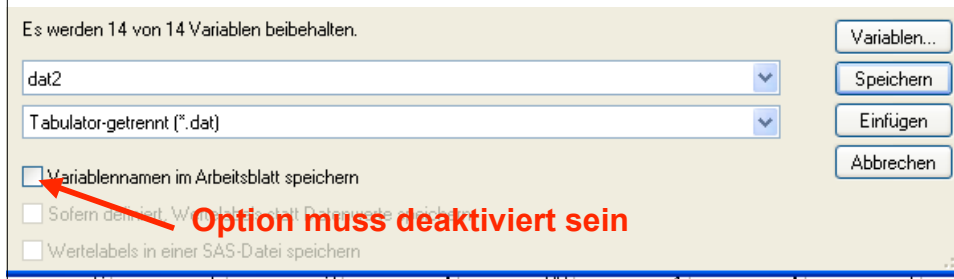


The screenshot shows the SPSS Recode dialog box for the variable 'geschlecht [sex]'. The 'Umkodieren in dieselben Variablen' (Recode into same variables) option is selected. The 'Alte und neue Werte...' (Old and new values...) button is highlighted. The 'Umkodieren in dieselben Variablen: Alte und neue Werte' (Recode into same variables: Old and new values) sub-dialog is open, showing the 'Alle anderen Werte' (All other values) option selected. The 'Neuer Wert' (New value) field is set to 'MISSING -> 99'. The background shows a data editor window with a table of data.

	1	2	3	4	5
10	2	1	9	4	5
11	9	9	1	8	9
12	4	5	8	7	5
13	4	..	7	9	5
14	6	6	5	9	9
15	4	4	1	7	8
16	1	1	9	9	3
17	2	3	7	9	4

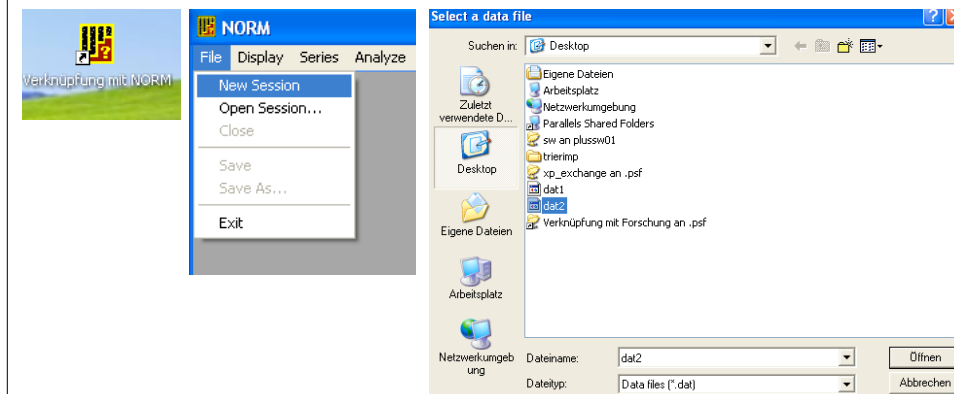
2. Datensatz speichern

- Datensatz als tab-getrennte .dat - Datei abspeichern (ASCII).

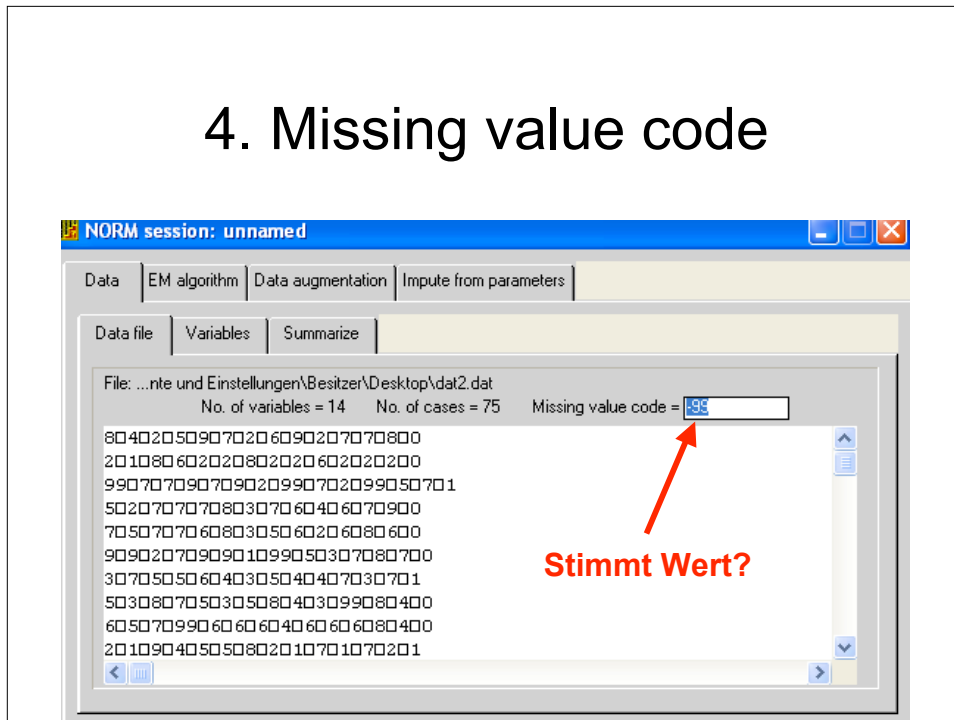


3. Norm starten

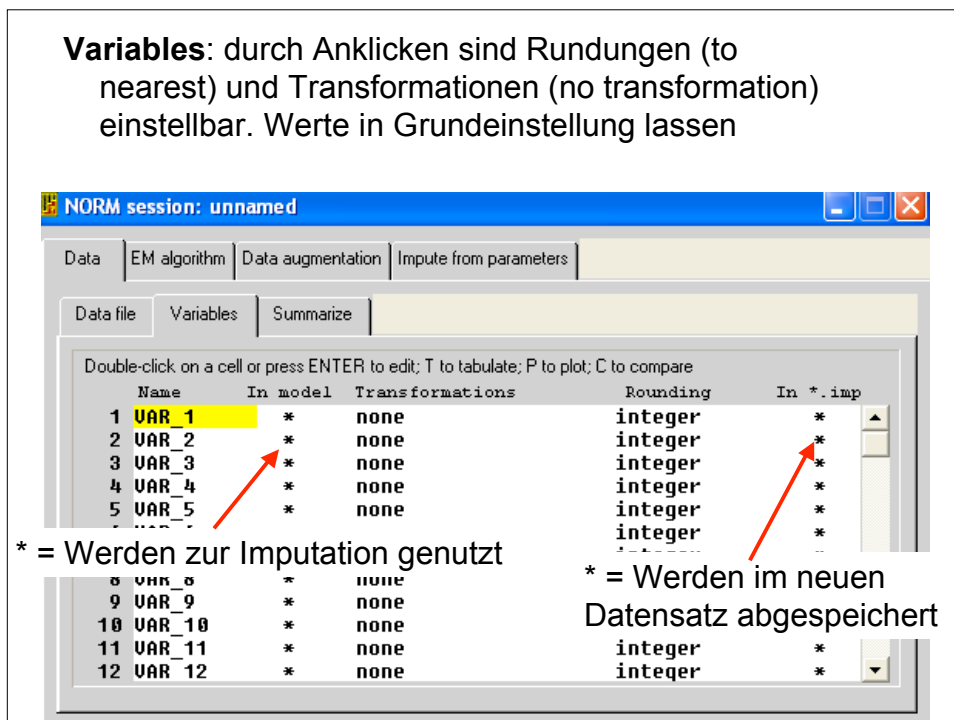
- Norm starten → new session → Datenfile (*.dat) öffnen



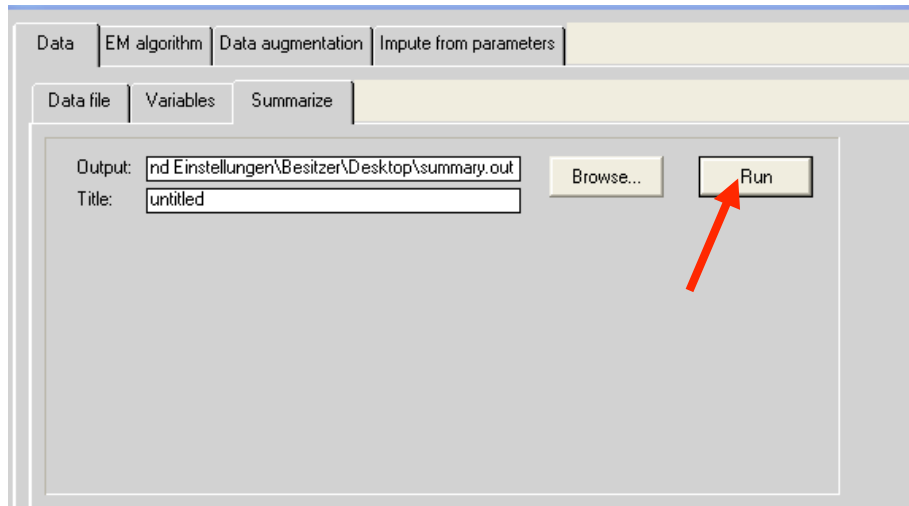
4. Missing value code



Variables: durch Anklicken sind Rundungen (to nearest) und Transformationen (no transformation) einstellbar. Werte in Grundeinstellung lassen



Summarize: durch Anklicken von "run" lässt sich eine Zusammenfassung des Datensatzes anzeigen



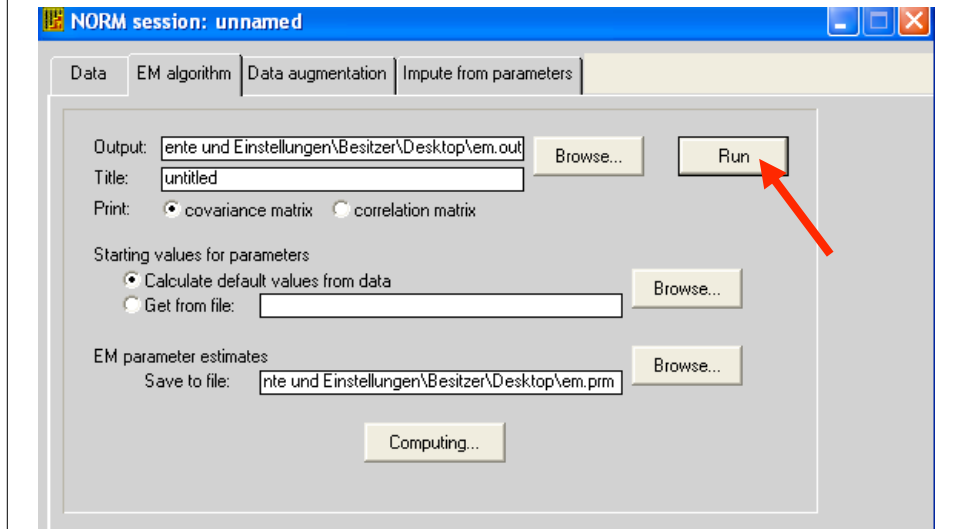
NUMBER OF OBSERVATIONS =	75	
NUMBER OF VARIABLES =	14	
	NUMBER MISSING	% MISSING
VAR_1	4	5.33
VAR_2	3	4.00
VAR_3	3	4.00
VAR_4	4	5.33
VAR_5	5	6.67
VAR_6	0	0.00
VAR_7	2	2.67
VAR_8	7	9.33
VAR_9	3	4.00
VAR_10	1	1.33
VAR_11	7	9.33
VAR_12	4	5.33
VAR_13	1	1.33
VAR_14	0	0.00

höchsten Wert notieren

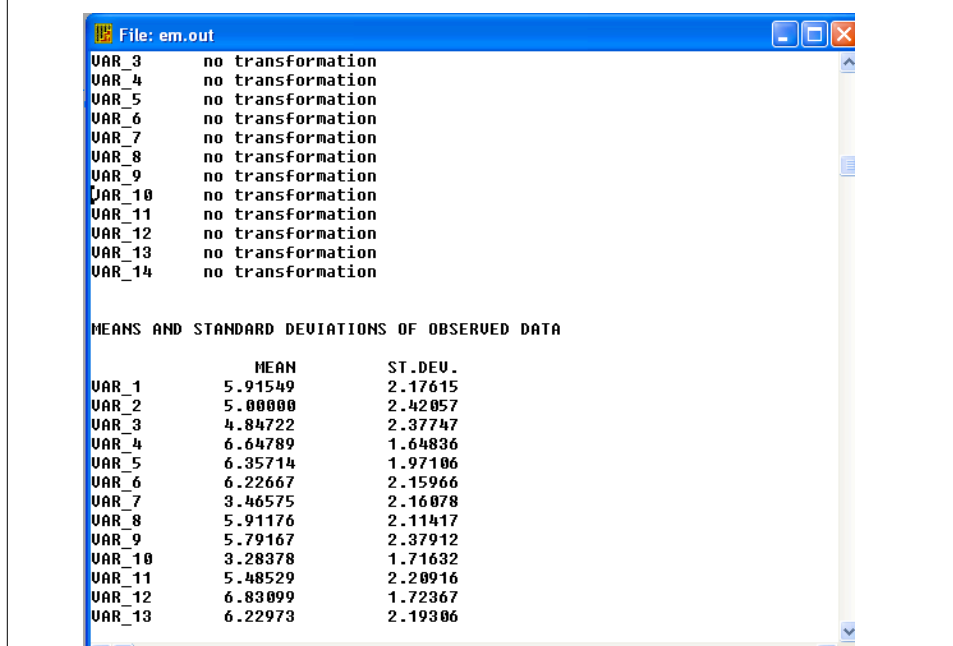
5. Ersetzung durch EM-Algorithmus:

Reiter "EM algorithm" → "Run"

(bei Computing kann das Konvergenz-Kriterium eingestellt werden, was aber nicht nötig ist)



Ergebnis: Imputierte Parameter werden angezeigt



6. Erzeugen des imputierten Datensatzes:

Reiter "impute from parameter" → "Run"

Im vorherigen Schritt erzeugte Parameter

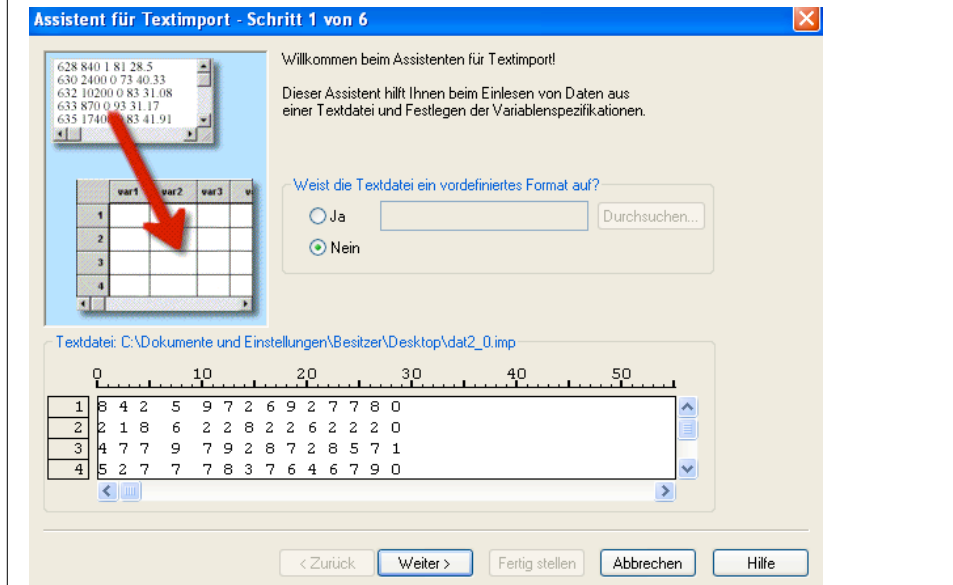
Hier wird imputierter Datensatz gespeichert (*.imp)

Imputierter Datensatz kann in Norm betrachtet werden:

Display → Imputation file

8	4	2	5	9	7	2	6	9	2	7	7	8	0
2	1	8	6	2	2	8	2	2	6	2	2	2	0
4	7	7	9	7	9	2	8	7	2	8	5	7	1
5	2	7	7	7	8	3	7	6	4	6	7	9	0
7	5	7	7	6	8	3	5	6	2	6	8	6	0
9	9	2	7	9	9	1	7	5	3	7	8	7	0
3	7	5	5	6	4	3	5	4	4	7	3	7	1
5	3	8	7	5	3	5	8	4	3	4	8	4	0
6	5	7	9	6	6	6	4	6	6	6	8	4	0
2	1	9	4	5	5	8	2	1	7	1	7	2	1
9	9	1	8	9	9	1	9	9	1	8	9	9	0
4	5	8	7	5	4	5	6	3	1	7	9	8	0
4	2	7	9	5	4	7	3	3	4	3	8	7	0
6	6	5	9	9	6	1	9	3	5	2	6	2	0
4	4	1	7	8	7	1	3	9	0	9	5	9	0
1	1	9	9	3	1	3	9	3	1	1	9	1	1
2	3	7	9	4	2	4	8	6	2	1	8	2	0
7	5	5	7	6	5	6	6	5	3	7	7	6	0
9	8	2	8	9	9	5	9	9	5	9	8	8	0
7	5	2	5	8	5	4	6	4	4	5	6	8	0

Assistent für Import öffnet sich
 normalerweise können alle Voreinstellungen
 übernommen werden



9. Einfügen der Variablennamen in den imputierten Datensatzes in SPSS

	Name	Typ	Spaltenformat	De
1	V1	Numerisch	1	0
2	V2	Numerisch	1	0
3	V3	Numerisch	1	0
4	V4	Numerisch	1	0
5	V5	Numerisch	2	1
6	V6	Numerisch	1	0
7	V7	Numerisch	1	0
8	V8	Numerisch	1	0
9	V9	Numerisch	1	0
10	V10	Numerisch	1	0
11	V11	Numerisch	1	0

Berichtbeispiel

Materials and Methods

Participants

Participants with more than 30% of missing values were excluded from analysis. The remaining maximum percentage of missing values in a single variable was 5.73%. Missing values of other participants were imputed using expectation-maximization algorithm (EMA) by the Software Norm (Version 2.03) leaving two hundred twenty-seven employees (age: $M = 36.71$, $SD = 9.66$; 98 female, 129 male) for analysis. Participants had worked for an average of 7.68 years ($SD = 7.17$) for their company and had a supervisor, were recruited as volunteers from a range of different companies. Branches comprised services (55 persons, 24.2 %), healthcare and welfare (26 persons, 11.5 %), manufacturing industry (24 persons, 10.6 %), banking and insurance industry (19 persons, 8.4 %), education and tuition (29 persons, 12.8 %), and other sectors (74 persons, 32.5 %).