

1 Datenaufbereitung in SPSS

In diesem Kapitel wird besprochen, was zu beachten ist, wenn man einen SPSS-Datensatz zum Einlesen in Mplus aufbereiten möchte. Das Programm SPSS wurde als Beispiel gewählt, da es in den Sozialwissenschaften zu den am weitesten verbreiteten Statistik-Software-Packages zu zählen ist. Die gezeigten Schritte lassen sich jedoch problemlos auch auf andere Statistikprogramme wie z.B. SAS, STATA, STATISTICA, Systat usw. übertragen.

Mplus kann keine Datensätze im SPSS-eigenen Datenformat (Dateiendung *.sav) korrekt verarbeiten. Man kann aber leicht mit Hilfe von SPSS (oder anderen gängigen Statistikprogrammen) Rohdatensätze (Individualdaten) als einfache Textdateien (sog. ASCII-Daten, *.txt oder *.dat) abspeichern lassen, die dann von Mplus gelesen werden können. Zu empfehlen ist hierbei insbesondere die Verwendung eines tabulator-getrennten („tab-delimited“) Formates in den ASCII-Dateien. Bei diesem Format sind die einzelnen Variablen Spalten mit Tabulatorzeichen voneinander getrennt. Aber auch andere Formate sind möglich (siehe dazu Mplus User's Guide; Muthén & Muthén, 1998-2007), wie auch das Einlesen von Summary-Daten (keine Rohdaten, sondern z.B. Kovarianz- oder Korrelationsmatrizen; siehe Kapitel 2.2).

Die aus Sicht des Autors schnellste, praktischste und fehlerunanfällige Variante der Aufbereitung von Individualdaten in SPSS und Abspeicherung einer ASCII-Datendatei wird im Folgenden Schritt für Schritt anhand des Beispieldatensatzes **KFT.sav** beschrieben. Es sei darauf hingewiesen, dass viele Wege nach Rom führen, d.h. das hier vorgeschlagene Verfahren ist nur eines von mehreren möglichen. Die Datendatei **KFT.sav** sowie eine SPSS-Syntaxdatei mit den entsprechenden Syntaxbefehlen befinden sich auf der Begleit-CD zum Buch.

Box 1.1: Datenbeispiel KFT.sav

Der Datensatz **KFT.sav** enthält vollständige Daten von 455 Schülern, die mit dem Kognitiven Fähigkeitstest (KFT; Heller, Gaedicke & Weinläder, 1976) getestet wurden. Der KFT ist ein schulnäher Intelligenztest. Die Schüler mussten jeweils Aufgaben zum logisch-schlussfolgernden Denken bearbeiten, wobei verschiedene Untertests mit verbalem, numerischem und figuralem Stimulusmaterial eingesetzt wurden: Zwei verbale Untertests (Variablen **kft_v1** und **kft_v3**), zwei numerische („quantitative“) Untertests (Variablen **kft_q1** und **kft_q3**), und zwei figurale („nonverbale“) Untertests (Variablen **kft_n1** und **kft_n3**). Die erreichten Punkte der Schüler wurden für jeden Untertest jeweils zu einem Summenscore aggregiert.

1.1 Ländereinstellungen

Zu beachten ist, dass Mplus als Dezimaltrennzeichen keine Kommas (so wie es im Deutschen üblich ist), sondern nur Punkte akzeptiert. Entsprechend muss darauf geachtet werden,

1 : kft_v1	18					Sichtbar: 6 von 6			
	kft_v1	kft_v3	kft_q1	kft_q3	kft_n1	kft_n3	var	var	var
10	6,00	9,00	9,00	3,00	7,00	12,00			
11	13,00	16,00	12,00	6,00	7,00	10,00			
12	7,00	7,00	7,00	4,00	7,00	5,00			
13	10,00	5,00	7,00	2,00	5,00	4,00			
14	5,00	5,00	10,00	4,00	5,00	4,00			
15	8,00	9,00	7,00	2,00	6,00	5,00			
16	17,00	13,00	12,00	5,00	16,00	12,00			
17	9,00	9,00	10,00	6,00	9,00	14,00			
18	12,00	10,00	8,00	5,00	5,00	7,00			
19	16,00	5,00	12,00	12,00	21,00	11,00			
20	20,00	14,00	15,00	10,00	15,00	13,00			
21	20,00	10,00	14,00	13,00	22,00	11,00			
22	7,00	6,00	9,00	8,00	6,00	5,00			
23	2,00	4,00	10,00	5,00	6,00	5,00			
24	17,00	14,00	16,00	9,00	20,00	9,00			
25	12,00	6,00	12,00	6,00	8,00	9,00			
26	14,00	5,00	8,00	7,00	9,00	9,00			
27	11,00	5,00	7,00	4,00	6,00	11,00			
28	7,00	4,00	10,00	4,00	5,00	8,00			
29	14,00	6,00	9,00	8,00	9,00	7,00			
30	9,00	10,00	12,00	6,00	10,00	7,00			
31	20,00	4,00	16,00	9,00	5,00	4,00			
32	14,00	3,00	15,00	8,00	7,00	8,00			
33			
34			
35			
36	8,00	3,00	16,00	3,00	2,00	3,00			
37	11,00	15,00	13,00	6,00	4,00	12,00			
38	4,00	2,00	8,00	4,00	6,00	10,00			

Abbildung 1.1. Ausschnitt aus dem SPSS-Beispieldatensatz mit Schülerdaten für 6 Subskalen aus dem Kognitiven Fähigkeitstest (KFT; Heller, Gaedicke & Weinläder, 1976). Nachkommastellen werden noch mit Kommas angezeigt. Fehlende Werte sind als leere Zellen („systemdefiniert fehlend“) dargestellt (siehe Zeilen 33-35). Beides würde in Mplus zu Dateneinlesefehlern führen. Das SPSS-Programm muss nun zunächst wieder geschlossen werden, und es müssen die Ländereinstellungen geändert werden (siehe Abb. 1.2 und 1.3).

dass in der zu erzeugenden ASCII-Datei Nachkommastellen durch Punkte abgetrennt werden. Die Anzeige (Komma vs. Punkt) hängt in älteren SPSS-Versionen von den Ländereinstellungen ab. Ist unter dem Windows-Betriebssystem in der Systemsteuerung unter „Län-

dereinstellungen“ beispielsweise „Deutsch (Deutschland)“ eingestellt, so zeigt SPSS Kommas als Dezimaltrennzeichen an. Auch in der mit SPSS zu erzeugenden ASCII-Datei (siehe Abschnitt 3.3) würden sich dann Kommas befinden, was zwangsläufig zu Datenlesefehlern in Mplus führen würde. Dies können Sie in älteren SPSS-Versionen leicht überprüfen, indem Sie den *.sav Datensatz, den Sie in Mplus einlesen möchte, in SPSS öffnen (siehe Abb. 1.1). Wenn Sie im SPSS-Dateneditor Kommas sehen (wie in Abb. 1.1), so müssen Sie zunächst das SPSS-Programm wieder schließen und die Ländereinstellungen ändern, z.B. auf „Englisch (USA)“. Wenn Sie danach SPSS erneut öffnen, sehen Sie Punkte anstelle der Kommas als Dezimaltrennzeichen (siehe Abb. 1.4). Die entsprechenden Schritte werden in den folgenden Screenshots verdeutlicht. Das in neueren SPSS-Versionen mögliche einfachere Vorgehen mittels Syntax wird im Anschluss gezeigt.

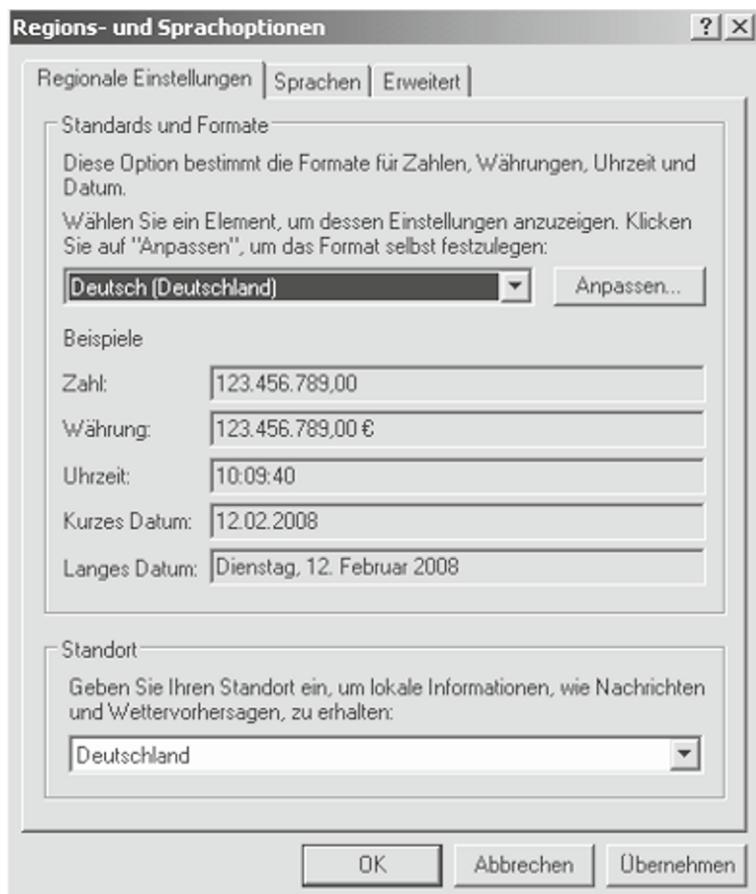


Abbildung 1.2. Im Betriebssystem Microsoft Windows gelangt man über die Option **Start → Einstellungen → Systemsteuerung → Regions- und Sprachoptionen** zu diesem Fenster. Man sieht, dass hier „Deutsch (Deutschland)“ eingestellt ist. Um Punkte als Dezimaltrennzeichen zu erhalten, muss diese Einstellung geändert werden, z.B. auf „Englisch (USA)“ (siehe Abb. 1.3).

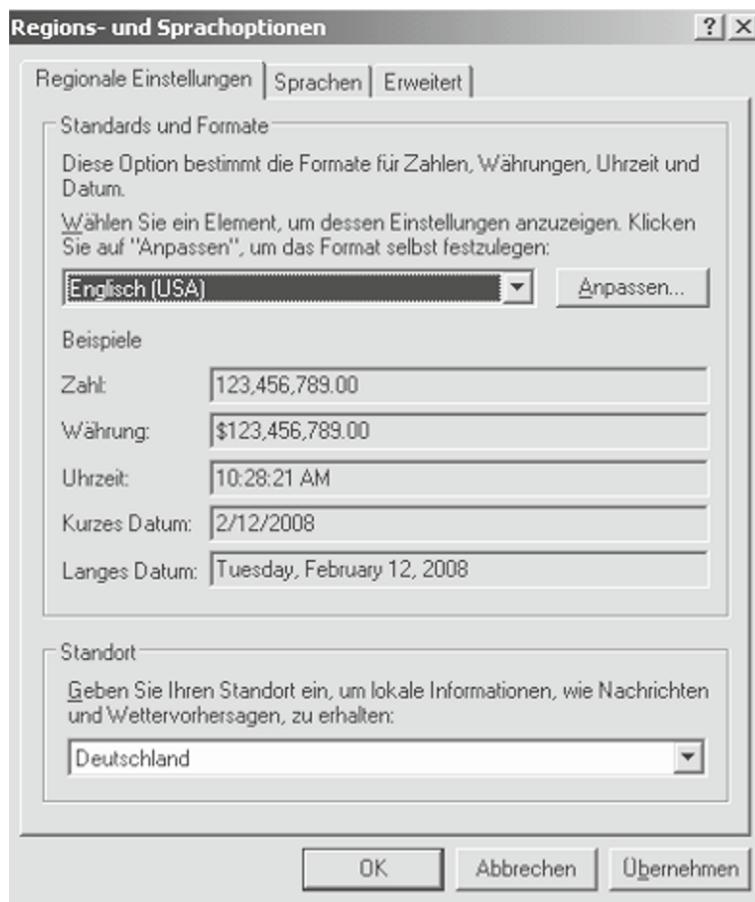


Abbildung 1.3. Geänderte regionale Einstellungen, damit Nachkommastellen mit Punkten abgetrennt werden. Danach muss der SPSS-Datensatz erneut geöffnet werden. Dezimaltrennzeichen werden nun in SPSS als Punkte dargestellt (siehe Abb. 1.4).

Dezimaltrennzeichen sind
nun Punkte!

	kft_v1	kft_v3	kft_q1	kft_q3	kft_n1	kft_n3	var	var	var
10	6.00	9.00	9.00	3.00	7.00	12.00			
11	13.00	16.00	12.00	6.00	7.00	10.00			
12	7.00	7.00	7.00	4.00	7.00	5.00			
13	10.00	5.00	7.00	2.00	5.00	4.00			
14	5.00	5.00	10.00	4.00	5.00	4.00			
15	8.00	9.00	7.00	2.00	6.00	5.00			
16	17.00	13.00	12.00	5.00	16.00	12.00			
17	9.00	9.00	10.00	6.00	9.00	14.00			
18	12.00	10.00	8.00	5.00	5.00	7.00			
19	16.00	5.00	12.00	12.00	21.00	11.00			
20	20.00	14.00	15.00	10.00	15.00	13.00			
21	20.00	10.00	14.00	13.00	22.00	11.00			
22	7.00	6.00	9.00	8.00	6.00	5.00			
23	2.00	4.00	10.00	5.00	6.00	5.00			
24	17.00	14.00	16.00	9.00	20.00	9.00			
25	12.00	6.00	12.00	6.00	8.00	9.00			
26	14.00	5.00	8.00	7.00	9.00	9.00			
27	11.00	5.00	7.00	4.00	6.00	11.00			
28	7.00	4.00	10.00	4.00	5.00	8.00			
29	14.00	6.00	9.00	8.00	9.00	7.00			
30	9.00	10.00	12.00	6.00	10.00	7.00			
31	20.00	4.00	16.00	9.00	5.00	4.00			
32	14.00	3.00	15.00	8.00	7.00	8.00			
33			
34			
35			
36	8.00	3.00	16.00	3.00	2.00	3.00			
37	11.00	15.00	13.00	6.00	4.00	12.00			
38	4.00	2.00	8.00	4.00	6.00	10.00			

Abbildung 1.4. Darstellung der Dezimaltrennzeichen als Punkte in SPSS (KFT-Datensatz aus Abb. 1.1). Fehlende Werte sind noch nicht kodiert.

In neueren SPSS-Versionen lässt sich das Problem der Dezimaltrennzeichen leichter lösen. Hier kann bei der Speicherung der ASCII-Datei für Mplus mittels eines SPSS-Syntaxbefehls direkt ein zusätzlicher Unterbefehl eingefügt werden, der dafür sorgt, dass in der zu speichernden ASCII-Datei Dezimalstellen mit Punkten abtrennt werden. Der entsprechende vollständige SPSS-Syntax-Befehl zur Abspeicherung der KFT-Daten in einer für Mplus lesbaren tabulator-getrennten Textdatei mit dem Dateinamen **KFT.dat** wird im Folgenden gezeigt:

```
SAVE TRANSLATE OUTFILE='C:\KFT.dat'
/TYPE=TAB
/textoptions decimal = dot
/MAP
/REPLACE
/CELLS=VALUES.
```

Die ASCII-Datei **KFT.dat** wird hier direkt auf der Festplatte „C“ abgespeichert. Wichtig ist hierbei der zusätzlich eingefügte Unterbefehl `/textoptions decimal = dot`, welcher dafür sorgt, dass Punkte als Dezimaltrennzeichen verwendet werden. Nach Abspeicherung der ASCII-Datei empfiehlt es sich, diese zunächst zu öffnen (in Windows z.B. mit dem Editor- oder WordPad-Programm), um die Anzeige der Dezimaltrennzeichen zu kontrollieren.

1.2 Kodierung fehlender Werte

Es empfiehlt sich, systemdefiniert fehlende Werte (in SPSS: leere Zellen mit Punkt, siehe Abb. 1.1 bzw. Abb. 1.4) vor der Abspeicherung der ASCII-Datei in SPSS in „benutzerdefiniert fehlend“ umzukodieren und diese nicht als leere Zellen abzuspeichern, da dies ebenfalls zu Datenlesefehlern in Mplus führen kann. „Benutzerdefiniert fehlend“ bedeutet, dass man einen Missing-Value-Code für jeden fehlenden Wert vergibt. Der Code kennzeichnet die fehlenden Werte eindeutig als fehlend. Es sollte hierbei am besten ein numerischer Code verwendet werden, also eine Zahl. Diese Zahl muss natürlich so gewählt werden, dass sie in den Daten nicht als gültiger Wert vorkommen kann. Beliebte numerische Missing-Value-Codes sind z.B. „-9“ „-99“ oder „999“. Wir wählen für unseren KFT-Beispieldatensatz die Zahl „-99“ als Missing-Value-Code für alle 6 Variablen. Die Zahl „-99“ kann bei keiner der KFT-Skalen als gültiger Wert vorkommen.

Keinesfalls sollte man Merkmalsträger mit einzelnen fehlenden Werten komplett aus den Daten herauslöschen. Mplus bietet nützliche Optionen zur Behandlung fehlender Werte (u.a. *Full Information Maximum Likelihood Schätzung* mit fehlenden Werten, siehe z.B. Enders, 2010; Geiser et al., in Vorbereitung; Reinecke, 2005). Zudem kann auch *listenweiser Fallausschluss* in Mplus als Option gewählt werden. Bei listenweisem Fallausschluss werden nur solche Fälle in die Analyse einbezogen, die auf allen verwendeten Variablen gültige Werte haben.

Systemdefiniert fehlende Werte können in SPSS leicht mit Hilfe der Option **Transformieren → Umkodieren in dieselben Variablen** in benutzerdefiniert fehlende Werte umkodiert werden. Das entsprechende Vorgehen zeigen die Abbildungen 1.5-1.8 für die KFT-Daten. Die entsprechenden Syntaxbefehle (RECODE-Befehl) zur Kodierung fehlender Werte können der SPSS-Syntaxdatei auf der Begleit-CD entnommen werden.

Achtung: Nach dem Umkodieren von „systemdefiniert fehlend“ in „benutzerdefiniert fehlend“ muss man SPSS unbedingt noch mitteilen, dass nun ein benutzerdefinierter Missing-Value-Code vergeben wurde. Dies erfolgt entweder manuell über die Variablenansicht (siehe Abb. 1.9-1.12) oder mit Hilfe des SPSS-Syntax Commands MISSING VALUES (siehe Begleit-CD). Es ist wichtig, dies zu beachten, da SPSS ansonsten den Missing-Value-Code

(hier: -99) wie einen realen Wert behandeln und falsche statistische Berechnungen ausgeben würde. Die folgenden Screenshots zeigen das korrekte Vorgehen exemplarisch anhand der KFT-Daten.

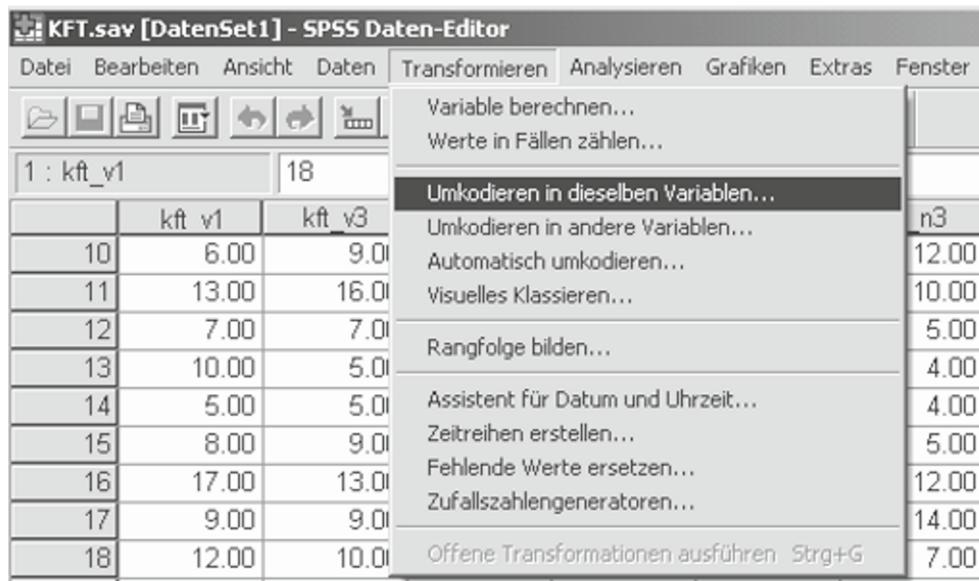


Abbildung 1.5: Option Transformieren → Umkodieren in dieselben Variablen in SPSS, hier benutzt zum Definieren eines Missing-Value-Codes.

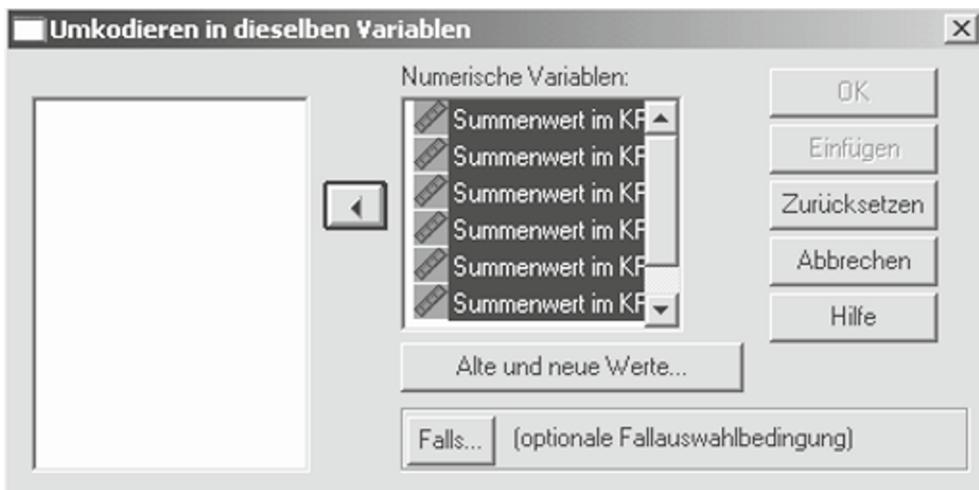


Abbildung 1.6: Alle Variablen werden zum Umkodieren in das rechte Fenster geklickt. Dann definiert man „Alte und neue Werte“ (siehe Abb. 1.7).

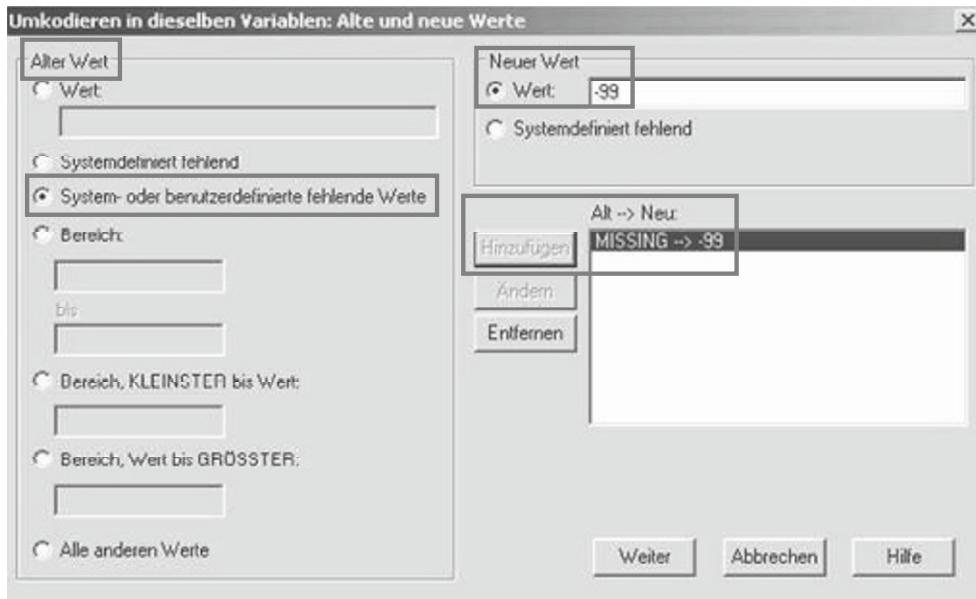


Abbildung 1.7: Option „Alte und neue Werte“ in der Umkodieroption von SPSS. Links werden alte (umzukodierende) Werte angegeben, rechts die neuen Werte (hier: Missing-Value-Code). Im gezeigten Beispiel wird die Zahl „-99“ als Missing-Value-Code für alle Werte vergeben, die bislang als systemdefiniert fehlend eingegeben waren oder einen anderen benutzerdefinierten Missing-Value-Code aufwiesen (z.B. -9). Dies dient der Vereinheitlichung zur Vermeidung von Fehlern. Nach dem Klicken auf Weiter und OK erhält man folgende veränderte Datenansicht (siehe Abb. 1.8).

*KFT.sav [DatenSet1] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Fenster Hilfe

Sichtbar: 6 von E

	kft_v1	kft_v3	kft_q1	kft_q3	kft_n1	kft_n3	Var	Var	Var
10	6.00	9.00	9.00	3.00	7.00	12.00			
11	13.00	16.00	12.00	6.00	7.00	10.00			
12	7.00	7.00	7.00	4.00	7.00	5.00			
13	10.00	5.00	7.00	2.00	5.00	4.00			
14	5.00	5.00	10.00	4.00	5.00	4.00			
15	8.00	9.00	7.00	2.00	6.00	5.00			
16	17.00	13.00	12.00	5.00	16.00	12.00			
17	9.00	9.00	10.00	6.00	9.00	14.00			
18	12.00	10.00	8.00	5.00	5.00	7.00			
19	16.00	5.00	12.00	12.00	21.00	11.00			
20	20.00	14.00	15.00	10.00	15.00	13.00			
21	20.00	10.00	14.00	13.00	22.00	11.00			
22	7.00	6.00	9.00	8.00	6.00	5.00			
23	2.00	4.00	10.00	5.00	6.00	5.00			
24	17.00	14.00	16.00	9.00	20.00	9.00			
25	12.00	6.00	12.00	6.00	8.00	9.00			
26	14.00	5.00	8.00	7.00	9.00	9.00			
27	11.00	5.00	7.00	4.00	6.00	11.00			
28	7.00	4.00	10.00	4.00	5.00	8.00			
29	14.00	6.00	9.00	8.00	9.00	7.00			
30	9.00	10.00	12.00	6.00	10.00	7.00			
31	20.00	4.00	16.00	9.00	5.00	4.00			
32	14.00	3.00	15.00	8.00	7.00	8.00			
33	-99.00	-99.00	-99.00	-99.00	-99.00	-99.00			
34	-99.00	-99.00	-99.00	-99.00	-99.00	-99.00			
35	-99.00	-99.00	-99.00	-99.00	-99.00	-99.00			
36	8.00	3.00	16.00	3.00	2.00	3.00			
37	11.00	15.00	13.00	6.00	4.00	12.00			
38	4.00	2.00	8.00	4.00	6.00	10.00			

Datenansicht Variablenansicht SPSS Prozessor ist bereit

Fehlende Werte sind nun mit -99 kodiert!

Abbildung 1.8: KFT-Beispieldatensatz. Alle fehlenden Werte sind nun bei allen 6 Variablen mit der Zahl „-99“ kodiert. Dass es sich bei den Werten „-99“ um fehlende Werte handelt, muss man SPSS nun unbedingt noch über die Variablenansicht oder mit Hilfe der Syntax mitteilen (siehe Abb. 1.9-1.12).

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte
1	kft_v1	Numerisch	8	2	Summenwert i	Kein	Kein
2	kft_v3	Numerisch	8	2	Summenwert i	Kein	Kein
3	kft_q1	Numerisch	8	2	Summenwert i	Kein	Kein
4	kft_q3	Numerisch	8	2	Summenwert i	Kein	Kein
5	kft_n1	Numerisch	8	2	Summenwert i	Kein	Kein
6	kft_n3	Numerisch	8	2	Summenwert i	Kein	Kein
7							

Abbildung 1.9: Die Zahl „-99“ ist in der Variablenansicht noch nicht als benutzerdefiniert fehlender Wert (Missing-Value-Code) eingetragen. Dies muss man unbedingt noch nachholen (siehe Abb. 1.10), da SPSS ansonsten die -99er Werte wie gültige Werte behandeln würde.

Um SPSS mitzuteilen, dass der Wert „-99“ nun bei allen Variablen fehlende Werte indiziert, kann man entweder den Syntax-Befehl MISSING VALUES (siehe SPSS-Syntax auf der Beigleit-CD) verwenden, oder folgendermaßen über das SPSS-Menü vorgehen. Dazu klickt man in der SPSS-Variablenansicht bei einer beliebigen Variable in der Spalte „fehlende Werte“ rechts auf die drei kleinen Punkte, wie in Abbildung 1.10 gezeigt.

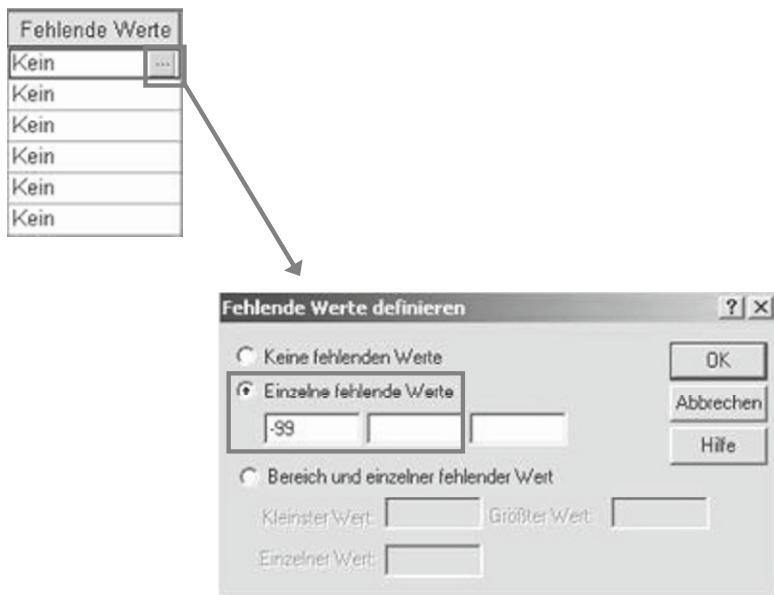


Abbildung 1.10: Der Wert -99 wird in SPSS als benutzerdefiniert fehlender Wert eingetragen. Diese Definition muss man praktischerweise nur für eine einzige Variable definieren. Dann kann man sie kopieren und für alle anderen Variablen ebenfalls einfügen (siehe Abb. 1.11).



Abbildung 1.11: Kopieren und Einfügen der Missing-Value-Definition in der SPSS-Variablenansicht. Abbildung 1.12 zeigt die korrekt definierten Variablen.

	Name	Typ	Spaltenformat	Dezimalstellen	Variablenlabel	Wertelabels	Fehlende Werte
1	kft_v1	Numerisch	8	2	Summenwert i	Kein	-99.00
2	kft_y3	Numerisch	8	2	Summenwert i	Kein	-99.00
3	kft_q1	Numerisch	8	2	Summenwert i	Kein	-99.00
4	kft_q3	Numerisch	8	2	Summenwert i	Kein	-99.00
5	kft_n1	Numerisch	8	2	Summenwert i	Kein	-99.00
6	kft_n3	Numerisch	8	2	Summenwert i	Kein	-99.00
7							

Abbildung 1.12: Korrekt eingetragener Missing Value Code (-99) für alle 6 Variablen des KFT-Datensatzes in der SPSS-Variablenansicht.

1.3 Abspeichern eines ASCII-Datensatzes für Mplus

Nachdem Dezimaltrennzeichen als Punkte im Datensatz eingetragen sind und man fehlende Werte mit einem numerischen Missing-Value-Code (hier: -99) gekennzeichnet hat, kann man die Daten in dem für Mplus benötigten ASCII-Format abspeichern. Dies geschieht in der SPSS-Datenansicht über **Datei → Speichern unter** (siehe Abb. 1.13-1.15) oder mit Hilfe des Syntaxbefehls **SAVE TRANSLATE** (siehe Seite 16).

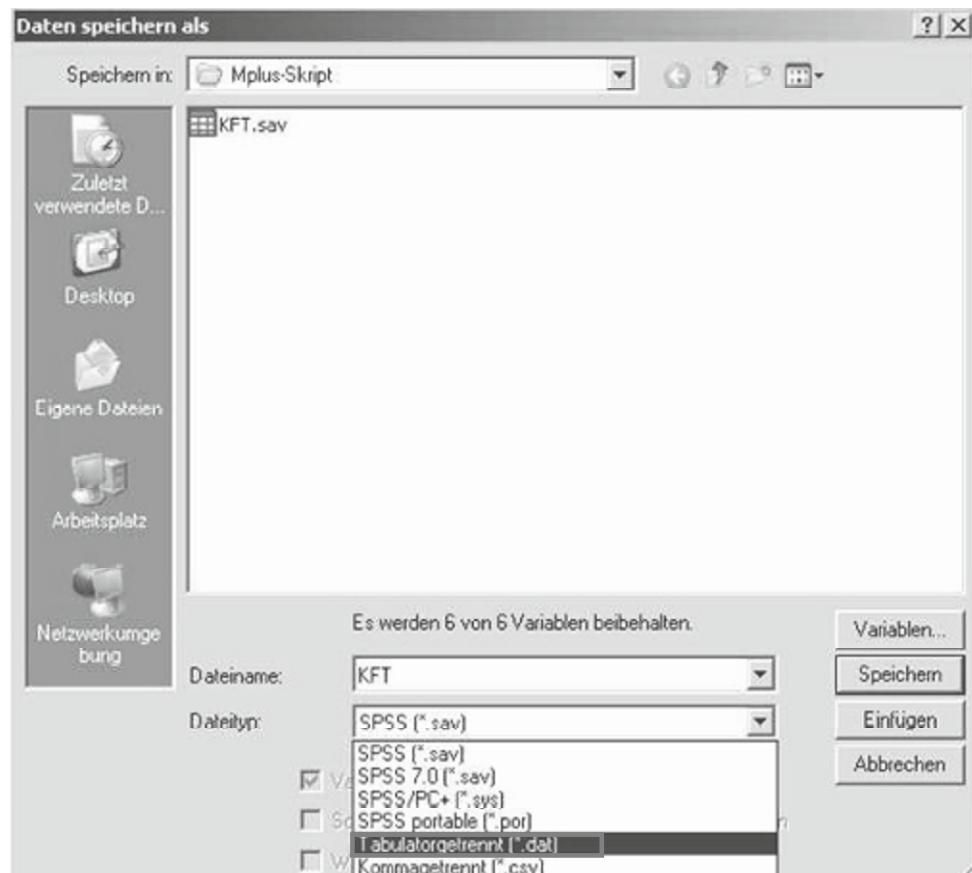


Abbildung 1.13: Abspeichern des KFT-Datensatzes im ASCII-Format. Der Autor empfiehlt hierbei die Verwendung der Variante „Tabulatortrennt (*.dat)“.

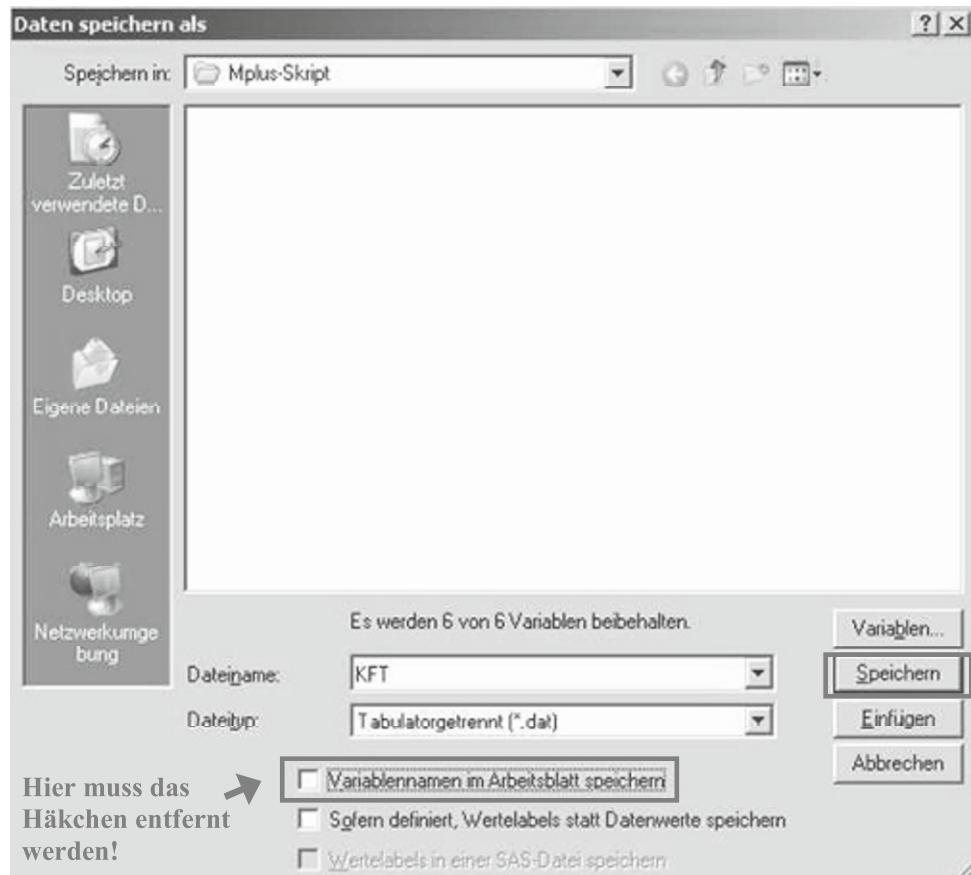


Abbildung 1.14: Abspeichern des KFT-Datensatzes im tabulator-getrennten ASCII-Format. Wichtig: Die Option **Variablennamen im Arbeitsblatt speichern** muss *deaktiviert* werden (Häkchen entfernen)! Dann auf **Speichern** klicken. Abbildung 1.15 zeigt einen Ausschnitt aus dem neu abgespeicherten ASCII-Datensatz mit dem Namen „KFT.dat“. Dieser lässt sich z.B. mit einem einfachen Textverarbeitungsprogramm (Editor, WordPad, Microsoft-Word etc.) öffnen.

18	13	14	7	9	5
16	7	4	2	12	10
17	13	8	8	9	10
20	17	13	9	20	12
13	6	19	9	10	10
6	5	8	6	3	5
16	7	19	9	9	10
12	5	10	4	13	9
7	10	9	7	9	9
6	9	9	3	7	12
13	16	12	6	7	10
7	7	7	4	7	5
10	5	7	2	5	4

Abbildung 1.15: Ausschnitt aus dem neu abgespeicherten ASCII-Datensatz (Variablenpalten sind mit Tabulatorzeichen getrennt) mit dem Namen „KFT.dat“. Dieser wurde hier mit dem Windows-Programm „Editor“ geöffnet (In Windows: **Start → Programme → Zubehör**). Man beachte, dass Nachkommastellen hier nicht berücksichtigt wurden, da es sich bei allen Werten um ganzzahlige Scores handelt. Bei nicht-ganzzahligen Werten würden die Nachkommastellen nun mit Punkten angezeigt werden (hat man Kommas als Dezimaltrennzeichen, so würde das Einlesen dieser Daten in Mplus Fehlermeldungen verursachen, siehe Abschnitt 1.2).

2 Einlesen von Daten in Mplus

Mplus kann unterschiedliche Arten von Daten einlesen und verarbeiten. Die beiden wichtigsten Varianten sollen hier besprochen werden. Die erste Variante ist in der Praxis wohl die Häufigste. Hierbei werden Rohdaten verwendet, d.h. sogenannte *Individualdaten (individual data)*. In Individualdatensätzen hat jede Person auf jeder Variable einen Wert (oder ggf. einen fehlenden Wert), so wie dies z.B. im KFT-Datensatz (siehe Abb. 1.15) der Fall ist. Die zweite Variante bezieht sich auf das Einlesen sogenannter *Summary-Daten*. Dabei handelt es sich z.B. um eine Kovarianz- oder Korrelationsmatrix (und ggf. die Mittelwerte und Standardabweichungen der Variablen). Die Möglichkeit, mit Summary-Daten arbeiten zu können, ist beispielsweise dann nützlich, wenn man Daten aus einem Forschungsbericht (z.B. einem Fachzeitschriftenartikel) reanalysieren möchte. In vielen Publikationen werden Kovarianz- oder Korrelationsmatrizen, die die Grundlage für die berichteten Analysen (z.B. Pfad- oder Strukturgleichungsanalysen) darstellen, mit angegeben (wohingegen die Rohdaten nur selten verfügbar gemacht werden). Für viele Arten von linearen Strukturgleichungsmodellen ist es nicht unbedingt erforderlich, Individualdaten zu analysieren (obwohl Individualdaten in vielerlei Hinsicht vorteilhaft sind, z.B. wenn fehlende Werte oder besondere Schätzverfahren berücksichtigt werden sollen). Als Input genügt bei vielen Modellen die Kovarianzmatrix, ggf. ergänzt durch die Mittelwerte der zu analysierenden Variablen. Es wird zunächst das Einlesen von Individualdaten gezeigt, und zwar anhand des in Abschnitt 1.3 abgespeicherten ASCII-Datensatzes mit den KFT-Daten.

2.1 Einlesen von Individualdaten (Rohdaten)

Das Einlesen von Rohdaten verläuft erfahrungsgemäß problemlos, wenn man die in Kapitel 1 gezeigten Schritte korrekt ausgeführt hat. Trotzdem ist es unabdingbar, dass man sich in Mplus zunächst Deskriptivstatistiken (Mittelwerte, Varianzen, Kovarianzen und Korrelationen) der beobachteten Variablen ausgeben lässt und diese mit den entsprechenden Werten in SPSS vergleicht, bevor man erste weiterführende statistische Analysen durchführt. Der Vergleich von Deskriptivstatistiken in Mplus und SPSS dient dazu, zu überprüfen, ob die Daten von Mplus korrekt gelesen werden. Diesen Punkt kann man gar nicht genug betonen, da er allzu oft übersprungen und dadurch nicht (oder zu spät!) bemerkt wird, dass Mplus die verwendeten Daten nicht korrekt verarbeitet und deshalb inkorrekte Schätzergebnisse produziert. Aus diesem Grund sollte man sich die Zeit nehmen, in Mplus zunächst eine sogenannte *basic-Analyse* durchzuführen, um das korrekte Einlesen der Daten sicherzustellen, bevor man mit der ersten tatsächlichen Analyse beginnt.

2.1.1 Aufbau der basalen Mplus-Syntax und basic-Analyse

Mplus ist ein syntaxbasiertes Programm, d.h. alle Funktionen werden über eine Befehlsprache angefordert und sind nicht wie z.B. in SPSS über das Menü erhältlich. Zur Erstellung der basalen Mplus-Syntax steht jedoch eine menügesteuerte Funktion zur Verfügung. Dabei handelt es sich um den sogenannten *Mplus Language Generator* (im Mplus-Menü **Mplus → Language Generator**). Diese menügesteuerte Option ermöglicht es, die wichtigsten Grundbefehle vom Programm automatisch erstellen zu lassen. Die Befehle zur Spezifikation eines bestimmten Modells (model-Befehl) muss man jedoch trotzdem noch selbst einfügen. Auf die Darstellung der Variante der Syntaxgenerierung über den Mplus Language Generator wird hier verzichtet, da Benutzer in aller Regel nach Erlernen der Grundbefehle nicht mehr auf diese Funktion zurückgreifen.

Im Folgenden wird ein sinnvolles Vorgehen zum Dateneinlesen und zur Datenkontrolle mit der **basic**-Option anhand des Beispieldatensatzes KFT.dat demonstriert. Zunächst muss man den Mplus-Editor öffnen (in Windows: **Start → Programme → Mplus → Mplus Editor**). Es öffnet sich ein leeres Editor-Fenster. Man kann dann entweder direkt die benötigten Syntaxbefehle eingeben oder das Vorgehen über den Mplus Language Generator wählen. Die Minimalvariante benötigter Syntaxbefehle wird in Abbildung 2.1 gezeigt.

Sobald man begonnen hat, eine Mplus-Syntax zu erstellen, sollte man diese als Mplus-Input-Datei abspeichern (siehe Abb. 2.1). Mplus-Input-Dateien erhalten die Dateiendung *.inp. Auch regelmäßiges Zwischenspeichern neu eingefügter oder modifizierter Befehle (mittels Klicken auf das Diskettensymbol im Mplus-Menü oder Drücken der Tastenkombination Strg + S) sollte nicht vergessen werden, damit bei einem etwaigen Computerabsturz keine Arbeit verloren geht.

Der **title**-Befehl dient dazu, die Analyse zu benennen bzw. ausführlich zu kommentieren. Dieser Befehl wird nicht zwingend benötigt, allerdings ist es zur Dokumentation der eigenen Arbeit ratsam, stets einen ausführlichen Titel für die zu spezifizierende Analyse zu vergeben. Optional ist auch das Einfügen von Kommentarzeilen. Kommentarzeilen müssen mit einem Ausrufezeichen (!) beginnen (und brauchen nicht mit einem besonderen Zeichen zu enden). Mplus ignoriert alles, was in einer Kommentarzeile erscheint. Jede neue Kommentarzeile muss erneut mit einem „!“-Zeichen beginnen. Kommentarzeilen sind in Mplus praktischerweise in grüner Schriftfarbe abgehoben. Bei Kommentarzeilen ist die Länge übrigens nicht kritisch, da es sich nicht um relevante Befehle handelt. Allerdings schneidet Mplus zu lange Kommentarzeilen nach 80 bzw. 90 Zeichen ab, so dass die Kommentare u.U. unvollständig im Output erscheinen.

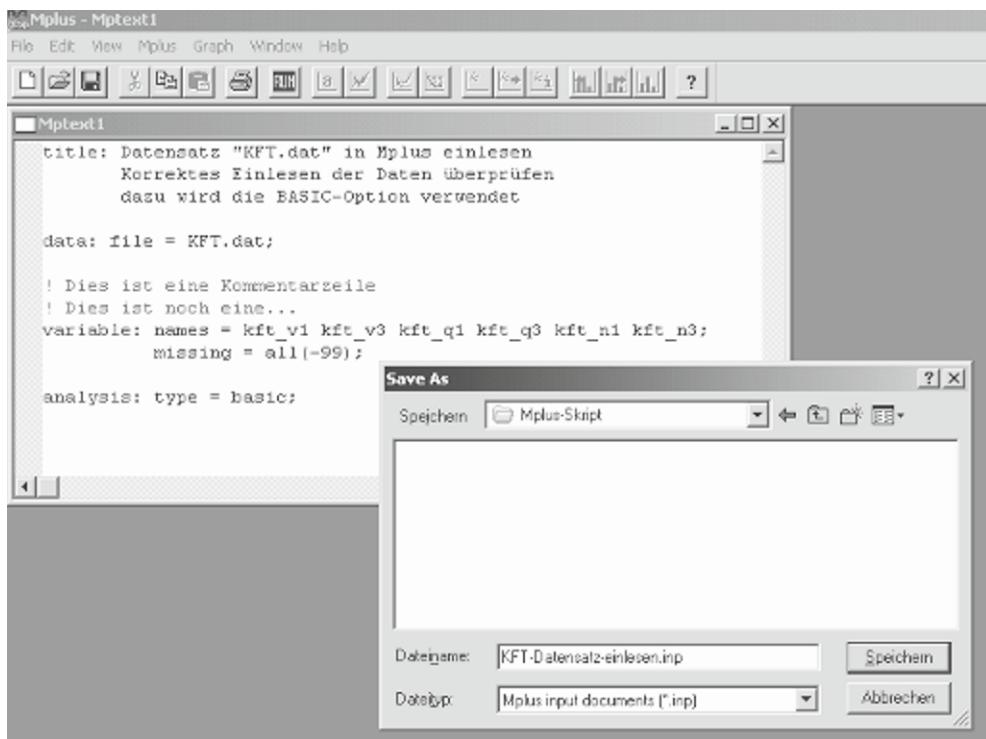


Abbildung 2.1: Syntax zur Überprüfung des korrekten Dateneinlesens des Datensatzes KFT.dat mittels basic-Analyse und Abspeichern des Mplus-Input-Files als *.inp-Datei. Hierbei wird der Input im selben Ordner abgespeichert, in welchem sich auch der Datensatz KFT.dat befindet. Deshalb ist eine genaue Spezifikation eines Pfades zu den Daten unter data: file = nicht erforderlich.

Der data-Befehl dient u.a. dazu, Mplus den Namen, Speicherort sowie die Art des zu verwendenden Datensatzes mitzuteilen. Wenn man den Datensatz auf dem Computer im gleichen Verzeichnis (Ordner) abspeichert wie die Mplus-Input-Datei, so braucht man den genauen Pfad zum Speicherort des Datensatzes nicht anzugeben (vgl. Abb. 2.1).

Der variable-Befehl dient u.a. zur Definition der im Datensatz enthaltenen und der zu analysierenden Variablen sowie zur Definition des verwendeten Missing-Value-Codes (vgl. dazu Abschnitt 1.2). Unter names = werden die Variablennamen vergeben. Wichtig ist hierbei die korrekte Reihenfolge der Variablennamen in der Liste hinter names =. Die Reihenfolge muss unbedingt der Reihenfolge der Variablen im einzulesenden ASCII-Datensatz entsprechen! Außerdem ist wichtig zu beachten, dass Variablennamen nicht länger als 8 Zeichen sein dürfen. Sehr bequem kann man die Variablennamen direkt aus SPSS mit Hilfe der SPSS-Menü-Option **Extras → Variablen** importieren. Diese Möglichkeit wird anhand von Abbildung 2.2 und 2.3 illustriert.

Ein weiterer Vorteil dieses Vorgehens ist, dass man die gleichen Variablennamen in beiden Programmen verwendet (Voraussetzung hierfür ist natürlich, dass man in SPSS auch

keine Variablennamen mit mehr als 8 Zeichen verwendet hat). Wichtig ist, hierbei darauf zu achten, dass SPSS die Variablennamen in der korrekten Reihenfolge in die Syntax einfügt, was leider nicht in jeder SPSS-Version der Fall ist.

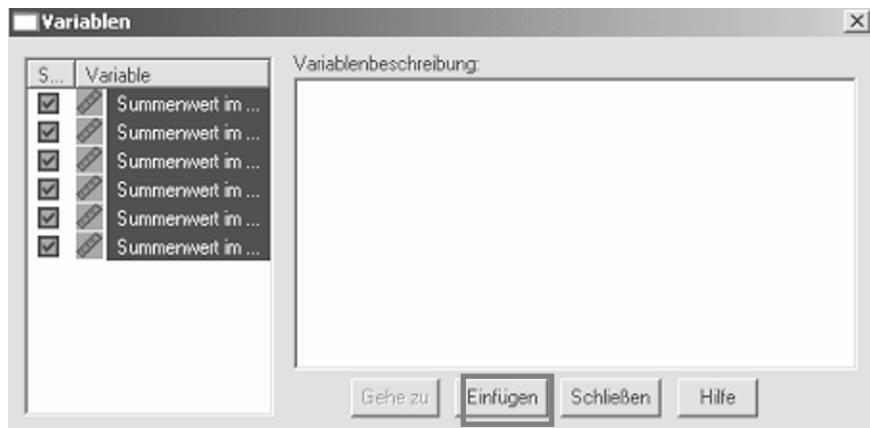


Abbildung 2.2. Exportieren der Variablennamen aus SPSS. Links werden alle Variablen markiert. Durch das Klicken auf **Einfügen** werden die Variablennamen in ein Syntaxfenster eingefügt (siehe Abb. 2.3).

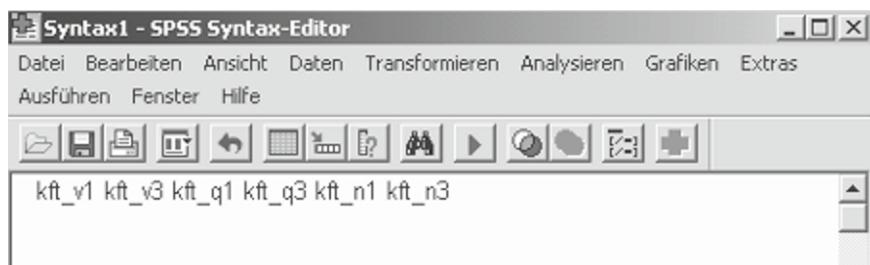


Abbildung 2.3. Exportieren der Variablennamen aus SPSS. Durch das Klicken auf **Einfügen** wurden die Variablennamen automatisch in das SPSS-Syntaxfenster eingefügt. Sie können nun aus dem Syntaxfenster herauskopiert und in Mplus-Editor im variable-Befehl hinter names = eingefügt werden. Hierbei muss man unbedingt auf die korrekte Reihenfolge der Variablennamen achten.

Weiterhin sollte beachtet werden, dass jeder Unterbefehl (mit Ausnahme des Titels) in Mplus mit einem Semikolon enden muss – das Fehlen eines Semikolons nimmt vermutlich in der Hitparade möglicher Fehlerquellen in Mplus den ersten Platz ein! Mit Hilfe des missing-Befehls teilt man Mplus mit, wie fehlende Werte kodiert sind („Missing-Value-Code“). Im vorliegenden Beispiel haben wir bei allen Variablen fehlende Werte mit -99 kodiert (siehe Abschnitt 1.2), also schreiben wir in Mplus unter variable: missing = all(-99);.

Mit Hilfe von analysis: type = basic; fordern wir Deskriptivstatistiken für alle Variablen an, die wir dann mit den entsprechenden Werten in SPSS vergleichen können.

Durch das Klicken auf RUN im Mplus-Menü (oder drücken der Tastenkombination **Alt + R**) wird die basic-Analyse ausgeführt. Box 2.1 gibt einen Überblick über einige grundlegende Regeln zur Mplus-Syntax. Im Anschluss werden die Ergebnisse der basic-Analyse besprochen.

Box 2.1: Einige wichtige Mplus-Syntaxregeln

- In der Mplus-Syntax wird nicht zwischen Groß- und Kleinschreibung unterschieden.
- Auch die Reihenfolge, mit welcher die einzelnen Befehle in der Syntax aufgelistet werden, ist weitgehend beliebig.
- Jede Befehlszeile muss mit einem Semikolon (;) abgeschlossen werden.
- Eine einzelne Befehlszeile darf nicht länger als 90 Zeichen sein (in älteren Mplus-Versionen nur 80 Zeichen!). Lange Befehlszeilen können zum Beispiel auftreten, wenn man viele Variablen einliest oder einen langen Pfad zum Datensatz angeben möchte. Die Folge ist, dass Mplus die Zeile nach 80 bzw. 90 Zeichen „abschneidet“, d.h. alle folgenden Zeichen ignoriert (was u.U. zu drastischen Fehlern führen kann!). Es erfolgt eine entsprechende Fehlermeldung im Mplus-Output, die auf dieses Problem hinweist. Man sollte sie keinesfalls ignorieren! Ein einfaches Mittel zum Beheben dieses Fehlers ist das simple Umbrechen der betreffenden zu langen Zeile mittels RETURN-Taste.
- Variablennamen dürfen nicht aus mehr als 8 Zeichen bestehen.
- Jede Kommentarzeile muss mit einem Ausrufezeichen (!) beginnen. Sie wird im Mplus-Input in grüner Farbe dargestellt. Kommentarzeilen müssen nicht mit einem besonderen Zeichen enden, jede neue Kommentarzeile muss jedoch wieder mit einem Ausrufezeichen beginnen.

2.1.2 Mplus-Output zur basic-Analyse

Nach dem Ausführen der Inputdatei zur basic-Analyse erscheint ein neues Fenster (sog. Outputfenster) mit dem entsprechenden Output. Der Output wird von Mplus automatisch als *.out-Datei mit demselben Dateinamen wie der Inputfile in dem Verzeichnis abgespeichert, in welchem sich der zugehörige Inputfile (*.inp) befindet. Der Output beginnt wie folgt:

```
Mplus VERSION 5
MUTHEN & MUTHEN
02/12/2008 3:40 PM

INPUT INSTRUCTIONS

title: Datensatz "KFT.dat" in Mplus einlesen
       Korrektes Einlesen der Daten überprüfen
       dazu wird die BASIC-Option verwendet

data: file = KFT.dat;

! Dies ist eine Kommentarzeile
! Dies ist noch eine...
variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
           missing = all(-99);

analysis: type = basic;
```

Zunächst werden die Mplus-Input-Befehle reproduziert. Dies ist nützlich, da man so noch einmal kontrollieren kann, ob die Inputspezifikation korrekt war und diese auch auf einem möglichen Ausdruck des Outputs noch einmal dokumentiert ist. Dann erscheint eine Warnmeldung, da es im Datensatz **KFT.dat** insgesamt 131 Schüler gibt, die auf allen 6 Variablen fehlende Werte aufweisen:

```
*** WARNING
Data set contains cases with missing on all variables.
These cases were not included in the analysis.
Number of cases with missing on all variables: 131
1 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS
```

Diese Meldung erscheint, da Mplus 5 standardmäßig (als „Default“) das Full Information Maximum Likelihood (FIML) Schätzverfahren zur Einbeziehung fehlender Werte verwendet (siehe z.B. Enders, 2010; Reinecke, 2005). Dieses Verfahren kann bei Personen, die auf *allen* verwendeten Variablen fehlende Werte aufweisen, nicht zum Einsatz gebracht werden, da für diese Personen in Bezug auf die zu analysierenden Variablen keinerlei Informationen zur Verfügung stehen.

Als nächstes erscheinen weitere technische Informationen zur Analyse, u.a. zur Stichprobengröße, zu den Variablen und zu den Daten allgemein:

```
Datensatz "KFT.dat" in Mplus einlesen
Korrektes Einlesen der Daten überprüfen
dazu wird die BASIC-Option verwendet
```

SUMMARY OF ANALYSIS

Number of groups	1
Number of observations	456

```

Number of dependent variables                                6
Number of independent variables                             0
Number of continuous latent variables                      0

Observed dependent variables
  Continuous
    KFT_V1      KFT_V3      KFT_Q1      KFT_Q3      KFT_N1      KFT_N3

Estimator                                                 ML
Information matrix                                         OBSERVED
Maximum number of iterations                            1000
Convergence criterion                                 0.5000D-04
Maximum number of steepest descent iterations          20
Maximum number of iterations for H1                  2000
Convergence criterion for H1                         0.1000D-03

Input data file(s)
  KFT.dat

Input data format  FREE

```

Diese sind nützlich, um noch einmal zu kontrollieren, dass tatsächlich der korrekte Datensatz sowie die intendierten Variablen verwendet wurden.

Nachfolgend erhält man nützliche Informationen zu den in den Daten beobachteten Mustern fehlender Werte („missing data patterns“).

```

SUMMARY OF DATA

  Number of missing data patterns                        2

SUMMARY OF MISSING DATA PATTERNS

  MISSING DATA PATTERNS (x = not missing)
    1   2
  KFT_V1     x   x
  KFT_V3     x   x
  KFT_Q1     x   x
  KFT_Q3     x   x
  KFT_N1     x   x
  KFT_N3     x

  MISSING DATA PATTERN FREQUENCIES

  Pattern  Frequency      Pattern  Frequency
        1         455           2           1

```

Weiter oben unter number of observations sehen wir, dass insgesamt 456 Schüler einen oder mehrere Werte beisteuern (hier sind diejenigen Fälle bereits ausgenommen, die fehlende Werte auf allen 6 Variablen aufweisen). Unter missing data patterns werden

alle beobachteten Ausfallmuster in einer eigenen Spalte dargestellt (mit Ausnahme des Musters, bei dem alle Werte fehlen). Die Spalte 1 (missing data pattern Nummer 1) enthält nur Kreuze, was anzeigt, dass es sich um dasjenige Muster handelt, bei welchem *keine* fehlenden Werte auftreten. In der zweiten Spalte (missing data pattern Nummer 2) sehen wir Kreuze bei allen Variablen, außer bei der letzten (**kft_n3**). Es handelt sich somit um das Ausfallmuster, bei welchem für alle Variablen Werte vorliegen außer für die letzte.

Im Abschnitt missing data pattern frequencies zeigt sich, dass von den 456 Schülern fast alle vollständige Daten (Werte auf allen 6 Variablen) liefern: 455 Schüler haben das missing data pattern Nummer 1 (Frequency = 455). Es gibt nur einen Schüler mit dem missing data pattern Nummer 2 (Frequency = 1), also einem fehlenden Wert auf der Variable **kft_n3**.

Die covariance coverage gibt an, wie viel Prozent der Fälle Werte zu einer bestimmten Varianz oder Kovarianz beisteuern. In unserem Fall sind dies für alle Varianzen und Kovarianzen 100% der Fälle, mit Ausnahme der Varianz und der Kovarianzen, die mit der Variable **kft_n3** assoziiert sind (dort nur 99.8%, da hier eine Person einen fehlenden Wert aufweist). Der minimal akzeptable Wert für die covariance coverage ist in Mplus 10% als Voreinstellung („Minimum covariance coverage value 0.100“).

COVARIANCE COVERAGE OF DATA					
Minimum covariance coverage value 0.100					
PROPORTION OF DATA PRESENT					
	Covariance Coverage				
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	1.000				
KFT_V3	1.000	1.000			
KFT_Q1	1.000	1.000	1.000		
KFT_Q3	1.000	1.000	1.000	1.000	
KFT_N1	1.000	1.000	1.000	1.000	1.000
KFT_N3	0.998	0.998	0.998	0.998	0.998
Covariance Coverage					
KFT_N3					
KFT_N3		0.998			

Achtung: Die covariance coverage gibt nicht die tatsächlichen Werte der Kovarianzen zwischen den Variablen an, sondern gibt nur Auskunft über die Datengrundlage, die zu deren Berechnung verfügbar ist. Die tatsächlich geschätzten Kovarianzen erhalten wir weiter unten unter der Rubrik estimated sample statistics.

Als nächstes erhalten wir Deskriptivstatistiken für die sechs beobachteten KFT-Variablen (d.h. die geschätzten Mittelwerte, die Kovarianzmatrix und die Matrix der Pearson Produkt-Moment-Korrelationen der Variablen unter der Rubrik estimated sample

`statistics`). Diese Statistiken wurden mit Hilfe des FIML-Verfahrens geschätzt. Dadurch sind die Statistiken nicht eins-zu-eins mit in SPSS berechneten Statistiken vergleichbar. Das liegt daran, dass die SPSS-Statistiken entweder auf paarweisem Fallausschluss (pairwise deletion) oder auf listenweisem Fallausschluss (listwise deletion; Ausschluss aller Personen, die auf wenigstens einer der sechs Variablen einen fehlenden Wert aufweisen) beruhen und nicht auf FIML-Schätzungen.

Um einen exakten Abgleich der Mplus- und SPSS-Ergebnisse vornehmen zu können, kann man in Mplus jedoch ebenfalls listenweisem Fallausschluss (anstelle der Voreinstellung FIML) anfordern. Listenweisem Fallausschluss erreicht man in Mplus, indem man im Input-Fenster unter dem `data`-Befehl den Unterbefehl `listwise = on` einfügt. Dies bewirkt, dass die FIML-Prozedur deaktiviert wird und alle Fälle, die auf wenigstens einer der verwendeten Variablen einen fehlenden Wert aufweisen, aus der Analyse ausgeschlossen werden (FIML ist erst seit der Mplus-Version 5 die Voreinstellung. Bei den Vorgängerversionen war listenweiser Fallausschluss die Voreinstellung und das FIML-Verfahren musste gesondert angefordert werden).

Die erweiterte Mplus-Syntax mit listenweisem Fallausschluss ist in Abbildung 2.4 dargestellt. Die resultierenden Deskriptivstatistiken sind darunter wiedergegeben. Zum Vergleich sind direkt unter dem Mplus-Output in Abbildung 2.5 die korrespondierenden SPSS-Statistiken gezeigt, die auf listenweisem Fallausschluss beruhen und mit Hilfe der SPSS-Option **Analysieren → Skalierung → Reliabilitätsanalyse → Statistik → Deskriptive Statistiken für Item/Zwischen Items Korrelationen/Kovarianzen** erzeugt wurden. Sie stimmen mit den Mplus-Ergebnissen unter `listwise deletion` (auf die dritte Nachkommastelle gerundet) überein, was uns zeigt, dass der Datensatz **KFT.dat** in Mplus korrekt gelesen wird. Auch kommt SPSS auf dieselbe Fallzahl ($N = 455$ listenweise gültige Fälle). Nach dieser Überprüfung können wir mit der Spezifikation des ersten Modells in Mplus beginnen (siehe Kapitel 3).

RESULTS FOR BASIC ANALYSIS [FIML-Schätzungen]					
ESTIMATED SAMPLE STATISTICS					
Means					
KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1	KFT_N3
11.904	8.978	12.377	7.730	11.088	8.277
Covariances					
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	21.592				
KFT_V3	10.761	17.938			
KFT_Q1	6.010	5.267	11.235		
KFT_Q3	4.645	4.406	4.027	6.754	
KFT_N1	11.072	11.644	6.870	6.111	29.826
KFT_N3	6.347	7.190	4.415	3.986	9.298
Covariances					
		KFT_N3			
KFT_N3		11.782			
Correlations					
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	1.000				
KFT_V3	0.547	1.000			
KFT_Q1	0.386	0.371	1.000		
KFT_Q3	0.385	0.400	0.462	1.000	
KFT_N1	0.436	0.503	0.375	0.431	1.000
KFT_N3	0.398	0.495	0.384	0.447	0.496
Correlations					
		KFT_N3			
KFT_N3		1.000			
MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS -7152.289					
Beginning Time: 15:40:25					
Ending Time: 15:40:25					
Elapsed Time: 00:00:00					
MUTHEN & MUTHEN					
3463 Stoner Ave.					
Los Angeles, CA 90066					
Tel: (310) 391-9971					
Fax: (310) 391-8971					
Web: www.StatModel.com					
Support: Support@StatModel.com					
Copyright (c) 1998-2007 Muthen & Muthen					

The screenshot shows the Mplus software window with the title bar "Mplus - [KFT-Datensatz-einlesen-listwise-deletion.inp]". The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. Below the menu is a toolbar with various icons. The main area contains the Mplus input code:

```

title: Datensatz "KFT.dat" in Mplus einlesen
      Korrektes Einlesen der Daten überprüfen
      dazu wird die BASIC-Option verwendet
      Abschalten der FIML-Prozedur, listenweiser Fallausschluss

data: file = KFT.dat;
      listwise = on; ! FIML-Prozedur wird abgeschaltet
                  ! Stattdessen listenweiser Fallausschluss

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;
          missing = all(-99);

analysis: type = basic;

```

The status bar at the bottom left says "Ready" and the bottom right says "Ln 4, Col 65".

Abbildung 2.4: Geänderter Mplus-basic-Input zum Abschalten der FIML-Prozedur mit Hilfe des Unterbefehls listwise = on;

SAMPLE STATISTICS [Berechnet gemäß listenweisem Fallausschluss]					
Means					
KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1	KFT_N3
11.899	8.974	12.387	7.732	11.112	8.281

Covariances					
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	21.677				
KFT_V3	10.799	18.008			
KFT_Q1	6.057	5.310	11.242		
KFT_Q3	4.669	4.429	4.038	6.783	
KFT_N1	11.172	11.745	6.794	6.120	29.686
KFT_N3	6.383	7.230	4.417	4.001	9.294

Covariances	
	KFT_N3
KFT_N3	11.811

Correlations					
	KFT_V1	KFT_V3	KFT_Q1	KFT_Q3	KFT_N1
KFT_V1	1.000				
KFT_V3	0.547	1.000			
KFT_Q1	0.388	0.373	1.000		
KFT_Q3	0.385	0.401	0.462	1.000	
KFT_N1	0.440	0.508	0.372	0.431	1.000
KFT_N3	0.399	0.496	0.383	0.447	0.496

Correlations	
	KFT_N3
KFT_N3	1.000

Itemstatistiken

	Mittelwert	Std.-Abweichung	Anzahl
Summenwert im KFT-Subtest V1	11.8989	4.65585	455
Summenwert im KFT-Subtest V3	8.9736	4.24360	455
Summenwert im KFT-Subtest Q1	12.3868	3.35293	455
Summenwert im KFT-Subtest Q3	7.7319	2.60434	455
Summenwert im KFT-Subtest N1	11.1121	5.44845	455
Summenwert im KFT-Subtest N3	8.2813	3.43665	455

Inter-Item-Kovarianzmatrix

	Summenwert im KFT-Subtest V1	Summenwert im KFT-Subtest V3	Summenwert im KFT-Subtest Q1	Summenwert im KFT-Subtest Q3	Summenwert im KFT-Subtest N1	Summenwert im KFT-Subtest N3
Summenwert im KFT-Subtest V1	21.677	10.799	6.057	4.669	11.172	6.383
Summenwert im KFT-Subtest V3	10.799	18.008	5.310	4.429	11.745	7.230
Summenwert im KFT-Subtest Q1	6.057	5.310	11.242	4.038	6.794	4.417
Summenwert im KFT-Subtest Q3	4.669	4.429	4.038	6.783	6.120	4.001
Summenwert im KFT-Subtest N1	11.172	11.745	6.794	6.120	29.686	9.294
Summenwert im KFT-Subtest N3	6.383	7.230	4.417	4.001	9.294	11.811

Inter-Item-Korrelationsmatrix

	Summenwert im KFT-Subtest V1	Summenwert im KFT-Subtest V3	Summenwert im KFT-Subtest Q1	Summenwert im KFT-Subtest Q3	Summenwert im KFT-Subtest N1	Summenwert im KFT-Subtest N3
Summenwert im KFT-Subtest V1	1.000	.547	.388	.385	.440	.399
Summenwert im KFT-Subtest V3	.547	1.000	.373	.401	.508	.496
Summenwert im KFT-Subtest Q1	.388	.373	1.000	.462	.372	.383
Summenwert im KFT-Subtest Q3	.385	.401	.462	1.000	.431	.447
Summenwert im KFT-Subtest N1	.440	.508	.372	.431	1.000	.496
Summenwert im KFT-Subtest N3	.399	.496	.383	.447	.496	1.000

Abbildung 2.5: Deskriptivstatistiken für die 6 KFT-Skalen, erzeugt in SPSS über Analysieren → Skalierung → Reliabilitätsanalyse → Statistik → Deskriptive Statistiken für Item/Zwischen Items Korrelationen/Kovarianzen. Die Werte stimmen mit den Mplus-Berechnungen überein, wenn in Mplus ebenfalls listenweiser Fallausschluss (listwise = on;) angefordert wird.

2.2 Einlesen von Summary-Daten (Kovarianz- oder Korrelationsmatrizen)

Wie zu Beginn von Kapitel 2 bereits erwähnt, ist es in manchen Fällen praktisch, Summary-Daten anstelle von Individualdaten einzulesen. Summary-Daten sind z.B. Kovarianz- oder Korrelationsmatrizen, ggf. ergänzt durch Mittelwerte und/oder Standardabweichungen aller Variablen. Wir zeigen hier das Einlesen von Summary-Daten für das Beispiel mit den sechs KFT-Variablen. Die einfachste Möglichkeit zum Einlesen von Summary-Daten ist das Eingeben der Matrix in eine einfache Textdatei. Dazu kann man in Windows z.B. das einfache Editor-Programm (**Start → Programme → Zubehör → Editor**) oder WordPad benutzen.

Abbildung 2.6 zeigt ein Editorfenster (Dateiname: **KFT_Summary-Daten.txt**) mit den eingegebenen Mittelwerten (1. Zeile), Standardabweichungen (2. Zeile) und der Produkt-Moment-Korrelationsmatrix der sechs beobachteten KFT-Variablen (die Datei befindet sich ebenfalls auf der Begleit-CD). Als Dezimaltrennzeichen müssen wie bei Individualdaten Punkte (keine Kommas) verwendet werden. Die Statistiken in einer solchen Textdatei können problemlos von Mplus gelesen und als Grundlage für die Schätzung beispielsweise eines Strukturgleichungsmodells verwendet werden.

11.899	8.974	12.387	7.732	11.112	8.281
4.656	4.244	3.353	2.604	5.448	3.437
1.000					
.547	1.000				
.388	.373	1.000			
.385	.401	.462	1.000		
.440	.508	.372	.431	1.000	
.399	.496	.383	.447	.496	1.000

Abbildung 2.6: Editor-Datei mit Summary-Daten der sechs KFT-Variablen. Die erste Zeile enthält die Mittelwerte der Variablen. Die zweite Zeile enthält deren Standardabweichungen. Darunter ist die Produkt-Moment-Korrelationsmatrix dargestellt. Abb. 2.7 zeigt, wie diese Daten korrekt in Mplus eingelesen werden können.

Die Mplus-Syntax zum Einlesen der Daten in der Datei **KFT_Summary-Daten.txt** ist in Abbildung 2.7 dargestellt. Unter **data** wird der Name des Datensatzes angegeben. Außerdem muss festgelegt werden, was für eine Art von Summary-Statistiken eingelesen werden, wobei die Reihenfolge der Befehle wichtig ist. Die Zeile **type = means std corr;** bedeutet, dass im Datensatz **KFT_Summary-Daten.txt** zunächst die Mittelwerte, dann die Standardabweichungen und dann die Korrelationen der Variablen eingegeben sind (zum Einlesen einer Kovarianzmatrix würde man den Befehl **cova** anstelle von **corr** benutzen).

Der Befehl **nobservations = 455;** teilt Mplus mit, auf wie vielen Beobachtungseinheiten die einzulesenden Summary-Daten beruhen (hier: $N = 455$). Diese Information kann aus Summary-Daten (im Gegensatz zu Individualdaten) nicht vom Programm erschlossen werden. Die Option **variable: names =** dient wiederum dazu, die Variablennamen zu definieren, die ebenfalls nicht in der Datendatei enthalten sind.

Die **basic**-Option steht für Summary-Daten nicht zur Verfügung. Aus diesem Grund werden zur Datenkontrolle Deskriptivstatistiken über **output: sampstat;** angefordert. Die Option **sampstat** steht hierbei für „sample statistics“ und liefert uns Deskriptivstatistiken für die sechs Variablen (in diesem Fall die beobachteten Mittelwerte sowie die Kovarianzmatrix). Da wir kein Modell explizit spezifizieren, schätzt Mplus automatisch ein sogenanntes „Nullmodell“ (oder Unabhängigkeitsmodell). Das Nullmodell nimmt an, dass zwischen den Variablen keinerlei Zusammenhänge bestehen. Als Modellparameter werden daher nur die Mittelwerte und Varianzen der Variablen geschätzt. Der Output zu diesem Modell ist nicht gesondert dargestellt, kann jedoch der Begleit-CD entnommen werden.

The screenshot shows the Mplus software interface. The title bar reads "Mplus - KFT-Summary-Daten-einlesen.inp". The menu bar includes File, Edit, View, Mplus, Graph, Window, and Help. A toolbar with various icons is visible above the main window. The main window displays the Mplus input script:

```
title: KFT-Datensatz einlesen
      Hier: Summary-Daten aus Datei "KFT_Summary-Daten.txt"
      Datei enthält die Mittelwerte, Standardabweichungen
      und Produkt-Moment-Korrelationen der 6 KFT-Variablen
      Berechnung von Deskriptivstatistiken und Nullmodell
      Annahme im Nullmodell: alle Variablen sind unkorreliert

data: file = KFT_Summary-Daten.txt;
      type = means std corr;
      nobservations = 455;

variable: names = kft_v1 kft_v3 kft_q1 kft_q3 kft_n1 kft_n3;

output: sampstat;
```

The status bar at the bottom left says "Ready" and at the bottom right says "Ln 1, Col 1".

Abbildung 2.7: Mplus-Input-Datei zum Einlesen der Summary-Daten aus Abb. 2.6 (Datei **KFT_Summary-Daten.txt**). Die basic-Option kann bei Summary-Daten nicht verwendet werden. Aus diesem Grund werden zur Datenkontrolle Deskriptivstatistiken über output: sampstat; angefordert. Da kein Modell explizit spezifiziert wird, schätzt Mplus automatisch ein „Nullmodell“ mit allen Variablen, die unter names = aufgelistet sind.

Im folgenden Kapitel wird anhand des Beispiels linearer Strukturgleichungsmodelle in die Modellspezifikation in Mplus eingeführt. Hierbei werden zunächst simple lineare Regressionsmodelle betrachtet und anschließend komplexere Strukturgleichungsmodelle wie konfirmatorische Faktorenanalysen und latente Pfadanalysen.