



Datei-Operationen mit Almo

Kurt Holm

Almo Statistik-System
www.almo-statistik.de
holm@almo-statistik.de
kurt.holm@jku.at

Im Text wird häufig auf das Dokument **P0** Bezug genommen. Dabei handelt es sich um das Almo-Dokument "Arbeiten mit Almo.PDF" (Dokument 0).

Siehe insbesondere auch das Almo-Handbuch
"Teil2: Almo-Programmiersprache"

Weitere Almo-Dokumente

Die folgenden Dokumente können alle von der Handbuchseite in www.almo-statistik.de heruntergeladen werden

0. Arbeiten_mit_Almo.PDF (1 MB)
- 1a. Eindimensionale Tabellierung.PDF (1.8 MB)
- 1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
2. Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
3. Nicht-parametrische Verfahren.PDF (0.9 MB)
4. Kanonische Analysen.PDF (1.8 MB)
Diskriminanzanalyse.PDF (1.8 MB)
enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate Korrespondenzanalyse, optimale Skalierung
5. Korrelation.PDF (1.4 MB)
6. Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
7. Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
- 7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
8. Tests auf Mittelwertsdifferenz, t-Test.PDF (1,6 MB)
9. Logitanalyse.pdf (1,2MB) enthält Logit- und Probitanalyse
10. Koeffizienten der Logitanalyse.PDF (0,06 MB)
11. Daten-Fusion.PDF (1,1 MB)
12. Daten-Imputation.PDF (1,3 MB)
13. ALM Allgemeines Lineares Modell.PDF (2.3 MB)
- 13a. ALM Allgemeines Lineares Modell II.PDF (2.7 MB)
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer, Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
17. Clusteranalyse.PDF (3 MB)
18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
20. Latent Structure Analysis.PDF (1 MB)
21. Statistische Algorithmen in C (80 KB)
22. Conjoint-Analyse (PDF 0,8 MB)
23. Ausreisser entdecken (PDF 170 KB)
24. Statistische Datenanalyse Teil I, Data Mining I
25. Statistische Datenanalyse Teil II, Data Mining II
26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-Datenanalyse-System
27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfachantworten (0.8 MB)
28. Metrische multidimensionale Skalierung (MDS) (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS) (0,4 MB)
31. Pfadanalyse (0,7 MB)
32. Datei-Operationen mit Almo (1,1 MB)
33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)

INHALTSVERZEICHNIS

| | |
|---|----|
| 0. Einführung: Mögliche Datei-Operationen | 4 |
| 0.1 Kopieren von Dateien..... | 4 |
| 0.2 Aus einer Datei eine Unterdatei entnehmen..... | 4 |
| 0.3 Eine Datei in Unterdateien auftrennen | 5 |
| 0.4 Eine Datei an eine schon bestehende Datei anhängen | 5 |
| 0.5 Parallele Dateien zusammenfügen | 5 |
| 0.6 Dateien über eine Verbindungsvariable zusammenfügen..... | 5 |
| 0.7 Verschmelzen zweier Dateien | 5 |
| 0.8 Daten fusionieren | 6 |
| 0.9 Daten importieren..... | 6 |
| 0.10 Daten exportieren | 6 |
| 0.11 Datenmatrix transponieren..... | 6 |
| 0.12 Daten sortieren | 7 |
| 1. Daten-Formate in Almo | 7 |
| 1.1 Format FREI (Werte sind durch Trennzeichen getrennt)..... | 7 |
| 1.2 FORMAT FIX (Werte sind spaltengebunden)..... | 12 |
| 1.3 FORMAT DIREKT (binäres Format)..... | 14 |
| 1.4 Fremdformate | 14 |
| 2 Kopieren | 15 |
| 2.1 Nur einige ausgewählte Variable in Zielformat übernehmen | 23 |
| 2.1.1 Wenn Ziel-Datei im Format FIX..... | 24 |
| 2.1.2 Wenn Ziel-Datei im Format FREI | 24 |
| 2.1.3 Wenn Ziel-Datei im Format DIREKT | 27 |
| 3. Aus einer Datei eine Unterdatei entnehmen..... | 29 |
| 3.1 Prog00mz als Kopierprogramm | 38 |
| 4 Eine Datei in Unterdateien aufteilen | 39 |
| 5. Eine Datei an eine schon bestehende Datei anhängen | 44 |
| 6. Parallele Dateien zusammenfügen | 49 |
| 7. Dateien über eine Verbindungsvariable zusammenfügen | 53 |
| 7.1 Ein Beispiel: | 53 |
| 7.2 Ein 2. Beispiel: | 54 |
| 7.3 Regeln für das Zusammenfügen..... | 56 |
| 7.4 Zulässige Datenformate für das Zusammenfügen..... | 57 |
| 7.5 So funktioniert das Zusammenfügen..... | 58 |
| 7.5.1 Umgeordneter Datensatz bei Prog00mg und Prog00mv | 59 |
| 7.6 Eingabe in Prog00mf..... | 61 |
| 7.6.1 Datensatz bei Prog00mg und Prog00mv umgeordnet speichern..... | 65 |
| 8 Verschmelzen zweier Dateien | 66 |
| 8.1 Beispiel..... | 66 |
| 8.2 Eingabe in Prog00mx und Prog00my | 68 |
| 9 Transponieren einer Datenmatrix | 69 |
| 10 Sortieren einer Datei..... | 72 |
| 10.1 Ein Beispiel | 72 |
| 10.2 Eingabe in Prog07m2 | 73 |
| 10.2.1 Regeln für das Sortieren von Zeichenvariable. | 77 |

0. Einführung: Mögliche Datei-Operationen

Almo ist zwar ein Statistik-Programm. Es bietet jedoch vielfältige Datei-Operationen an, sogar solche, die man eher nur den Datenbank-Programmen zutrauen würde.

Zum Begriff "Datei"

Unter "Datei" verstehen wir eine Daten-Matrix, deren Zeilen Untersuchungsobjekte (z.B. Personen) sind und deren Spalten Variable sind (z.B. Merkmale von Personen). Ist eine Datei "verkehrt herum" aufgebaut, dann kann sie durch das Almo-Programm Prog01m3 transponiert werden. In Abschnitt 9 wird das Programm erläutert.

Wir verwenden auch häufig mit gleicher Bedeutung den Begriff "Datenmatrix". Im Almo-Dokument 24 "Statistische Datenanalyse", Teil I, Abschnitt P45.0 haben wir diesen Begriff ausführlicher definiert.

Die Datei ist irgendwo auf einem Datenträger gespeichert oder sie wurde in Almo geladen und befindet sich jetzt in einem Almo-Fenster.

Zum Begriff "Datensatz"

Das kann z.B. eine Person mit ihren Variablen sein. Der Datensatz ist eine Zeile aus der Datenmatrix.

Zum Begriff "Variable"

Eine Variable ist eine Spalte aus der Datenmatrix. Eine Datei kann numerische Variable und Zeichenvariable enthalten. Zeichenvariable sind Worte oder allgemein Zeichen, die keinen numerischen Inhalt unmittelbar ausdrücken. Siehe dazu Almo-Handbuch Teil 2 "Almo-Programmiersprache", Abschnitt 45.

Zum Begriff "Datei-Format"

Almo kann Daten in drei Formaten verarbeiten. Dies sind (1) FREI, (2) FIX und (3) DIREKT. Im anschließenden Abschnitt 1 werden diese Formate ausführlich vorgestellt.

Dies sind nun die Datei-Operationen, die in Almo enthalten sind

0.1 Kopieren von Dateien.

Dateien, die in einem der drei Formate FREI, FIX, DIREKT geschrieben wurden, können in eine neue Datei in eines dieser drei Formate geschrieben werden. Das Format FREI ist das am weitesten verbreitete Datenformat. Die Daten können aus Zahlenwerten und/oder Zeichenketten (z.B. Worten) bestehen. Sie sind durch Leerzeichen oder Komma oder Tabulator-Zeichen voneinander getrennt. Das Format wird oft auch durch die Datei-Erweiterungen "*.dat" oder "*.txt" oder "*.tab" gekennzeichnet. Die drei Formate werden in Abschnitt 1 ausführlich vorgestellt. Prog00mp ist die Standard-Programm-Maske für das Kopieren. Das Programm und auch die im folgenden genannten Programm-Masken werden in Almo gefunden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Hauptfensters. Auch Fremdformate (z.B. das "*.sav-Format" von SPSS) können in eines der drei Almo-Formate transferiert werden. Siehe Abschnitt 0.9 und 1.4.

0.2 Aus einer Datei eine Unterdatei entnehmen

Wenn eine Datei außerordentlich groß ist und der Forscher sich aber nur für eine Untergruppe von Untersuchungsobjekten interessiert, dann kann es sich rentieren, aus der großen Datei eine kleinere Datei herauszunehmen, z.B. nur Frauen. Es können aber auch komplexere

Untergruppen gebildet werden. Also kann beispielsweise folgende Such-Aufgabe erfüllen: In einer Datei sollen alle Personen gesucht werden, die männlich sind, jünger als 23 Jahre sind und von Beruf "Facharbeiter" oder "angelernte Arbeiter" sind. Die gefundenen Datensätze werden dann in eine Unterdatei ausgegeben. Dies leistet die Programm-Maske Prog00mz.

0.3 Eine Datei in Unterdateien auftrennen

Beispiel: Eine Datei von Personen wird nach dem Geschlecht in zwei selbständige Dateien aufgetrennt. Oder die Datei wird nach dem Beruf der Personen in die 3 Dateien Arbeitnehmer, Selbständige und Sonstige aufgetrennt. Oder eine Datei wird nach den zwei kombinierten Variablen Beruf und Geschlecht in 6 Dateien aufgeteilt. Dies leistet Prog01m1.

0.4 Eine Datei an eine schon bestehende Datei anhängen

Beispiel: Die Schüler der Klasse 4a bilden eine Datei, deren Spalten die Unterrichtsfächer und deren Zeilen die Schüler mit ihren Noten in den einzelnen Fächern sind. Daran angehängt werden sollen nun die Schüler der Klasse 4b mit ihren Noten in den einzelnen Fächern.

Allgemein: An eine schon bestehende Datei wird an deren unteren Ende eine andere schon bestehende Datei angehängt. Voraussetzung ist, dass die aufeinander folgenden Spalten der beiden Dateien, dieselben Variablen enthalten. Siehe die Programm-Maske Prog00md

0.5 Parallele Dateien zusammenfügen

Beispiel: Wir verfügen über zwei Datei von Schülern. Datei 1 enthält die Schüler als Zeilen und ihre Schulnoten als Spalten. Datei 2 enthält dieselben Schüler als Zeilen - in derselben Reihenfolge - und ihre demografischen Merkmale (wie Geschlecht, Alter usw.) als Spalten.

Allgemein: Zwei Dateien sind "parallel", wenn die Personen dieselben sind und die Reihenfolge, in der sie in den Zeilen aufeinander folgen, ebenfalls dieselbe ist - jedoch die Variablen verschieden sind. Die beiden Dateien werden dadurch zusammengefügt, dass die eine rechts an die andere "angelegt" wird. Siehe Programm-Maske Prog00me

0.6 Dateien über eine Verbindungsvariable zusammenfügen

Beispiel: Wir haben eine Datei 1 mit Personen. Die Datei umfasse 1000 Personen. Für jede Person sind 50 Variable vorhanden. Daneben haben wir eine 2. Datei, in der für verschiedene Gemeinden bestimmte Strukturmerkmale angegeben sind; beispielsweise: Zahl der landwirtschaftlich Tätigen, Zahl der Arbeiter, Prozentzahl der Wähler der Partei A etc. Die Datei umfasse 20 Gemeinden mit je 10 Strukturmerkmalen. Diese Strukturmerkmale sollen nun jeder Person aus der Personen-Datei entsprechend ihrer Gemeindezugehörigkeit als zusätzliche Variable angefügt werden. Möglich wird dies dadurch, dass sowohl in der Personen-Datei als auch in der Gemeinde-Datei die Gemeinde-Nummer als "Verbindungs-Variable" enthalten ist. Also erzeugt eine neue gemeinsame Datei, die aus der Personen-Datei und den angefügten Gemeinde-Variablen besteht. Siehe die 4 Programm-Masken Prog00mf, Prog00mg, Prog00mv und Prog00mw

0.7 Verschmelzen zweier Dateien

Beispiel: In der Datei "Mai_Kauf" sind die Kunden enthalten, die im Mai die Produkte A, B,C (in einer bestimmten Menge) gekauft haben. In der Datei "Junikauf" diejenigen, die im Juni die Produkte A, B, D, E gekauft haben.

Ein Teil der Kunden hat sowohl im Mai als auch im Juni gekauft. Diese Kunden sind also in beiden Dateien enthalten. Die Zahl der Datensätze und der Variablen in den beiden Dateien dürfen verschieden sein. Also "verschmilzt" die beiden Dateien zu einer gemeinsamen. Siehe Programm-Maske Prog00mx

0.8 Daten fusionieren

Ein Beispiel: Wir verfügen über eine Datei der politischen Parteipräferenzen von 1000 Personen. Die Datei enthält u.a. auch Variable, die als Determinanten der Parteipräferenz gelten können, wie (1) Einkommen, (2) Bildung, (3) Gewerkschaftsmitgliedschaft, (4) Kirchengangshäufigkeit usw. Wir besitzen auch eine zweite Datei mit 200 anderen Personen, in denen einige dieser Determinanten dieser (anderen) Personen enthalten sind - nicht jedoch deren Parteipräferenz. Die Aufgabe besteht nun darin deren Parteipräferenz zu "errechnen", da ja die Determinanten der Parteipräferenz vorhanden sind. In der ersten Datei kann z.B. ein "Allgemeines Lineares Modell" (etwa vom Typ der Regressionsanalyse) gerechnet werden, das die Regressionskoeffizienten liefert, mit deren Hilfe dann in der zweiten Datei aus den dort vorhandenen Determinanten die Parteipräferenz der 200 Personen errechnet wird. Dieser Vorgang wird "Datenfusion" genannt. Das Verfahren und auch seine Problematik werden sehr ausführlich im Almo-Dokument 11 "Datenfusion" und im Almo-Dokument 25 "Statistische Datenanalyse Teil II", Abschnitt 45.8 dargestellt. Hier wird es deswegen nicht behandelt.

0.9 Daten importieren

Daten, die in SPSS mit dem Datentyp (*.dat) Tabstopgetrennt oder mit dem Datentyp (*.csv) Kommagetrennt ausgegeben wurden, können unkompliziert mit der Programm-Maske Prog00a5 in das Almo-Format FREI konvertiert werden. Das Programm findet man durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters.

Daten, die in SPSS mit dem Datentyp (*.sav) ausgegeben wurden, können mit einem Programm von Joachim Gerich sehr komfortabel in eines der drei Almo-Formate konvertiert werden. Das Programm wird aus Almo gestartet. Klick auf (1) das Menü "Almo", (2) dann "Daten-Import und -Export", (3) dann "Methode I: SPSS <--->Almo: Daten-Import und -Export". Das Handbuch zum Daten-Import und -Export findet man im Almo-Unterverzeichnis ALMO_SPS unter dem Namen "manual.pdf".

Sollen Daten aus einem anderen Programm nach Almo importiert werden, dann geht dies nur wenn die Daten dort im "tabulator-getrennten" Format ausgegeben werden können. Letzteres Format kann dann in Almo im Spezialformat "-2" eingelesen werden oder mit der Programm-Maske Prog00m@ in ein Almo-Format konvertiert werden. Siehe nachfolgenden Abschnitt 1.1.

Alle in Almo angebotenen Programme zum Daten-Import findet man durch Klick auf den Knopf "Verfahren", dann "Daten-Import/Export".

0.10 Daten exportieren

Sollen Daten aus Almo z.B: nach SPSS exportiert werden, dann sollte die Programm-Maske Prog00m5 eingesetzt werden oder das oben unter 0.9 beschriebene Programm zum Daten-Import und -Export von Joachim Gerich benutzt werden.

0.11 Datenmatrix transponieren

Gelegentlich werden Daten "verkehrt herum" geschrieben. Die Untersuchungsobjekte (z.B. Personen) wurden als Spalten und die Variablen (z.B. deren Merkmale) als Zeilen geschrieben. Eine solche Datenmatrix kann durch viele Almo-Maskenprogramme (auch durch die meisten anderen Statistikprogramme) nicht ausgewertet werden. Sie muss zuvor transponiert werden. Das leistet Prog01m3. In Abschnitt 9 wird das Programm erläutert.

0.12 Daten sortieren

Almo enthält zwei Programm-Masken zum Sortieren von Dateien. Prog07m1 sortiert eine Datei, die nur aus numerischen Variablen besteht. Das allgemeine Prog07m2 sortiert eine Datei, die numerische Variable und/oder Zeichenvariable enthält. Dabei können bis zu 10 numerische Variable und/oder Zeichenvariable hintereinander "geschachtelt" als Sortiervariable eingesetzt werden.

1. Daten-Formate in Almo

In fast allen Almo-Programm-Masken werden Daten in folgender Eingabe-Box eingelesen:

File name: ".\Testdat\TESTDAT.FRE"
Format: frei
Variables: V1:20

Format der Daten
der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

Neben dem Datei-Namen muss der Benutzer das Format angeben, in dem die Daten vorliegen. Almo unterscheidet 3 Formate Es sind dies die Formate:

FREI,
FIX
DIREKT

Siehe die ausführliche Darstellung im Almo-Handbuch, Teil 2 "Almo-Programmiersprache", Abschnitt 9.2 und 14. Siehe auch das Almo-Dokument Nr. 0 "Arbeiten mit Almo.pdf".

1.1 Format FREI (Werte sind durch Trennzeichen getrennt)

Beispiel:

```
1 2 23.5 15 0 4
2 5 35 33 1 7
1 4 42 47 0 3
```

Dieses Format wird oft auch als "free-field format" oder "Ascii data" bezeichnet.

Die Werte der aufeinander folgenden Datensätze müssen nicht, wie in unserem Beispiel, exakt untereinander stehen.

Die Werte eines Datensatzes sind durch ein oder mehrere *Blanks* (Leerzeichen) oder ein *Komma* voneinander getrennt. In einem Datensatz dürfen beide Trennzeichen vorkommen. Die 1. Zeile könnte auch so geschrieben werden:

```
1 2 23.5 15, 0, 4
```

Wir würden eher davon abraten, Daten mit gemischten Trennzeichen zu schreiben, da sie dann von anderen Programmsystemen möglicherweise nicht gelesen werden können.

Almo akzeptiert auch das *Tabulatorzeichen* als Trennzeichen. Siehe nachfolgend.

Der *Punkt* dient als Dezimalpunkt. Das Komma hinter 15 und hinter 0 ist also ein Trennzeichen. Der Punkt in 23.5 ist ein Dezimalpunkt.

Das Komma als Dezimalzeichen ist durchaus erlaubt, muss aber dann extra deklariert werden. Wir werden das weiter unten im Text noch zeigen. Dann ist aber das Komma als Trennzeichen natürlich nicht erlaubt.

Wird das *Komma* durchgängig als Trennzeichen eingesetzt, dann würden unsere obigen Beispieldaten so aussehen:

```
1,2,23.5,15,0,4
2,5,35,33,1,7
1,4,42,47,0,3
```

Hinter den letzten Zahlenwert wird kein Komma geschrieben.

Tabulator als Trennzeichen

Daten, die das Tabulatorzeichen als Trennzeichen verwenden, können in Almo auch im Format FREI eingelesen werden. Das gelingt aber nur durch einen Trick. Man gibt in obiger Eingabe-Box als Format nicht FREI ein, sondern die Zahl -2. So können sie problemlos gelesen werden. Sollen Daten neu geschrieben werden, dann raten wir aber eher ab, den Tabulator als Trennzeichen einzugeben. Der Grund dafür: Lädt der Benutzer die Datei in ein Almo-Fenster, dann können sie dort nicht korrekt abgebildet werden, weil das Tabulatorzeichen nicht darstellbar ist. Die Daten wirken, wie wenn sie zerstört worden wären. Allerdings können mit Prog02ma die Daten korrekt gelesen und dann in einem Fenster gezeigt werden. Mit Prog00m@, Prog00a5, Prog00m4 und Prog00m6 liegen kurze Programm-Masken vor, die eine (z.B. aus SPSS oder Excel ausgegebene) Datei mit Tabulator als Trennzeichen in eine Almo-Datei im Format FREI mit Blank als Trennzeichen wandeln. Die Programme findet man durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Mit Prog02ma können, wie schon erwähnt, die Daten auch in korrekt lesbarer Form ausgegeben werden. Das Programm findet man ebenfalls durch Klick auf den Knopf "alle Progs".

Neben numerischen Variablen (Variable mit Zahlenwerten) sind auch Zeichenvariable (z.B. Worte) erlaubt.

Zahlenwerte

Die Zahlenwerte der numerischen Variablen können in folgenden Formen vorliegen.

Beispiele:

```
24          Ganzzahl
24.25       Dezimalzahl mit Punkt als Dezimalzeichen
24,25       Dezimalzahl mit Komma als Dezimalzeichen
            dann darf aber das Komma nicht als Trennzeichen zwischen
            den Werten verwendet werden
-24         negative zahl
+24         Zahl mit Plus-Zeichen
1.25e2      Zahl in Exponentialdarstellung (entspricht: 125)
12.3e-2     Zahl in Exponentialdarstellung (entspricht: 0.123)
kw          kein Wert; fehlender Wert für numerische Variable
            zulässig sind auch: KW k K
```

Exponentialdarstellung

Die Exponentialdarstellung von Zahlen ist nur für Daten erlaubt; oder anders formuliert:

Zahlen im Exponentialformat dürfen nur innerhalb einer Datenmatrix auftreten, nicht jedoch im Rahmen der Almo-Programmiersprache. Beispiel: Eine Umkodierungsanweisung wie

```
Wenn V5 kleiner 1.25e2 dann V5=1; Sonst V5=2; EndeWenn
```

ist unzulässig. Erstzweise kann man schreiben

```
Wenn V5 kleiner 1.25 * 10**2 dann V5=1; Sonst V5=2; EndeWenn
```

****** =ist in der Almo-Programmiersprache das Potenzierungszeichen.

Der Wert $12.3e-2$ kann in der Almo-Syntax ersetzt werden durch $12.3 * 10**-2$

Werte von Zeichenvariablen

Beispiele:

| | |
|---------|--|
| a | einzelnes Zeichen |
| aBc5D | Zeichenkette |
| 'a' | einzelnes Zeichen zwischen einfachen Apostrophen |
| 'abCD3' | Zeichenkette zwischen einfachen Apostrophen |
| 'ab CD' | Zeichenkette mit einem oder mehreren Blanks (muss zwischen einfachen Apostrophen stehen, wenn Blank als Trennzeichen zwischen Werten verwendet wird) |
| ' ' | Blank zwischen einfachen Apostrophen bedeutet: Zeichenvariable besitzt Kein-Wert leere Zeichenkette |

Der übliche Doppelapostroph wird von Almo nicht als Begrenzer sondern als Bestandteil der Zeichenkette interpretiert.

"Standard-FREI" und "erweitert-FREI"

In Almo wird noch zwischen "Standard-FREI" und "erweitert-FREI" unterschieden. Das Format "erweitert-FREI" bezeichnen wir auch als

"Nicht-Standard-FREI".

WICHTIG:

Befinden sich ihre Daten im Format "erweitert-FREI", dann geben Sie in der Box "Datei aus der gelesen wird" als Format "frei" an. Sie müssen dann aber zusätzlich die nachfolgende Optionsbox öffnen und entsprechende Einträge vornehmen. Siehe dort die verschiedenen Hilfeknöpfe.

The screenshot shows a dialog box with the following elements:

- Text: "Datei aus der gelesen wird" with a "Hilfe" button to its right.
- Text: "bei Datei-Problemen" below the first "Hilfe" button.
- Text input field: ".\TESTDAT\Meinedaten.fre"
- Text input field: "frei" with a "Hilfe" button to its right.
- Text: "Format der Daten" below the second "Hilfe" button.
- Text input field: "V1:15"
- Text: "der Datensatz enthält diese Variablen" and "Bei Format DIREKT schreiben Sie: alle_V" below the third input field.
- Text input field: "Wenn Dateiformat FIX oder Nicht-Standard-FREI" with a "Hilfe" button to its right.

Standard-FREI liegt vor, wenn ...

- a. die einzulesenden Datensätze nur aus numerischen Variablen (=Zahlenwerten) bestehen, also keine Zeichenvariablen (=Worte) vorhanden sind
 - b. wenn die Variablenwerte durch ein Trennzeichen voneinander getrennt sind, und zwar durch
 - a. ein Blank (oder auch mehrere Blanks)
 - b. ein Komma
 - c. als Sonderfall: das Tabulatorzeichen (siehe Anmerkung oben)
 - i. Blank und Komma dürfen gemischt in einem Datensatz vorkommen (nur bei Almo)
 - ii. Zwei Kommas hintereinander sind nicht erlaubt.
Beispiel: 24,,35 <--- verboten
 - iii. Zwei Kommas hintereinander mit einem Blank dazwischen sind bei numerischen Variablen (Zahlenwerten) nicht erlaubt, jedoch im "erweitert-frei"-Format bei Zeichenvariablen (Worten). Sie signalisieren dort ein fehlendes Wort bzw. Zeichen.
Beispiel: 24, ,35 <--- bedingt erlaubt bei Zeichenvariablen
- c. wenn bei Dezimalvariablen als Dezimalzeichen der Punkt (und nicht das Komma) verwendet wurde
- d. wenn der jeweils nächste Datensatz in einer neuen Zeile beginnt - wobei ein langer Datensatz durchaus über mehrere Zeilen gehen darf. Ein neuer Datensatz muss aber prinzipiell in einer neuen Zeile beginnen. Beispiel:

```
1 2 23.5 15 0 4 77 88 2 5
7 3 4
2 5 35 33 1 7 77 93 4 2 13
21 3
:
```

Beim 1. Datensatz stehen in der 2. Zeile noch die drei Werte 7 3 4. Der 2. Datensatz mit den Werten 2 5 35 usw. beginnt in einer neuen Zeile die letzten Werte 21 3 stehen in der nächsten Zeile. Wird ein Datensatz auf mehrere Zeilen verteilt, dann müssen diese nicht gleich lang sein.

Erweitert-FREI oder **Nicht-Standard-FREI**

Die Daten befinden sich im erweiterten FREI-Format, also nicht im "Standard-FREI - Format", wenn

- a. Zeichenvariable (=Worte) vorhanden sind. Beispiel:

```
1 2 23.5 Beamter 15 0 4
2 5 35 Bauer 33 1 7
1 4 42 Angestellter 47 0 3
```

Das Wort "Beamter" ist eine Ausprägung der Zeichenvariablen V4 "Beruf". Die Werte (die Ausprägungen) von Zeichenvariablen dürfen Blanks enthalten. Sie müssen dann zwischen Apostrophe gestellt werden. Beispiel:

```
1 2 23.5 'höherer Beamter' 15 0 4
```

Verwendet wird der einfache Apostroph (wie im Beispiel) oder der Doppelapostroph " Siehe die ausführliche Darstellung von Zeichenvariablen im Almo-Handbuch Teil2 "Almo-Programmiersprache", Abschnitt 45.

Wird als Trennzeichen zwischen den Werten das Komma (und nicht das Blank) verwendet, dann darf der Apostroph fehlen. Beispiel:

```
1,2,23.5,höherer Beamter,15,0,4
```

- b. wenn 2 oder mehr Datensätze hintereinander in einer Zeile stehen; d.h. wenn ein neuer Datensatz nicht in einer neuen Zeile beginnt. Siehe dazu Almo-Handbuch Teil 2 "Almo-Programmiersprache", Abschnitt 13.1
- c. wenn bei Dezimalvariablen als Dezimalzeichen nicht der *Punkt* sondern das *Komma* verwendet wurde. WICHTIG: Dann darf als Trennzeichen nicht das Komma sondern nur das Blank verwendet werden. Beispiel:

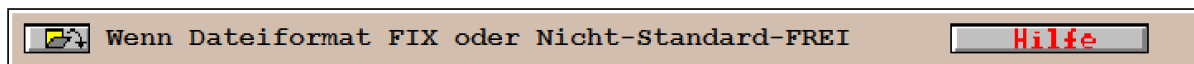
```
1 2 23,5      Beamter 15 0 4
2 5 35       Bauer 33 1 7
1 4 42       Angestellter 47 0 3
```

Das Komma in 23,5 ist ein Dezimalzeichen.

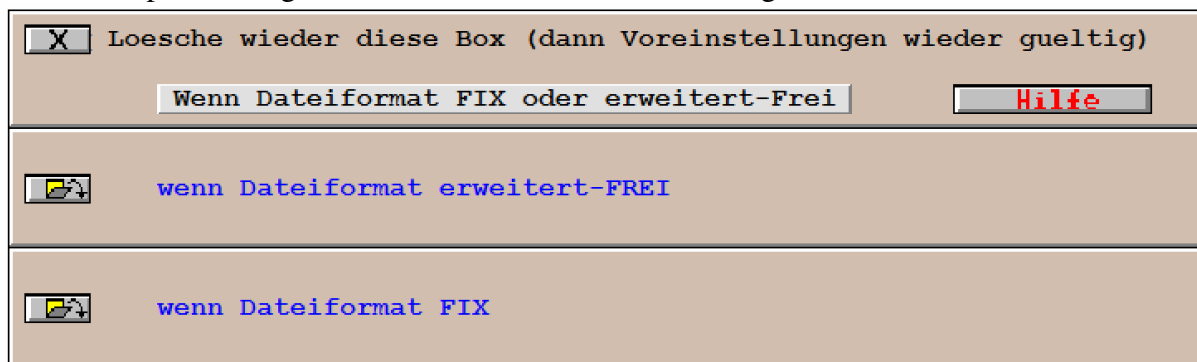
Wenn nur eine dieser Bedingungen gegeben ist, dann befinden sich die Daten im "erweitert-FREI-Format" bzw. "Nicht-Standard-FREI-Format".

Wenn Ihre Daten sich im Format "Standard-FREI" aber auch im Format "erweitert-FREI" befinden, dann geben Sie in der Box "Datei aus der gelesen wird" im 2. Eingabefeld als Format "frei" an.

Befinden sich ihre Daten im Format "Nicht-Standard-FREI" (erweitert-FREI), dann müssen Sie zusätzlich die bereits oben abgebildete sandfarbene Optionsbox öffnen und entsprechende Einträge vornehmen.



Wird die Optionsbox geöffnet, dann sieht man zunächst folgende 2 Sub-Boxen:



Geöffnet wird nun die Sub-Box für "erweitert-FREI". Wir füllen sie aus für die oben zuletzt abgebildete Daten-Matrix

Hier kommen noch die Anweisung hinzu:

```
Feld 1,1,4,12,2,1,1
```

wenn in einer Feld-Anweisung viele gleiche Zahlen sich wiederholen, z.B.

```
Feld 1,1,1,1,1,1,5,12,2,1,1,1
```

dann kann man kurz schreiben:

```
Feld 6*1,5,12,2,3*1
```

Im 1. Datensatz besitzt die 3. Variable den Wert 23.5. Für diese Variable ist also eine Feldbreite von 4 anzugeben. Der Dezimalpunkt nimmt einen Platz im Feld ein. Im 2. und 3. Datensatz ist der Variablenwert 35 und 42. Sie benötigen nur 2 Plätze von den vier verfügbaren. Sie werden korrekt gelesen. Der Dezimalpunkt (oder alternativ das Dezimalkomma) muss geschrieben sein, damit der Wert als Dezimalwert erkannt wird. Würde er fehlen, dann würde die Variable als Ganzzahl 235 gelesen.

Die Feldbreite der einzulesenden Variablen muss angegeben werden. Diese schreibt der Benutzer in der nachfolgenden Optionsbox. In unserem Beispiel hat die erste Variable eine Feldbreite von 1, die 2. Variable von 1, die 3. Variablen (23.5) von 4 etc. Die 4. Variable ist eine Zeichenvariable. Ihre Ausprägung lautet " Beamter". Sie besitzt eine Feldbreite von 12 wobei 5 Blanks links vor dem Wort "Beamter" stehen. Die Feldbreite muss 12 sein, damit auch der Beruf "Angestellter" in das Feld hinein passt. Die Feldbreite muss mindestens so groß sein wie der "breiteste" Wert. Der Benutzer muss die Feldbreite aller Variablen kennen.

Die Daten werden üblicherweise "rechtsbündig" in ihrem Feld stehen. Sie können aber auch "linksbündig" oder "mitten hinein" in das Feld geschrieben sein. Der 2. Datensatz könnte also auch so geschrieben sein:

```
rechtsbündig:      25..35.....Bauer3317
"mitten hinein":  25.35.....Bauer...3317
linksbündig:      2535..Bauer.....3317
```

Um die Stellung der Ziffern bzw. Buchstaben in ihrem jeweiligen Feld besser sichtbar zu machen, haben wir Punkte eingesetzt. Sie sollen die Blanks repräsentieren. Die Punkte dürfen natürlich nicht in den realen Daten stehen.

Beachte: Für die 3. Variable ist eine Feldbreite von 4 vorhanden. Die Zahl 35, die nur 2 Ziffern besitzt kann also rechtsbündig so geschrieben sein:

```
..35
"mitten hinein" .35.
linksbündig     35..
```

Nach der Feldbreite muss die Zahl der Dezimalstellen angegeben werden. Im Maskenprogramm klickt der Benutzer einfach im letzten Eingabefeld auf den Knopf mit dem nach unten weisenden Pfeil. Also setzt dann das Wort "Dezimal" ein.

In unserem Beispiel mit den Werten (im freien Format geschrieben)

```
1 2 23.5      Beamter 23 0 4
2 5 35       Bauer 33 1 7
1 4 42       Angestellter 47 0 3
```

lautet die Dezimal-Anweisung dann

Dezimal 0,0,1,0,0,0,0

Wenn in einer Dezimal-Anweisung viele gleiche Zahlen sich wiederholen, z.B.

Dezimal 0,0,0,0,0,0,2,1,1,1

dann kann man kurz schreiben:

Dezimal 6*0,2,3*1

Ganzzahlvariable und Zeichenvariable werden mit 0 angegeben.

1.3 FORMAT DIREKT (binäres Format)

Dies ist ein Almo-spezifisches binäres Format. Es ist sehr schnell, erlaubt den DIREKT-Zugriff auf ausgewählte Datensätze (sogar Variable) und ermöglicht das platzsparende Speichern der Daten auf Platte (oder sonstigem Datenträger) mit 1 bis 8 Byte (je nach Größe des Variablenwertes).

Eine von Ihnen im Format FREI oder FIX geschriebene Datei kann mit den Maskenprogrammen Prog45mh oder Prog00mp in das Format DIREKT gewandelt werden. Diese Maskenprogramme finden Sie durch Klick auf den Knopf "Verfahren" und dann Klick auf den Eintrag "Datei-Operationen". Dort finden Sie auch Programme für den umgekehrten Transfer von DIREKT zu FREI oder FIX. Sie können auch auf den Knopf "alle Progs" am Oberrand des Almo-Hauptfensters klicken. Almo präsentiert Ihnen dann eine alphabetisch geordnete Liste aller vorhandenen Programme.

EMPFEHLUNG: Wenn Sie mit denselben Daten mehrere Analysen durchführen wollen, dann ersparen Sie sich Arbeit, wenn Sie die Daten zuerst in eine Datei im Format DIREKT einschreiben.

1.4 Fremdformate

Wenn Ihre Daten in einem Fremdformat vorliegen, beispielsweise im sav-Format von SPSS, dann können Sie diese in ein Almo-Format konvertieren. Klicken Sie im Menü ALMO auf den Eintrag

"Daten-Import und -Export".

Lesen Sie dort zuerst die Hilfe. Und lesen Sie insbesondere im Almo-Handbuch "Teil 1: Bedienungsanleitung" Kapitel 8. Almo startet nach Klick auf den Eintrag ein Transferprogramm, das auch Variablen- und Ausprägungsnamen nach Almo überführt.

In Excel besteht die Möglichkeit, Daten im Format

"Tabs getrennt (*.txt)"

auszugeben. Auch im SPSS besteht diese Möglichkeit. Dort wird dieses Format genannt:

"Tabulator-getrennt (*.dat)"

Diese Formate können sehr einfach in ein Almo-Format übertragen werden. Klicken Sie auf den Knopf "Verfahren" und wählen Sie dann "Datei-Import/Export" und laden dann das Programm

Prog00m4 bzw. Prog00m6

Diese beiden Programme übertragen auch die Variablennamen nach Almo - nicht jedoch die Ausprägungsnamen.

2 Kopieren

Almo bietet mehrere Kopierprogramme an. Eine Datei, die in einem der drei Almo-Formate FREI, FIX und DIREKT vorliegt, wird kopiert in eine neue Datei in einem der drei Formate. Die Kopierprogramme sind zu finden nach Klick auf den Knopf "Verfahren" am Oberrand des Almofensters, dann "Datei-Operationen".

Hier soll nur die sehr leistungsfähige Programm-Maske **Prog00mp** erläutert werden. Sie erbringt folgende Kopier-Leistung:

| | | |
|---------------------|--------|---------------------|
| Quelldatei | | Zieldatei |
| Vorhandene Datei in | | Neue Datei in |
| einem der 3 Formate | | einem der 3 Formate |
| FREI | | FREI |
| FIX | -----> | FIX |
| DIREKT | | DIREKT |

Neben dem im folgenden abgebildeten und beschriebenen Prog00mp ist noch **Prog00mz** empfehlenswert. Es wird im Abschnitt 3 (Unterdateien bilden) vorgestellt.

Prog00mp.Msk
 Allgemeines Kopierprogramm
 Eine Quell-Datei im Format FREI oder FIX oder DIREKT
 in
 eine Zieldatei im Format FREI oder FIX oder DIREKT
 schreiben

 Die Quelldatei kann Zahlen- und Zeichenvariable enthalten

Die Daten können durch

1. die Ein- und Ausschluss-Option sowie
2. durch die Umkodierungs- und KeinWert-Option

modifiziert in die Zieldatei geschrieben werden

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare Variable= 20

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Quell-Datei Hilfe

bei Datei-Problemen

" .\TESTDAT\Almdez.fre"

frei Format der Daten Hilfe

V1:10 der Datensatz enthält diese Variablen
 Bei Format DIREKT schreiben Sie: alle_v

Wenn Dateiformat FIX oder Nicht-Standard-FREI Hilfe

Variable die von Quell-Datei zur Ziel-Datei übertragen werden

V1:10 diese Variablen aus der Quell-Datei
 sollen in die Ziel-Datei übernommen werden

Hilfe

Erläuterungen zu den Dialogboxen

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Normalerweise ist dies die Nummer der letzten Variable des eingelesenen Datensatzes. Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen. Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 oder mehr (z.B. 100) an. Sie dürfen großzügig sein.

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Quelldatei aus der gelesen wird

In der Eingabebox werden der Dateiname, das Format und die einzulesenden Variablen angegeben. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

Optionsbox: Wenn Dateiformat FIX oder Nicht-Standard-Frei (erweitert frei)

In unserem Beispiel enthält die Quelldatei nur numerische Variable in der von Almo vorgeschriebenen Form. Die Optionsbox muss nicht geöffnet werden.

Siehe oben Abschnitt 1 und Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.1.

Eingabebox: Variable, die von Quell- zur Ziel-Datei übertragen werden

Sie können die Variablen „von Hand“ in das Eingabefeld schreiben oder Sie klicken auf den Knopf mit den 2 kleinen symbolischen Fenstern. Almo öffnet dann die Dialogbox „Variable für Analyse auswählen“. In dieser können Sie die Variable, die in das Eingabefeld eingeschrieben werden sollen per Mausklick selektieren. Siehe die ausführliche Beschreibung dieser Dialogbox im Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.11.

Im Normalfall wird man alle Variable der Quell-Datei in die Ziel-Datei übertragen. Dann gibt es keine Probleme - wenn die Zieldatei im Format FREI oder DIREKT sein soll.

Soll sie im Format FIX sein, dann muss der Benutzer jedoch die Optionsbox

"Wenn Dateiformat der Zieldatei FREI oder FIX"

öffnen und die Feldbreite der Variablen eingeben. Dazu muss er sie auch kennen - es sei denn, alle Variable haben eine Feldbreite von 1 oder der Benutzer gibt eine so breite Feldbreite an, dass diese durch keinen Variablenwert überschritten wird.

Der Benutzer kann jedoch auch nur einige ausgewählte Variable übertragen wollen. Dann kann es Komplikationen geben. Diese werden anschliessend in Abschnitt 2.1 ausführlich erörtert.

Optionsbox: Ein- und Ausschliessen von Untersuchungseinheiten.

Siehe die ausführliche Darstellung dieser Optionsbox im nachfolgenden Abschnitt 3 und im Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.7. Wird diese Eingabebox geöffnet, dann wird das Kopierprogramm Prog00mp zu einem Programm, das eine Unterdatei bildet.

Optionsbox: Umkodierungen und Kein-Wert-Angaben

Siehe Almo-Dokument Nr.0 „Arbeiten mit Almo“, Abschnitt P0.5.

Eingabebox: Zieldatei

The image shows two overlapping dialog boxes. The top one is titled "Ziel-Datei, die erzeugt werden soll". It has a text input field containing the path ".\Progs\Neudat.fre" (highlighted in red), a dropdown menu showing "frei", and a label "Format der Daten in Ziel-Datei". A "Hilfe" button is on the right. The bottom dialog box is titled "Wenn Dateiformat der Zieldatei FREI oder FIX" and also has a "Hilfe" button.

Für die Zieldatei muss ein Pfad- und Dateiname angegeben werden. Zulässig ist auch die Windows Kurzform, bei der bis einschliesslich *Almo* der Pfad durch einen Punkt vertreten werden kann (wie oben in Abbildung).

Danach muss eines der 3 Almo-Dateiformat eingesetzt werden.

Wird FREI oder FIX als Format gewählt, dann muss die nachfolgende Optionsbox geöffnet werden. Wenn jedoch die Daten keine Zeichenvariable enthalten und im Standard-FREI-Format gespeichert werden sollen, dann kann die Optionsbox geschlossen bleiben.

Sonderfall: "Tabulator-getrenntes" Format FREI

Sollen die Daten zwar im Format FREI, jedoch durch Tabulator getrennt gespeichert werden, dann ist folgendermaßen vorzugehen:

1. In das 2. Eingabefeld wird als Format nicht FREI eingeschrieben sondern: -2
2. Die nachfolgende Optionsbox wird geöffnet. Dort wird als Trennzeichen "Tabulator" eingesetzt.

Wird die Optionsbox "Wenn Dateigormat der" geöffnet, dann sieht man zunächst

folgendes

| |
|--|
| <input checked="" type="checkbox"/> Loesche wieder diese Box (dann Voreinstellungen wieder gueltig) |
| Wenn Dateiformat der Zieldatei FREI oder FIX <input type="button" value="Hilfe"/> |
| <input type="checkbox"/> öffnen, wenn Dateiformat der Ziel-Datei FREI sein soll |
| <input type="checkbox"/> öffnen, wenn Dateiformat der Ziel-Datei FIX sein soll |

1. Sub-Box: Öffnen, wenn Format der Zieldatei FREI.

Diese Sub-Box muss geöffnet werden, wenn Die Zieldatei im freien Format geschrieben werden soll. Das kann jedoch unterbleiben, wenn keine Zeichenvariablen vorhanden sind und im Standard-Frei-Format geschrieben wird.

| | |
|--|---|
| <input checked="" type="checkbox"/> Loesche wieder diese Sub-Box (Voreinstellungen wieder gueltig) | |
| Dateiformat der Ziel-Datei soll FREI sein | |
| <input type="button" value="↔"/> <input type="text" value="120"/> | Zeilenlänge Wieviel Spalten sollen in einer Zeile maximal beschrieben werden. Benötigt der Datensatz mehr Spalten, dann erfolgt ein Zeilenumbruch Der Datensatz wird dann in eine 2. bzw. 3. Zeile usw. geschrieben <input type="button" value="Hilfe"/> |
| <input type="button" value="▼"/> <input type="text" value="Punkt"/> | Dezimalzeichen <input type="button" value="Hilfe"/> Punkt= Dezimalwerte werden mit dem Punkt als Dezimalzeichen geschrieben Komma= Dezimalwerte werden mit dem Komma als Dezimalzeichen geschrieben |
| <input type="button" value="↔"/> <input type="text" value="16"/> | Genauigkeit <input type="button" value="Hilfe"/> Bei Format FREI werden die Zahlenwerte mit maximal dieser Zahl von signifikanten Ziffern in die Datei geschrieben Mögliches Maximum: 16 |
| <input type="button" value="▼"/> <input type="text" value="Blank"/> | Trennzeichen <input type="button" value="Hilfe"/> hier angegeben, welches Trennzeichen zwischen die Zahlen- bzw. Zeichenvariable geschrieben werden soll Möglich sind: Blank, Komma, Tabulator |

1. Eingabefeld: Zeilenlänge

Geben Sie an, wieviel Zeichen maximal in eine Zeile der zu erzeugenden Datei geschrieben werden sollen. Benötigt ein Datensatz eine längere Zeile, dann bricht Also um und schreibt den Rest des Datensatze in die 2. (oder weitere) Zeile. Wird der so auf mehrere Zeilen verteilte Datensatz später wieder eingelesen, dann entstehen keine Probleme.

Beispiel: In die zu erzeugende Datei sollen 50 einziffrige Datenwerte mit einem Blank als Trennzeichen geschrieben werden. Dann benötigen Sie 50 Zeichen für die

Datenwerte und 50 Zeichen für die Blanks. Es reicht also eine Zeilenlänge von 100 aus. So sparsam braucht man jedoch nicht zu sein. Geben Sie besser eine größere Zahl an, z.B. 120. Sie können aber auch eine kleinere Zahl angeben, z.B. 80. Es tritt aber trotzdem kein Problem auf. Also bricht nach 40 Datenwerte (und 40 Blanks) um und schreibt den Rest von 10 Datenwerte (und 10 Blanks) in eine 2.Zeile. Dabei wird ein Datenwert jedoch nicht auseinander gerissen. Die Zahl der Zeilenumbrüche ist nicht beschränkt.

Beachte: Die Zeilenlänge die der Benutzer angibt, darf nicht länger sein als die im Menü "Allerlei / Einstellungen" vom Benutzer festgelegte maximale Zahl von Zeichen je Zeile !

2. Eingabefeld: Dezimalzeichen

Siehe oben Abschnitt 1.1

3. Eingabefeld: Genauigkeit

Wenn der Benutzer beispielsweise für die Genauigkeit 7 oder 16 angibt, dann werden folgende Variablenwerte in folgender Weise in die Datei geschrieben

| Wert | in die Datei wird geschrieben | |
|--------------------------|-------------------------------|-------------------------|
| | bei Genauigkeit=7 | bei Genauigkeit=16 |
| (1) 12.34 | 12.34 | 12.34 |
| (2) 1.000 | 1 | 1 |
| (3) 1234.567890123456 | 1234.568 | 1234.567890123456 |
| (4) 0.001234567890123456 | 0.0012346 | 0.0012345678901235 |
| (5) 1234567.890123456 | 1.234568e+006 | 1.234567890123456e+006 |
| (6) 0.00000000123456789 | 1.234568e-009 | 1.2345678900000000e-009 |
| (7) 0.0000000000000001 | 0 | 0 |

Es gelten folgende Regeln

1. Bei Format FREI wird zuerst überprüft aus wie vielen signifikanten Ziffern die Zahl besteht. In Beispiel 1 besteht die Zahl 12.34 aus 4 signifikanten Ziffern. Diese werden geschrieben. Bei Beispiel 2 existiert nur 1 signifikante Ziffer. Diese wird geschrieben.
2. Die Zahl in Beispiel 3 besteht aus 16 signifikanten Ziffern. Bei einer Genauigkeit=7 werden dann nur maximal 7 Ziffern geschrieben, Die letzte Ziffer wird gerundet.
3. Bei Zahlen kleiner 1, wie in Beispiel 4, werden bei Genauigkeit=7 maximal 7 Kommastellen geschrieben, bei Genauigkeit=16 maximal 16. Die letzte Ziffer wird gerundet.
4. Werte größergleich 1.0e6, wie in Beispiel 5, werden im e-Format geschrieben. Dabei umfasst die Mantisse so viele Ziffern, wie als Genauigkeit gewünscht werden. Die letzte Ziffer der Mantisse wird gerundet
5. Werte kleiner 9.536743e-06, wie in Beispiel 6, werden ebenfalls im e-Format geschrieben. Die Ziffernzahl der Mantisse ist gleich der gewünschten Genauigkeit. Die letzte Ziffer der Mantisse wird gerundet
6. Werte kleiner 2.22044605e-15, wie in Beispiel 7, werden als 0 geschrieben
7. Ist die zu schreibende Zahl beispielsweise 1234.56 und als Genauigkeit wird nur 2 angegeben, dann wird der Vorkommateil auf jeden Fall geschrieben. In die Datei wird also 1234 geschrieben.

EMPFEHLUNG: Bei "normalen" Daten genügt eine Genauigkeit von 7. Da diese in Also voreingestellt ist, brauchen Sie keine Angabe zu machen.

BEACHTEN: Als Genauigkeit können Werte zwischen 1 und 16 angegeben werden. Da die Computer-interne Darstellung der Zahlen maximal 15 bis 16 Ziffern umfasst, ist eine Genauigkeit von größer als 16 sinnlos.

4. Eingabefeld: Trennzeichen

Siehe oben Abschnitt 1.1

2. Sub-Box: Öffnen, wenn Format der Zieldatei FIX

Wird die Sub-Box geöffnet, dann sieht man folgendes:

Loesche wieder diese Sub-Box (Voreinstellungen wieder gueltig)

Dateiformat der Ziel-Datei soll FIX sein

Zeilenlänge
Wieviel Spalten sollen in einer Zeile maximal beschrieben werden. Benötigt der Datensatz mehr Spalten, dann erfolgt ein Zeilenumbruch. Der Datensatz wird dann in eine 2. bzw. 3. Zeile usw. geschrieben

Punkt
Dezimalzeichen
Punkt= Dezimalwerte werden mit dem Punkt als Dezimalzeichen geschrieben
Komma= Dezimalwerte werden mit dem Komma als Dezimalzeichen geschrieben

FELD-Anweisung
Geben Sie hier die Feldbreite je Variable an, z.B. FELD 14,3,5,3,5,3,4,5,6,12

DEZIMAL-Anweisung
Wenn die Variablen Dezimalwerte besitzen, dann geben Sie hier die Kommastellen je Variable an, z.B. DEZIMAL 6*0,1,0,0,9

1. Eingabefeld: Zeilenlänge

siehe oben bei 1. Sub-Box

2. Eingabefeld: Dezimalzeichen

Siehe oben Abschnitt 1.1

3. Eingabefeld: Feld-Anweisung

Siehe auch oben Abschnitt 1.2

Beim Format FIX stehen die Werte ohne Trennzeichen direkt hintereinander.

Die Werte könnten z.B. folgende sein

```
1 2 4 0 23.55 Beamter 23 0 4
```

Beachten Sie: Es dürfen auch Zeichenvariable (Worte) vorkommen. Im fixen Format würden die Werte dann so geschrieben sein:

```
124023.55 Beamter2304
```

Damit diese Wertefolge von Almo richtig gelesen werden kann muss noch die Feldbreite angegeben werden. In unserem Beispiel hat die erste Variable eine Feldbreite von 1, die 2. 3. und 4. Variable ebenfalls von 1, die 5. Variablen (23.55) von 5 etc. Die 6. Variable ist eine Zeichenvariable. Ihre Ausprägung lautet "Beamter". Sie besitzt eine Feldbreite von 12 wobei 5 Blanks links vor dem Wort "Beamter" stehen.

Die Feldanweisung würde in diesem Fall also lauten:

```
Feld 1,1,1,1,5,12,2,1,1  
oder kürzer: Feld 4*1,5,12,2,1,1
```

Sie können also anstelle 1,1,1,1 kurz schreiben: 4*1

Die Feldbreite muss mindestens so groß sein wie der "breiteste" Wert. Wenn beispielsweise die Zeichenvariable "Beruf" die Ausprägungen

```
Beamter  
Arbeiter  
Angestellter
```

besitzt, dann muss die Feldbreite mindestens gleich 12 sein. Die Ausprägung "Beamter" wird deswegen rechtsbündig mit 5 Blanks davor geschrieben. Entsprechendes gilt natürlich auch für numerische Variable. Der Benutzer muss die Feldbreite aller Variablen kennen. Dazu kann die Quelldatei zuerst in ein Fenster geladen werden und inspiziert werden. Befindet sich die Quelldatei im nicht lesbaren Format Direkt, dann kann sie mit Hilfe von Prog02ma sichtbar gemacht werden.

Soll eine im Format FREI oder DIREKT vorliegende Datei im Format FIX geschrieben werden dann kann der Benutzer für jede Variable eine Feldbreite angeben, die so groß ist, dass kein Fehler entstehen kann. Dem Sinn des fixen Formats, möglichst kleine Dateien zu erzeugen, wird dadurch nicht entsprochen.

Vergessen Sie nicht das Wort "Feld". Im Maskenprogramm klicken Sie einfach auf den Knopf mit dem nach unten weisenden Pfeil. Almo setzt dann das Wort "Feld" ein.

4. Eingabefeld: Dezimal-Anweisung

Nach der Feldbreite muss die Zahl der Dezimalstellen angegeben werden. Beim Einlesen der Daten werden keine Dezimalstellen angegeben. Im Maskenprogramm klicken Sie einfach im letzten Eingabefeld auf den Knopf mit dem nach unten weisenden Pfeil. Almo setzt dann das Wort "Dezimal" ein.

Wenn beispielsweise die ursprünglich eingelesenen Wert folgende waren:

```
1 2 4 0 23.55 Beamter 23 0 4  
dann lautet die Dezimal-Anweisung
```

```
Dezimal 0,0,0,0,2,0,0,0,0  
oder kurz: Dezimal 4*0,2,4*0
```

Ganzzahlvariable und Zeichenvariable werden mit 0 angegeben. Der Benutzer besitzt für Zahlenvariable die Freiheit mehr oder weniger Dezimalstellen anzugeben. Er kann auch für Ganzzahlvariable Dezimalstellen angeben. Er muss jedoch darauf achten, dass die zuvor festgelegte Feldbreite nicht überschritten

wird.

2.1 Nur einige ausgewählte Variable in Zieldatei übernehmen

Der Benutzer kann jedoch auch nur einige ausgewählte Variable übertragen wollen. Dann kann es Komplikationen geben.

Betrachten wir ein Beispiel: Die Quelldatei ist folgende:

| V1 | V2 | V3 | V4 | V5 |
|-------|----|------|----|----|
| Hofer | 31 | ja | 21 | 85 |
| Taler | 32 | nein | 22 | kw |
| Mayer | 33 | nein | 19 | 74 |
| : | : | : | : | : |

V1 und V3 sind Zeichenvariable, die anderen sind numerische Variable.

Die Namensdatei zu dieser Daten-Datei ist folgende (sie wird für das Kopierprogramm nicht gebraucht)

```
Name 1=Familiename;  
Name 2=PersonNr;  
Name 3=TreibtSport:ja,nein;  
Name 4=Lebensalter;  
Name 5=Körpergewicht;
```

Die Datei ist im Almo-Ordner TESTDAT unter dem Namen "Quell1.fre" zu finden, die Namensdatei unter dem Namen "Quell1.nam".

Die Quell-Datei umfasst die Variablen V1:5
In die Ziel-Datei sollen übertragen werden V1, 3, 5, 4

Die entsprechende Eingabebox im Kopierprogramm ist folgende:

Variable die von Quell-Datei zur Ziel-Datei übertragen werden

↔ [] [] V1,3,5,4

diese Variablen aus der Quell-Datei
sollen in die Ziel-Datei übernommen werden

WICHTIG ----> **Hilfe**

Die so ausgefüllte Programm-Maske Prog00mp ist unter dem Namen "Kopier.Alm" zu finden (durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters).

V2 soll also nicht übertragen werden. Im Datensatz in der Ziel-Datei folgt also auf V1 unmittelbar V3. Auch werden V4 und V5 in der Aufeinanderfolge vertauscht.

Als Zieldatei wird in "Kopier.Alm" die Datei "Durcheinander.xxx" angegeben. Für xxx ist je danach, ob die Datei im Format FREI oder FIX oder DIREKT erzeugt werden soll einzugeben: "fre" oder "fix" oder "dir". Für die beiden ersten muss in der Programm-Maske die nachfolgende Optionsbox geöffnet werden und die Zeichenvariablen V1 und V3 angegeben werden. Bei FIX müssen auch noch die Feldbreiten eingetragen werden.

Die durch "Kopier.Alm" entstandene Ziel-Datei "Durcheinander.xxx" ist folgende:

| | | | |
|-------|------|----|----|
| Hofer | ja | 85 | 21 |
| Taler | nein | kw | 22 |
| Mayer | nein | 74 | 19 |
| . | . | . | . |
| . | . | . | . |

Die seitherige Variable V2 fehlt jetzt, die seitherigen V4 und V5 sind vertauscht.

2.1.1 Wenn Ziel-Datei im Format FIX

Werden alle Variable der Quell-Datei in die Ziel-Datei übertragen, dann muss (wie oben schon ausgeführt) der Benutzer die Optionsbox "Wenn Dateiformat der Zieldatei FREI oder FIX" öffnen und die Feldbreite der Variablen eingeben. Dazu muss er sie auch kennen.

Sollen nur einige ausgewählte Variable übertragen wollen, dann gilt das was im folgenden Abschnitt für das Format FREI ausgeführt wird. Für den Benutzer kann es doppelt mühsam werden, da er auch die Feldbreite der Variablen (für deren jeweils veränderte Position) mitteilen muss - es sei denn alle Variable haben eine Feldbreite von 1 oder der Benutzer gibt eine so breite Feldbreite an, dass sie durch keinen Variablenwert überschritten wird.

Wir raten davon ab, für die Zieldatei das Format FIX anzufordern. Die Fehleranfälligkeit ist zu groß.

2.1.2 Wenn Ziel-Datei im Format FREI

Wenn Sie für die Ziel-Datei das Format FREI festgelegt haben und danach mit dieser neuen Datei arbeiten und rechnen, dann können Sie in den verschiedenen Almo-Programmen die Daten lesen mit der Angabe

1. Möglichkeit: V1, 3, 5, 4
2. Möglichkeit: V1:4

Im Prinzip könnten die Daten auch mit V7, 9, 1, 25 oder sonstigen Nummern gelesen werden. Es müssen nur 4 sein. Beim Format FREI merkt sich Almo beim *Schreiben* der Daten *nicht* deren Variablennummern. Beim *Lesen* der Daten wird dann ein Wert nach dem anderen eingelesen und der Variablennummer zugewiesen, die (in der Eingabebox für die Datei im 3. Eingabefeld) angegeben wurde.

Wird die Datei "Durcheinander.fre" nun im Format FREI mit der Anweisung

V1, 3, 5, 4

mit der sie ursprünglich erzeugt wurde, in irgend ein anderes Almo-Programm eingelesen, dann werden den Daten folgende Variablennummern zugewiesen:

| | | | |
|-------|------|-----|----|
| V1, | 3, | 5, | 4 |
| --- | -- | --- | -- |
| Hofer | ja | 85 | 21 |
| Taler | nein | kw | 22 |
| Mayer | nein | 74 | 19 |
| . | . | . | . |
| . | . | . | . |

Wird sie hingegen mit

V1:4

eingelesen, dann findet folgende Zuweisung statt

| | | | |
|-----|----|-----|----|
| V1, | 2, | 3, | 4 |
| --- | -- | --- | -- |

| | | | |
|-------|------|----|----|
| Hofer | ja | 85 | 21 |
| Taler | nein | kw | 22 |
| Mayer | nein | 74 | 19 |
| . | . | . | . |
| . | . | . | . |

Die Datenmatrix selbst bleibt unverändert. Nur die zugeordneten Variablennummern haben sich geändert.

Wenn der Benutzer eine Datei der Variablennamen angelegt hat, dann bekommt er nun bei der 2. Möglichkeit Probleme, da unser Kopierprogramm Prog00mp die Namens-Datei unangetastet lässt. Er muss dann die Variablennamen-Datei ändern. Verwenden Sie dazu **Prog06m1**. Das Programm weist automatisch den Variablennamen die neuen Nummer zu.

```

N1=Familiename;
N2=TreibtSport:ja,nein;
N3=Körpergewicht:
N4=Lebensalter;

```

Das Programm wird gefunden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Das Problem umgehen Sie selbstverständlich, wenn Sie alle Variable in die Ziel-Datei übertragen - was wir sehr empfehlen.

Für unser Zahlenbeispiel wird die Programm-Maske Prog06m1 wie nachfolgend gezeigt ausgefüllt. Es entsteht das Programm "NeuNameNummer.Alm"

```

NeuNameNummer.Alm
Variablen-Namen sollen (fortlaufend) so unnummeriert werden
dass sie mit Name1 beginnen, dann Name2, 3, 4, ...

Beispiel: Es liegt folgende Datei der Variablennamen vor

Name 1=Familienname;
Name 2=PersonNr;
Name 3=TreibtSport:ja,nein;
Name 4=Lebensalter;
Name 5=Körpergewicht;

Für eine neue Datei werden gebraucht die Namen 1,3,4,5
jedoch in der Reihenfolge                          1,3,5,4
mit den neuen Nummern                              1,2,3,4

Das Programm erzeugt folgende neue Namensdatei


Name 1=Familienname;
Name 2=TreibtSport:ja,nein;
Name 3=Körpergewicht;
Name 4=Lebensalter;


S. Almo-Dokument 32 "Datei-Operationen mit Almo", Abschnitt 2

```

Programm-Bedienung --->


Vereinbare Variable=

 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert



in Ursprungsdatei

für neue Datei



0 = neue Namensdatei zusätzlich in Ergebnisliste zeigen
1 = nur neue Namensdatei speichern

Die so ausgefüllte Programm-Maske wird durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fenster unter dem Namen "NeuNameNummer.Alm" gefunden. Als neue Namensdatei "Neudat.nam" wird korrekt ausgegeben

```
N1=Familienname;
N2=TreibtSport:ja,nein;
N3=Körpergewicht:
N4=Lebensalter;
```

Die Nummern der Variablennamen wurden also an die veränderte Nummern der Variablen angepasst.

Im nachfolgenden Abschnitt 3 wird **Prog00mz** erläutert. Es wird dazu verwendet aus einer Datei eine Unterdatei heraus zu lösen. Es kann jedoch auch als Kopierprogramm eingesetzt werden. Dabei hat es den großen Vorteil, dass die Nummern der Variablennamen automatisch angepasst werden.

2.1.3 Wenn Ziel-Datei im Format DIREKT

Wenn Sie im Kopierprogramm "Kopier.Alm" für die Ziel-Datei das Format DIREKT festgelegt haben und danach mit dieser neuen Datei "Durcheinander.dir" arbeiten und rechnen, dann steht Ihnen nur die Möglichkeit zur Verfügung, die Daten mit der Anweisung

```
alle_V
```

einzulesen. Wenn Daten im Direkt-Format erzeugt werden, dann legt Almo einen "Dateikopf" an, in dem vermerkt ist, welche Variablennummer zu welchem Platz in der Aufeinanderfolge der Variablen gehört. Almo weist so der 2. eingelesenen Variablen die Variablennummer V3 zu. D.h. die Angabe "alle_V" entspricht in diesem Beispiel der Angabe "V1, 3, 5, 4" beim Format FREI.

Hier kann der Benutzer aber durch 2 Almo-Maskenprogramme klare Verhältnisse schaffen. Will der Benutzer eine derartige Direktdatei *um-nummerieren*, so dass sie mit V1 beginnend fortlaufend nummeriert ist, dann kann er dies mit dem eigens dafür geschaffenen **Prog00mw** tun. Das Programm wird gefunden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Das Programm ordnet automatisch die Variablen nach ihrer Variablennummer hintereinander an. Reihenfolge und Nummer stimmen dann wieder überein. Wir wollen an obigem Beispiel zeigen, was das Programm Prog00mw leistet.

Nachdem das Kopierprogramm "Kopier.Alm" gerechnet wurde, stehen in der Zieldatei, also in der neuen Datei im Format DIREKT "Durcheinander.dir" folgende Werte:

| | V1, | 3, | 5, | 4 |
|-------|------|----|-----|----|
| | --- | -- | --- | -- |
| Hofer | ja | 85 | 21 | |
| Taler | nein | kw | 22 | |
| Mayer | nein | 74 | 19 | |
| : | : | : | : | |
| : | : | : | : | |

Das für unser Beispiel ausgefüllte Prog00mw ist folgendes. Es ist zu finden unter dem Namen "UmNummer.Alm" durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters.

UmNummer.Alm

Eine DIREKT-Datei lesen. Die Datei besitzt eine diskontinuierliche Reihenfolge der Variablen-Nummern mit V1,3,5,4

lesen ----->

Das Programm leistet folgendes:
 Die Variable werden um-nummeriert, beginnend mit V1,
 dann fortlaufend V2,3,...

dann

in neue Datei im Format DIREKT und FREI speichern
 Für numerische Variable und Zeichenvariable

Hilfe zu Almo-Datenformate ---->

Programm-Bedienung --->

Vereinbare Variable=

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Geben Sie den Dateinamen ohne Erweiterung an
 Der Pfad- und Dateiname muss anders lauten als der der alten Datei

Almo erzeugt 2 Dateien:

1. eine nicht lesbare Almo-Arbeitsdatei mit der Erweiterung .dir
2. eine anschaulbare Datei im Format FREI mit der Erweiterung .fre

Das Programm erzeugt eine neue Datei "Geordnet.dir" im Format Direkt und zusätzlich "Geordnet.fre" im Format FREI mit folgender Reihenfolge der Variablen

| V1, | 2, | 3, | 4 |
|-------|------|-----|----|
| --- | -- | --- | -- |
| Hofer | ja | 85 | 21 |
| Taler | nein | kw | 22 |
| Mayer | nein | 74 | 19 |
| : | : | : | : |

Die Datenmatrix bleibt unverändert, die zugeordneten Variablennummern beginnen mit V1

und sind dann fortlaufend nummeriert mit V2,3,4.

Auch die Datei der Variablenamen kann so umgeordnet werden, dass die Variablenamen mit "Name 1=..." beginnend fortlaufend nummeriert sind. Das wird automatisch geleistet durch das bereits erwähnte **Prog06m1**, das wir bereits oben (ausgefüllt für unser Beispiel) unter dem Namen "NeuNameNummer.Alm" abgebildet haben.

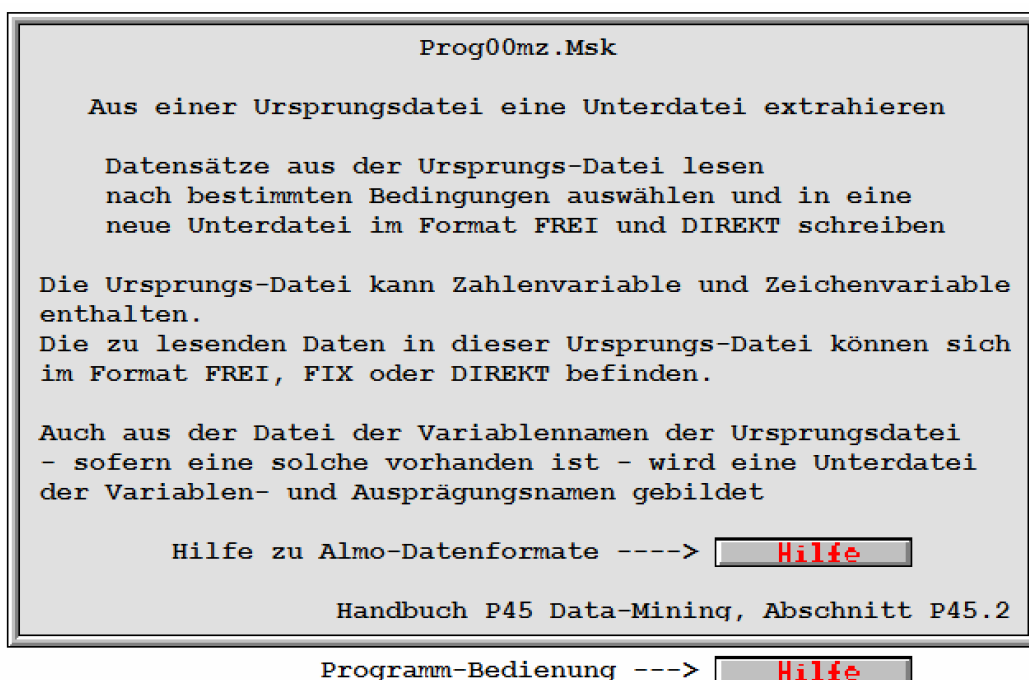
Das Programm wird ebenfalls gefunden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters.

Im nachfolgenden Abschnitt 3 wird **Prog00mz** erläutert. Es wird dazu verwendet aus einer Datei eine Unterdatei heraus zu lösen. Es kann jedoch auch als Kopierprogramm eingesetzt werden. Dabei hat es gegenüber Prog00mp den großen Vorteil, dass die Nummern der Variablen und der Variablenamen automatisch mit 1 beginnend und kontinuierlich fortlaufend angepasst werden.

3. Aus einer Datei eine Unterdatei entnehmen

Wenn eine Datei außerordentlich groß ist und der Forscher sich aber nur für eine Untergruppe von Untersuchungsobjekten interessiert, dann kann es sich rentieren, aus der großen Datei eine kleinere Datei herauszunehmen. Betrachten wir ein Beispiel: Die weltweite Pisa-Studie der Leistung von Schülern umfasst mehrere Gigabyte an Daten. Wenn der Forscher aber nur die europäischen Länder vergleichen will, dann ist es sinnvoll eine Unterdatei der europäischen Schüler zu erzeugen und nur mit dieser zu arbeiten. Almo bietet für diesen Zweck die Programm-Maske Prog00mz an. Sie wird gefunden durch Klick auf den Knopf "Verfahren" am Oberrand des Almo-Fensters, dann Klick auf "Datei-Operationen".

Wird die Eingabebox "Ein- und Ausschliessen von Untersuchungseinheiten" nicht geöffnet, dann wird Prog00mz zu einem sehr leistungsfähigem Kopierprogramm.



Speicher fuer x Variable

Vereinbare Variable=

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Ursprungsdatei aus der gelesen wird

Format der Daten

der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

Datei der Variablennamen der Ursprungsdatei

Leeren Sie das Eingabefeld
wenn keine Namensdatei
vorhanden ist

Wenn Dateiformat FIX oder Nicht-Standard-FREI

Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)
 Leere alle Eingabefelder

Ein- und Ausschliessen von Untersuchungseinheiten

nur diese Datensätzen verwenden
wenn Eingabefeld leer, dann alle

Untersuchungseinheiten in Analyse
nur einschließen
wenn

Untersuchungseinheiten aus Analyse
ausschließen
wenn

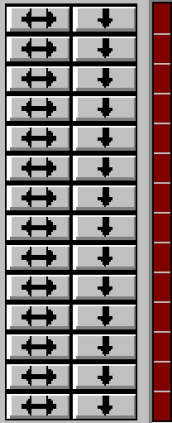
nur eine Zufallsstichprobe von x %
der Datensätze für Analyse verwenden

Startwert für Zufallsgenerator

Kein-Wert-Angabe und Umkodierungen von Zahlenvariablen
Die Variablen werden mit ihren umkodierten Werten bzw.
mit dem Kein-Wert-Code in die neue Datei übernommen

Kein-Wert-Angabe
Umkodierungen

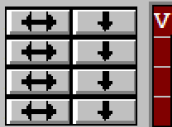
Hilfe
Hilfe



[...] erzeuge zusätzliche Felder für Umkodierungen / Kein_Wert-Angaben

Umkodierungen von Zeichenvariablen in Zahlenvariablen.
Zeichenvariablen, die nicht umkodiert werden, gehen
wieder als Zeichenvariablen in die neue Datei ein


Hilfe



V8 ('männl'=1; 'weibl'=2)

[...] erzeuge zusätzliche Felder für Umkodierungen

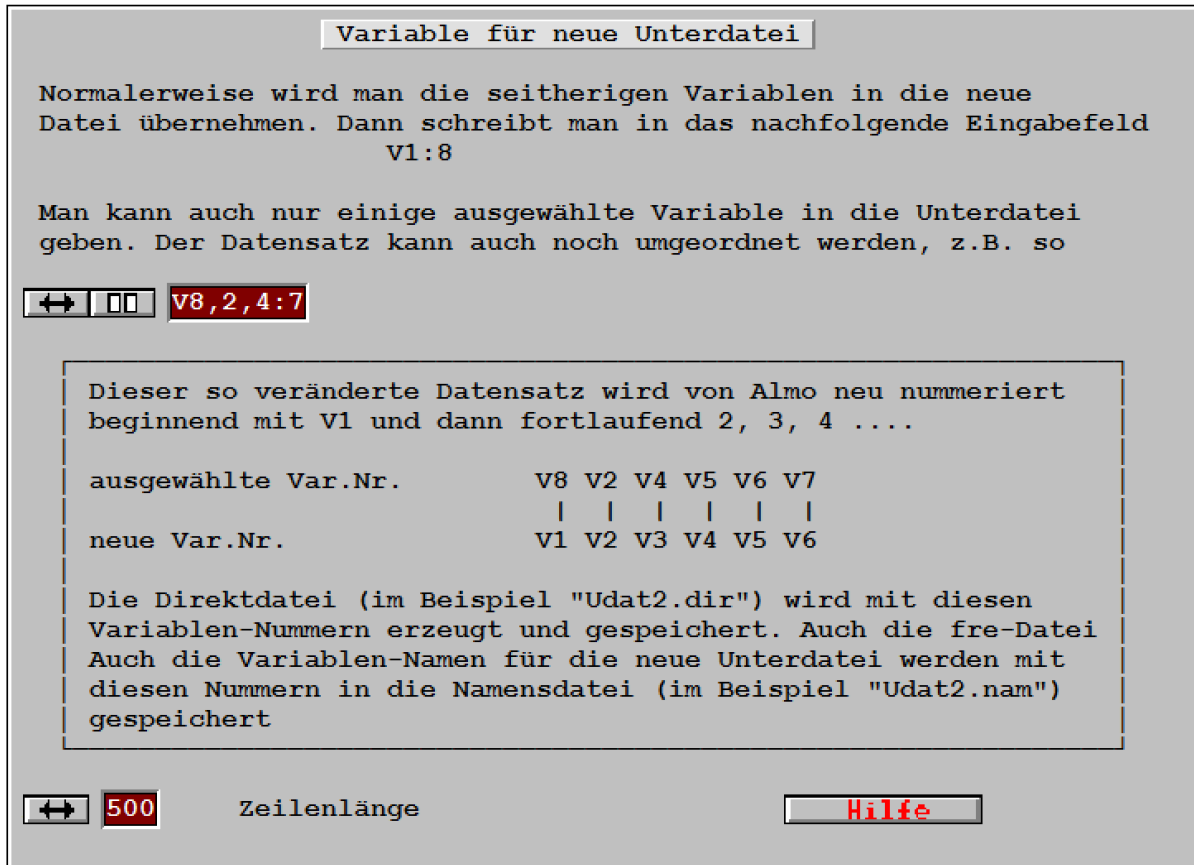
Neue Datei (im Format DIREKT und FREI), die erzeugt werden soll

 ".\Progs\Udat2"

Geben Sie den Dateinamen ohne Erweiterung an

Almo erzeugt 3 Dateien:

1. eine nicht lesbare Almo-Arbeitsdatei
mit der Erweiterung ____.dir
2. eine anschaubare Datei im freien Format
mit der Erweiterung ____.fre
3. eine Datei der Variablen- u. Ausprägungsnamen
mit der Erweiterung ____.nam



Erläuterungen zu den Dialogboxen

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Normalerweise ist dies die Nummer der letzten Variable des eingelesenen Datensatzes. Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen. Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 an

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Ursprungsdatei aus der gelesen wird

Im oberen Teil der Box werden der Dateiname, das Format und die einzulesenden Variablen angegeben. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

Im unteren Teil wird der Dateiname der Namensdatei angegeben. Ist keine Namensdatei vorhanden, dann bleibt das Eingabefeld leer. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.3.

Optionsbox: Wenn Dateiformat FIX oder Nicht-Standard-Frei (erweitert frei)

In unserem Beispiel enthält die Ursprungsdatei nicht nur numerische Variable sondern auch Zeichenvariable (Worte). Die Optionsbox muss also geöffnet werden und entsprechen ausgefüllt werden. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.1.

Optionsbox: Ein- und Ausschließen von Untersuchungseinheiten

| | | |
|--|---|--|
| <input type="button" value="X"/> | Loesche wieder diese Box (dann Voreinstellungen wieder gueltig) | |
| <input type="button" value="↔"/> | Leere alle Eingabefelder | |
| Ein- und Ausschliessen von Untersuchungseinheiten | | |
| | | <input type="button" value="Hilfe"/> |
| | | <input type="button" value="Hilfe"/> |
| <hr/> | | |
| <input type="button" value="↔"/> | <input type="checkbox"/> | nur diese Datensätzen verwenden wenn Eingabefeld leer, dann alle |
| | | <input type="button" value="Hilfe"/> |
| <hr/> | | |
| | | Untersuchungseinheiten in Analyse <input type="button" value="Hilfe"/> |
| | | nur einschließen wenn |
| <input type="button" value="↔"/> | <input type="button" value="V4 gleich 1"/> | |
| <hr/> | | |
| | | Untersuchungseinheiten aus Analyse <input type="button" value="Hilfe"/> |
| | | ausschließen wenn |
| <input type="button" value="↔"/> | <input type="checkbox"/> | |
| <hr/> | | |
| <input type="button" value="↔"/> | <input type="checkbox"/> | nur eine Zufallsstichprobe von x % der Datensätze für Analyse verwenden |
| | | <input type="button" value="Hilfe"/> |
| <input type="button" value="↔"/> | <input type="checkbox"/> | Startwert für Zufallsgenerator |
| | | <input type="button" value="Hilfe"/> |

Dies ist die zentrale Eingabebox für das Bilden von Unterdateien. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.7.1. Wir übernehmen teilweise den Text aus diesem Dokument.

Wichtig ist folgende Regel:

Das Ein- und Ausschliessen von Untersuchungseinheiten erfolgt vor eventuellen Umkodierungen

Almo bietet 4 Möglichkeiten, Datensätze in die Unterdatei zu übernehmen.

1. Nur die Datensätze mit der Nummer xy übernehmen. Beispiel nur die Datensätze 1 bis 100 (in der Reihenfolge des Einlesens aus der Datei) für die Unterdatei verwenden. Die restlichen Datensätze überlesen.
2. Das "Einschliessen": Ein Datensatz wird nur dann übernommen, wenn er eine bestimmte Bedingung erfüllt, z.B. wenn eine Untersuchungsperson männlich ist.
3. Das "Ausschliessen": Ein Datensatz wird ausgeschlossen, wenn er eine bestimmte Bedingung erfüllt, z.B. wenn eine Person jünger als 20 ist.
4. Nur eine Zufallsstichprobe von x % der Datensätze für die Unterdatei übernehmen

Folgende Regeln gelten:

- a. Der Benutzer kann nur eine der Möglichkeiten oder mehrere oder alle 4 Möglichkeiten benutzen.
- b. Die vom Benutzer selektierten Möglichkeiten werden von Almo in der Reihenfolge, wie sie oben untereinander stehen, abgearbeitet. Selektiert der

Benutzer z.B. alle 4 Möglichkeiten, dann verfährt Almo folgendermaßen:

- zuerst wird überprüft, ob die eingelesene Datensätze die vom Benutzer vorgegebenen Datensatznummern besitzen. Nur diese werden übernommen
- dann wird das Einschliessen durchgeführt,
- dann das Ausschliessen,
- dann wird aus den noch verbliebenen Datensätzen eine Zufallsstichprobe von x % gezogen.

Betrachten wir ein Beispiel zu Regel b:

- Im Eingabefeld "nur diese Datensätze verwenden ..." steht

100 : 200 (der Doppelpunkt heißt: bis)

Almo nummeriert die Datensätze (beginnend mit 1) durch, die es einen nach dem anderen aus der Ursprungsdatei einliest. Dem Benutzer bleibt das verborgen. Die obige Anweisung bedeutet also: Es werden nur Datensätze für die Unterdatei übernommen, deren fortlaufende Datensatznummer zwischen 100 und 200 liegt

- Im Eingabefeld "nur EINSchließen wenn ..." steht

Alter groesser 20

Das bedeutet: Von den eingelesenen Personen 100 bis 200 gehen nur die in die Unterdatei ein, die älter sind als 20. Alle anderen werden überlesen.

- Im Eingabefeld "AUSSchließen wenn ..." steht

Beruf gleich Beamter

Das bedeutet: Von den Personen 100 bis 200, die älter als 20 sind, werden alle Personen, die Beamte sind, ausgeschlossen. Alle anderen Berufe der über 20-Jährigen mit der fortlaufenden Datensatznummer 100 bis 200 werden analysiert.

- Im Eingabefeld "nur eine Zufallsstichprobe von x % der Datensätze für Analyse verwenden" steht

50

Das bedeutet: 50 % der noch übrig gebliebenen Personen werden zufällig ausgewählt und für die Unterdatei verwendet

Die 4 Möglichkeiten, Datensätze für die Unterdatei zu verwenden bzw. nicht zu verwenden werden von Almo durchgeführt bevor eventuelle Umkodierungen vorgenommen werden.

Das ist nur für die 2. und 3. Möglichkeit, d.h. das Einschliessen und das Ausschliessen von Bedeutung

Beispiel:

Die Variable „Beruf“ ist so kodiert

- (1) Arbeiter
- (2) Angestellter

- (3) Beamter
- (4) Freier Beruf
- (5) Unternehmer

Es sollen nur die Beamten und die Freien Berufe in die Unterdatei übernommen werden. Der Benutzer trägt also in das 2. Eingabefeld ein:

Beruf gleich 3 oder Beruf gleich 4

In der nachfolgenden Box "Umkodierungen und Kein-Wert-Angaben" wird vom Benutzer geschrieben

Beruf (1,2,3=1; 4,5=2)

Almo verfährt dann folgendermaßen: Es berücksichtigt nur Beamte und Freie Berufe. Personen mit anderen Berufen werden übersprungen. Danach wird Beamter in 1 umkodiert und Freier Beruf in 2.

Wir wollen die Eingabefelder der Optionsbox nochmals erläutern.

Eingabefeld 1: Nur diese Datensätzen verwenden

Almo nummeriert die Datensätze (beginnend mit 1) durch, die es einen nach dem anderen aus der Ursprungsdatei einliest. Dem Benutzer bleibt das verborgen.

Geben Sie die Datensätze an, die Sie für die Unterdatei verwenden wollen. Sie können z.B. durch die Anweisung

1:20, 50:60, 99

eine Auswahl treffen. Der Doppelpunkt heißt "bis".

Wenn Sie angeben

1:50

dann werden die ersten 50 eingelesenen Datensätze für die Unterdatei verwendet, die nachfolgenden nicht mehr.

Wenn Sie das Eingabefeld leer lassen, dann teilen Sie Almo dadurch mit, dass Sie diese Einschluß-Möglichkeit nicht benutzen wollen. Es werden dann alle Datensätze verwendet.

Eingabefeld 2: Untersuchungseinheiten in Analyse nur einschließen wenn

Eingabefeld 3: Untersuchungseinheiten aus Analyse ausschließen wenn

Betrachten wir ein Beispiel: Im Eingabefeld "nur Einschließen" steht

V8 kleiner 9

Das bedeutet: In die Unterdatei gehen nur Untersuchungseinheiten ein, die in der Variablen V8 einen Wert kleiner als 9 besitzen. Alle anderen werden überlesen.

Ein weiteres Beispiel: Im Eingabefeld "Ausschließen" steht

V6 gleich 1.1

Das bedeutet: Alle Untersuchungseinheiten, die in der Variablen V6 den Wert 1.1

besitzen werden ausgeschlossen. Alle anderen werden analysiert.

Ein drittes Beispiel: In beide Eingabefeld wird ein Eintrag gesetzt

Im Eingabefeld "nur Einschließen" steht V8 kleiner 9

Im Eingabefeld "Ausschließen" steht V6 gleich 1.1

Das bedeutet: In die Unterdatei gehen nur Untersuchungseinheiten ein, die in der Variablen V8 einen Wert kleiner als 9 besitzen. Von diesen werden dann diejenigen doch ausgeschlossen, die in der Variablen V6 den Wert 1.1 besitzen.

Wenn Sie keine Bedingungen setzen wollen, unter denen Untersuchungseinheiten ein- oder ausgeschlossen werden sollen, dann machen Sie beide Eingabefeld leer. Klicken Sie dazu auf den Rein-Raus-Knopf vor dem Eingabefeld.

Wollen Sie jedoch eine Ausschluß-Bedingung und/oder eine Einschluß-Bedingung setzen, dann müssen Sie einen logischen Ausdruck in das betreffende Eingabefeld schreiben. Möglich sind z.B. folgende Anweisungen

```
V5 gleich 7
V6 groesser 8
V7 kleiner 0.5

V8 nichtgleich V9
V10 groessergleich 2.5
V11 kleingleich 2
```

Auch UND sowie ODER sind möglich. Beispiel

```
V5 gleich 7 UND V10 groesser V20
V6 kleiner V5 ODER V12 nichtgleich 2
```

Die Also-Schlüsselworte **Wenn**, **Und**, **Oder** werden im Handbuch, Teil 2, "Also Programmiersprache", Abschnitt 29 - 32 mit vielen Beispielen erläutert.

Anstelle von V.. können Sie selbstverständlich auch einen Variablennamen schreiben, z.B. Einkommen groesser 4000

Eingabefeld 4 und 5: Nur eine Zufallsstichprobe von x % verwenden

Wenn Sie hier z.B. 25 angeben, dann zieht Also (indem es einen Zufallsgenerator verwendet) aus den eingelesenen Datensätzen eine Stichprobe von 25 %. Nur diese werden für die Unterdatei verwendet.

Also besitzt einen eingebauten Zufallsgenerator. Dieser Zufallsgenerator muss durch eine Startzahl (die dann in komplizierter Weise transformiert wird) in Betrieb gesetzt werden. Die Folge der Zufallszahlen, die Also dann erzeugt, ist bei jeder Wiederholung des Programms dieselbe - es sei denn der Startwert für den Zufallsgenerator wird verändert. Verwenden Sie eine ungerade 5- oder 6-stellige Zahl.

Bleibt das Eingabefeld leer, dann verwendet Also die Startzahl 78125

Eingabebox: Kein-Wert-Angabe und Umkodierungen von Zahlenvariablen

Eingabebox: Umkodierungen von Zeichenvariablen in Zahlenvariablen

Siehe dazu Also-Dokument Nr. 0 "Arbeiten mit Also", Abschnitt P0.5.

In der Programm-Maske muss der Benutzer eventuelle Umkodierungen von Zahlen- und Zeichenvariablen getrennt in 2 Eingabeboxen vornehmen. Wichtig ist folgende Regel:

Das Ein- und Ausschliessen von Untersuchungseinheiten erfolgt vor eventuellen Umkodierungen. Zuerst wird ein- bzw. ausgeschlossen, dann umkodiert


Sollen Zeichenvariable ausgewertet werden, dann müssen sie zuvor in Zahlenwerte umkodiert werden. Beispiel:

```
Geschlecht ('männlich'=1; 'weiblich'=2)
```

Erst nach dieser Umkodierung kann beispielsweise ausgezählt werden, wieviele der Untersuchungspersonen männlich bzw. weiblich sind.

Eingabebox: Neue Datei, die erzeugt werden soll

Neue Datei (im Format DIREKT und FREI), die erzeugt werden soll

 ".\Progs\Udat2"

Geben Sie den Dateinamen ohne Erweiterung an
Der Pfad- und Dateiname muss anders lauten
als der der alten Datei !!

Almo erzeugt 3 Dateien:

1. eine nicht lesbare Almo-Arbeitsdatei
mit der Erweiterung ____.dir
2. eine anschauliche Datei im freien Format
mit der Erweiterung ____.fre
3. eine Datei der Variablen- u. Ausprägungsnamen
mit der Erweiterung ____.nam

In unserem Beispiel erzeugt Almo folgende 3 neue Dateien

```
"C:\Almo\Progs\Udat2.dir"
```

```
"C:\Almo\Progs\Udat2.fre"
```

```
"C:\Almo\Progs\Udat2.nam"
```

Eingabebox: Variable für neue Unterdatei



Die Erläuterungen in der Eingabebox bedürfen keiner weiteren Ergänzung

Zeilenlänge:

Der Benutzer muss für die neue Datei angeben, wieviele Zeichen maximal in eine Zeile geschrieben werden sollen. Benötigt ein Datensatz eine längere Zeile als vom Benutzer angegeben, dann bricht Almo um und schreibt den Rest des Datensatzes in die 2. (oder weitere) Zeile. Wird der so auf mehrere Zeilen verteilte Datensatz später wieder eingelesen, dann entstehen dadurch keine Probleme.

3.1 Prog00mz als Kopierprogramm

Prog00mz kann auch als ein sehr leistungsfähiges Kopierprogramm verwendet werden. Der wichtige Unterschied zum im Abschnitt 2 dargestellten Kopierprogramm Prog00mp ist folgender:

1. Die neue Datei wird im Format FREI und DIREKT angelegt. Format FIX ist nicht verfügbar
2. Es werden - wie in obiger letzter Eingabebox - nur einige ausgewählte Variable (die eventuell auch noch ungeordnet sind) in die neue Datei übernommen. Das geschieht zwar auch bei Prog00mp. Bei Prog00mz werden dann aber die Variablen-Nummern unnummeriert, und zwar so dass sie mit v1 beginnen und dann kontinuierlich weiter nummeriert werden mit v2, 3, 4, ...
Bei Prog00mp wird nicht kontinuierlich unnummeriert. Dort muss, wenn kontinuierlich unnummeriert werden soll, anschließend Prog00mw gestartet werden.
3. Ein weiterer möglicher Vorteil von Prog00mz besteht darin, dass eine neue Datei der

Variablen-*Namen* angelegt werden kann, die nur die ausgewählten Variablen mit den neuen kontinuierlich fortlaufenden Nummern enthält.

4 Eine Datei in Unterdateien aufteilen

Betrachten wir einige Beispiele: Eine Datei von Personen wird nach dem Geschlecht in zwei Unterdateien aufgetrennt. Oder die Datei wird nach dem Beruf der Personen in die 3 Dateien Arbeitnehmer, Selbständige und Sonstige aufgetrennt. Oder eine Datei wird nach den zwei kombinierten Variablen Beruf und Geschlecht in die 6 Unterdateien aufgeteilt:

| | Arbeiter | | Selbständig | | Sonstig | |
|----------------|----------|-------|-------------|-------|---------|-------|
| | männl | weibl | männl | weibl | männl | weibl |
| Unterdatei Nr. | 1 | 2 | 3 | 4 | 5 | 6 |

Eingabe in Prog01m1 "Quelldatei in Unterdateien auftrennen"

Prog01m1.Msk
Unterdateien erzeugen

Eine Quell-Datei im Format FREI oder FIX oder DIREKT einlesen und aus ihre mehrere selbständige Unterdateien erzeugen. Diese erhalten von Almo automatisch die Dateinamen

".\Progs\Unterdatei_1.fre"
".\Progs\Unterdatei_2.fre"
.
.

Der Benutzer sollte nachträglich diese Dateien umbenennen, ihnen sinnvolle Namen geben, z. B.

".\Progs\Arbeiter.fre"
".\Progs\Selbständig.fre"
".\Progs\SonstigerBeruf.fre"

Bedingungen:

1. Die Variablen, nach denen die Unterdateien gebildet werden, sollen ganzzahlig sein; sie werden sonst automat. gerundet
2. Es können maximal die Unterdateien 1 bis 18 gebildet werden

Die Daten können durch

1. die Ein- und Ausschluss-Option sowie
2. durch die Umkodierungs- und KeinWert-Option

modifiziert in die Unterdateien geschrieben werden

Siehe Handbuch, Teil 2, Almo-Programmiersprache, Abschnitt 40

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare
Variable= 40

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Quell-Datei Hilfe
bei Datei-Problemen

".\TESTDAT\Almdez.fre"

frei Format der Daten Hilfe

V1:10 der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

Wenn Dateiformat FIX oder Nicht-Standard-FREI Hilfe

Option: Ein- und Ausschliessen von Untersuchungseinheiten

Option: Umkodierungen und Kein-Wert-Angaben

aus den Ausprägungen dieser Variablen
werden die Unterdateien gebildet

wenn 2 oder mehr Variable eingesetzt werden
dann müssen sie doch "MIT" kombiniert werden
Bei nur 1 Variablen bleibt Eingabefeld leer

Erläuterungen zu den Dialogboxen

Die Programm-Maske enthält keine Eingabebox für eine Datei von Variablenamen. Wird die Quelldatei in Unterdateien aufgetrennt, dann wird an der Aufeinanderfolge der Variablen nichts geändert. Also wird auch nichts in der Datei der Variablenamen geändert - sofern eine solche überhaupt vorhanden ist.

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Normalerweise ist dies die Nummer der letzten Variable des eingelesenen Datensatzes. Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen. Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 an

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Quell-Datei aus der gelesen wird

Der Benutzer muss den Dateiname, das Format und die einzulesenden Variablen angeben. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

Optionsbox: Wenn Dateiformat FIX oder Nicht-Standard-Frei (erweitert frei)

In unserem Beispiel enthält die Ursprungsdatei nur numerische Variable im Standard-FREI-Format, so dass die Optionsbox nicht geöffnet ist. Zu den Almo-Daten-Formaten siehe oben Abschnitt 1 oder Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.1.

Optionsbox: Ein- und Ausschließen von Untersuchungseinheiten

Siehe dazu oben Abschnitt 3 oder Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.7.1.

Eingabebox: Kein-Wert-Angabe und Umkodierungen

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.5.

In der Programm-Maske muss der Benutzer eventuelle Umkodierungen vornehmen. Wichtig ist folgende Regel:

Das Ein- und Ausschliessen von Untersuchungseinheiten erfolgt vor eventuellen Umkodierungen. Zuerst wird ein- bzw. ausgeschlossen, dann umkodiert

Eingabebox: Variable für Unterdateien

Variable für Unterdateien Hilfe

↔ V3,4

aus den Ausprägungen dieser Variablen werden die Unterdateien gebildet

↔ V3 MIT V4

wenn 2 oder mehr Variable eingesetzt werden dann müssen sie durch "MIT" kombiniert werden
Bei nur 1 Variablen bleibt Eingabefeld leer

Hilfe

In dieser Eingabebox werden die Variable eingetragen, nach denen die Quelldatei aufgetrennt wird. In unserem Beispiel sind dies die beiden Variablen V3 und V4.

Im einfachsten Fall wird man jedoch nur eine "Auftrennungs"-Variable verwenden, z.B. nur V3. Wir unterstellen, dass V3 die Variable *Berufsgruppe* mit den Ausprägungen *Arbeitnehmer*, *Selbständiger*, *Sonstiger* enthält. Almo erzeugt dann 3 Unterdateien, denen es die Namen gibt:

```
"C:\Almo\Progs\Unterdatei_1.fre"
"C:\Almo\Progs\Unterdatei_2.fre"
"C:\Almo\Progs\Unterdatei_2.fre"
```

Der Benutzer sollte nachträglich diese Dateien umbenennen, ihnen sinnvolle Namen geben, z. B.

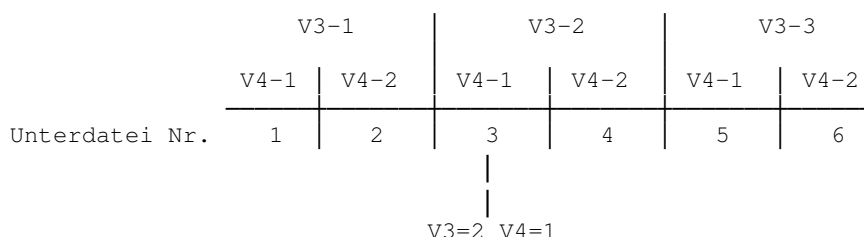
```
"C:\Almo\Progs\Arbeiter.fre"
"C:\Almo\Progs\Selbständig.fre"
"C:\Almo\Progs\SonstigerBeruf.fre"
```

Werden 2 oder mehr Variable eingesetzt, dann müssen diese (verbunden durch das Almo-Schlüsselwort "MIT") auch in das 2. Eingabefeld eingetragen werden. Bei 3 Variablen würde die Eingabe beispielsweise lauten:

V3 mit V4 mit V5

Sehr wichtig ist es, die Reihenfolge zu berücksichtigen. Es wurde oben gezeigt, welche Unterdateien entstehen, wenn eingegeben wird: V3 MIT V4

Wir wiederholen nochmals die Grafik



5. Eine Datei an eine schon bestehende Datei anhängen

```
Prog00md.Msk
Datensätze an eine bestehende Datei anhängen

Beispiel:
Die schon bestehende Datei sei "Schon_da.fre".
An sie sollen die Datensätze aus "Anhang.fre" angehängt
werden

      "Schon_da.fre"
-----
V1  V2  V3          V4  V5
Nr. Alter Beruf      Leistung  Geschlecht

  1  21  Arbeiter      1.5   m
  2  23  Angestellter  3.3   w
  3  27  Bauer          7.0   w
  4  22  Beamter       2.6   m
  5  29  Beamter       4.1   m

      "Anhang.fre"
-----
V1  V2  V3          V4  V5
Nr. Alter Beruf      Leistung  Geschlecht

  6  25  Angestellter  3.8   w
  7  31  Handwerker    0.8   w
  8  21  Beamter       7.3   m

Die neue Datei ist dann

      1  21  Arbeiter      1.5  m
      2  23  Angestellter  3.3  w
      3  27  Bauer          7.0  w
      4  22  Beamter       2.6  m
      5  29  Beamter       4.1  m
      6  25  Angestellter  3.8  w
      7  31  Handwerker    0.8  w
      8  21  Beamter       7.3  m

Voraussetzung: Die Variablen der beiden Dateien müssen
übereinstimmen. D.h. beispielsweise: V5 muss in beiden
Dateien das Geschlecht sein.
Almo kann das nicht überprüfen !!

Die beiden Dateien dürfen unterschiedliches Format besitzen
Die schon vorhandene Datei kann z.B. im Format FREI vor-
liegen und die anzuhängende Datei im Format DIREKT.


      Siehe Handbuch, Abschn. 46.4
```

Programm-Bedienung --->

Hilfe


Speicher fuer x Variable

Vereinbare Variable=


 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert


Schon vorhandene Datei

bei Datei-Problemen

 ".\TESTDAT\Schon_da.fre"

Format der Daten

 der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

 Loesche wieder diese Box

weitere Angaben zur schon vorhandenen Datei -nicht wenn Format DIREKT
löschen Sie diese Box, wenn sich die schon vorhandene Datei
im Format DIREKT befindet

Zeichenvariable in der schon vorhandenen Datei

  **V3, 5**
 **12,10**
 **Blank**

1. Eingabefeld:
Geben Sie hier die Variablen an, die
Zeichenvariable (also Worte) sind
2. Eingabefeld:
Geben Sie hier die (maximale) Länge an, mit
der die Zeichenvariablen in die neue
gemeinsame Datei geschrieben werden sollen
3. Eingabefeld:
Geben Sie hier das Trennzeichen an,
das hinter den Zeichenvariablen in der
schon vorhandenen Datei steht
- relevant nur bei Format FREI

 **Punkt**

Wenn Format FREI oder FIX

Dezimalzeichen

 **Hilfe**

Punkt= Dezimalwerte sind mit dem Punkt
als Dezimalzeichen geschrieben
Komma= Dezimalwerte sind mit dem Komma
als Dezimalzeichen geschrieben

Nur bei Format FIX

 **Hilfe**

schreiben Sie hier die Feldanweisung


  

Nur bei Format FREI

 **Hilfe**

Editfeld leer = jeder Datensatz ist vom
nachfolgenden durch einen Zeilenumbruch
getrennt
ohne_Zeilenumbruch = mehrere Datensätze
stehen hintereinander in einer Zeile

Anzuhängende Datei

 **"..\TESTDAT\Anhang.fre"**


 **frei**

Format der Daten

 **Hilfe**

  **V1:5**

der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

 Loesche wieder diese Box

weitere Angaben zur Anhänge-Datei - nicht wenn Format DIREKT
löschen Sie diese Box, wenn sich die Anhänge-Datei
im Format DIREKT befindet

Zeichenvariable in der Anhänge-Datei

    v3, 5
  12,10
  Blank

1. Eingabefeld:

Geben Sie hier die Variablen an, die
Zeichenvariable (also Worte) sind

2. Eingabefeld:

Geben Sie hier die (maximale) Länge an, mit
der die Zeichenvariablen in die neue
gemeinsame Datei geschrieben werden sollen

3. Eingabefeld:

Geben Sie hier das Trennzeichen an,
das hinter den Zeichenvariablen in der
Anhänge-Datei steht
- relevant nur bei Format FREI

  Punkt

Wenn Format FREI oder FIX

Dezimalzeichen



Punkt= Dezimalwerte sind mit dem Punkt
als Dezimalzeichen geschrieben
Komma= Dezimalwerte sind mit dem Komma
als Dezimalzeichen geschrieben

Nur bei Format FIX



schreiben Sie hier die Feldanweisung


Nur bei Format FREI



Editfeld leer = jeder Datensatz ist vom
nachfolgenden durch einen Zeilenumbruch
getrennt

ohne_Zeilenumbruch = mehrere Datensätze
stehen hintereinander in einer Zeile


Neue Datei

 ".\Progs\NeuDat"

Geben Sie einen neuen Dateinamen
ohne Erweiterung an
Almo erzeugt 2 Dateien:

1. eine nicht lesbare Almo-Arbeitsdatei
mit der Erweiterung ____.dir
2. eine anschauliche Datei im freien Format
mit der Erweiterung ____.fre

Optionen

 1
8

Zur Kontrolle die Datensätze

vom x. Datensatz
bis zum y. Datensatz

..... ausgeben.

wenn 2. Eingabefeld 0 oder leer
dann keine Kontroll-Ausgabe

Die Erläuterungen und Hilfen in den Eingabeboxen sind so ausführlich, dass wir uns weitere Erläuterungen ersparen können.

6. Parallele Dateien zusammenfügen

```

                                Prog00me.Msk
                                2 parallele Dateien zusammenfügen

Dateien sind 'parallel', wenn die Reihenfolge der Datensätze
in beiden Dateien dieselbe ist.

Beispiel:

      Stamm-Datei                      Anfüge-Datei

Die 1. Datei sei                      Hinzugefügt werden soll
"Stamm.fre"                            "Anfugdat.fre"

Hofer  31 ja  21  1.5      156 Hofer  3.3  122.5  männl
Taler  32 nein 22  kw      356 Taler  3.2    kw  weibl
Mayer  33 nein 19  4       456 Mayer  3.4  422.1  weibl
Maier  34 ja  20  5       556 Maier  3.1  522.1  männl
Müller 35 ja  31  2.5     222 Müller 2.1  321.7  männl
Maler  36 ja  22  2       121 Maler  3.0  432.8  weibl
Berg   37 nein 21  6       345 Berg   1.7  112.3  weibl
Berger 38 ja  23  7       233 Berger 4.1  231.7  männl
Tal    39 ja  34  8       111 Tal    5.8  435.0  männl
Meier  40 nein 22  9       643 Meier  1.7   23.8  weibl
Friedl 41 ja  22  10      kw Friedl 1.0  147.5  weibl

kw = das ist Code für Kein-Wert

V1 in "Stamm.fre" ist der Personen-Name.
V2 in "Anfugdat.fre" ist ebenfalls der Personen-Name
Die Reihenfolge der Personen in beiden Dateien ist also
offensichtlich dieselbe

Nach dem Zusammenfügen ist die neue Datei "Zusamm.fre"
folgende:

Hofer  31  ja  21  1.5  156 Hofer  3.3  122.5  männl
Taler  32 nein 22  kw  356 Taler  3.2    kw  weibl
Mayer  33 nein 19  4   456 Mayer  3.4  422.1  weibl
Maier  34  ja  20  5   556 Maier  3.1  522.1  männl
Müller 35  ja  31  2.5 222 Müller 2.1  321.7  männl
Maler  36  ja  22  2   121 Maler  3.0  432.8  weibl
Berg   37 nein 21  6   345 Berg   1.7  112.3  weibl
Berger 38  ja  23  7   233 Berger 4.1  231.7  männl
Tal    39  ja  34  8   111 Tal    5.8  435.0  männl
Meier  40 nein 22  9   643 Meier  1.7   23.8  weibl
Friedl 41  ja  22  10  kw Friedl 1.0  147.5  weibl

Die zusammengefügte Datei umfasst nunmehr 10 Variable. Dabei
sind          V1:5 aus der Stamm-Datei
              V6:10 aus der Anfüge-Datei.

                                Siehe Handbuch, Abschn. 46.6
```

Speicher fuer x Variable
(BEACHTTE: Um mindestens 1 höher vereinbaren als notwendig)
Vereinbare Variable=


Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Stamm-Datei

Format der Daten

der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

letzte Variablen aus Stamm-Datei
hinter der angefügt werden soll

 Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

weitere Angaben zur Stamm-Datei - nur wenn Format FREI

löschen Sie diese Box, wenn sich die Stamm-Datei
im Format DIREKT befindet

Zeichenvariable in der Stamm-Datei

     **V1, 3**
  **12,10**
 **Blank**

1. Eingabefeld:

Geben Sie hier die Variablen an, die
Zeichenvariable (also Worte) sind

2. Eingabefeld:

Geben Sie hier die (maximale) Länge an, mit
der die Zeichenvariablen in die neue
gemeinsame Datei geschrieben werden sollen

3. Eingabefeld:

Geben Sie hier das Trennzeichen an,
das hinter den Zeichenvariablen in der
Stamm-Datei steht
- relevant nur bei Format FREI

 **Punkt**

Dezimalzeichen

 **Hilfe**

Punkt= Dezimalwerte sind mit dem Punkt
als Dezimalzeichen geschrieben

Komma= Dezimalwerte sind mit dem Komma
als Dezimalzeichen geschrieben


   

 **Hilfe**

Eingabefeld leer = jeder Datensatz ist vom
nachfolgenden durch einen Zeilenumbruch
getrennt

ohne_Zeilenumbruch = mehrere Datensätze
stehen hintereinander in einer Zeile

Anfüge-Datei

 **".\TESTDAT\Anfugdat.fre"**

 **frei**


Format der Daten

 **Hilfe**

  **V1:5**

der Datensatz enthält diese Variablen

Bei Format DIREKT schreiben Sie: alle_V

 Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)

 weitere Angaben zur Anfüge-Datei - nur wenn Format FREI

löschen Sie diese Box, wenn sich die Anfüge-Datei
im Format DIREKT befindet

Zeichenvariable in der Anfüge-Datei



1. Eingabefeld:

Geben Sie hier die Variablen an, die
Zeichenvariable (also Worte) sind

2. Eingabefeld:

Geben Sie hier die (maximale) Länge an, mit
der die Zeichenvariablen in die neue
gemeinsame Datei geschrieben werden sollen

3. Eingabefeld:

Geben Sie hier das Trennzeichen an,
das hinter den Zeichenvariablen in der
Anfüge-Datei steht
- relevant nur bei Format FREI



Dezimalzeichen



Punkt= Dezimalwerte sind mit dem Punkt
als Dezimalzeichen geschrieben

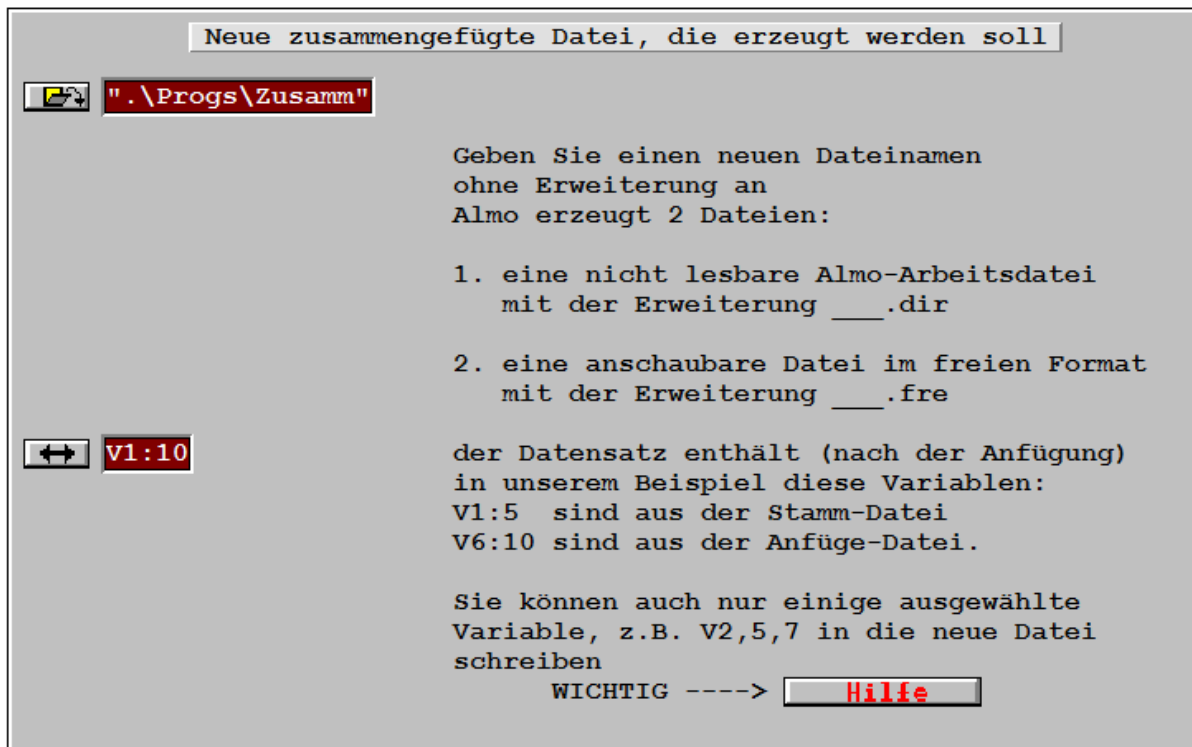
Komma= Dezimalwerte sind mit dem Komma
als Dezimalzeichen geschrieben





Eingabefeld leer = jeder Datensatz ist vom
nachfolgenden durch einen Zeilenumbruch
getrennt

ohne Zeilenumbruch = mehrere Datensätze
stehen hintereinander in einer Zeile



Die Erläuterungen und Hilfen in den Eingabefeldern sind so ausführlich, dass wir uns weitere Erläuterungen ersparen können.

7. Dateien über eine Verbindungsvariable zusammenfügen

7.1 Ein Beispiel:

Wir haben eine Datei mit Personen. Die Datei umfasse 1000 Personen. Für jede Person sind 50 Variable vorhanden. Daneben haben wir eine 2. Datei, in der für verschiedene Gemeinden bestimmte Strukturmerkmale angegeben sind; beispielsweise:

Zahl der landwirtschaftlich Tätigen
Zahl der Arbeiter,
Prozentzahl der Wähler der Partei A etc.

Die Datei umfasse 20 Gemeinden mit je 10 Strukturmerkmalen.

Diese Strukturmerkmale sollen nun jeder Person aus der Personen-Datei entsprechend ihrer Gemeindezugehörigkeit als zusätzliche Variable angefügt werden. Möglich wird dies dadurch, dass sowohl in der Personen-Datei als auch in der Gemeinde-Datei die Gemeinde-Nummer als Variable enthalten ist.

In unserer Terminologie bezeichnen wir die Personen-Datei als "*Stamm-Datei*" und die Gemeinde-Datei als "*Anfüge-Datei*". Die Gemeinde-Nummer dient als "*Verbindungsvariable*". Almo erzeugt eine neue Datei, eine gemeinsame Datei, die aus der Personen-Datei und den angefügten Variablen besteht.

WICHTIG:

1. Gehört eine Person einer Gemeinde an, die nicht in der Gemeinde-Datei enthalten ist, dann werden für diese Person zwar auch die Strukturmerkmale als Variable angefügt - jedoch auf "Kein-Wert" gesetzt.
2. Ist in der Gemeindedatei eine Gemeinde enthalten, die nicht gebraucht wird, weil keine

Person aus der Personendatei in ihr wohnt, dann geht sie nicht in die gemeinsame Datei ein. Es gilt also: Es wird nur angefügt was für der Stammdatei gebraucht wird. Also verfügt auch über Programm-Masken, die eine vollständige "Verschmelzung" zweier Dateien bewirken. Das wird in Abschnitt 46.7 vorgetragen.

7.2 Ein 2. Beispiel:

In der nachfolgend abgebildeten Programm Maske Prog00mf wird folgendes Beispiel verwendet:

| Stammdatei | | | | | | | | | | Anfügedatei | | | | | |
|------------|---|---------------------|---|---|---|---|---|---|----|---------------------|---|---|---|---|----|
| 2 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 16 |
| 1 | 6 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | kw | 3 | 1 | 1 | 1 | 1 | 5 |
| 3 | 1 | 3 | 1 | 1 | 1 | 0 | 1 | 3 | 3 | 4 | 1 | 1 | 1 | 2 | 4 |
| kw | 2 | 4 | 1 | 1 | 1 | 1 | 0 | 1 | 3 | 8 | 1 | 1 | 1 | 2 | 2 |
| 1 | 8 | 5 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 9 | 1 | 1 | 2 | 1 | kw |
| 1 | 9 | 6 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 10 | 1 | 1 | 2 | 1 | 4 |
| 3 | 3 | 7 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 11 | 1 | 1 | 2 | 2 | 2 |
| 4 | 1 | 8 | 1 | 0 | 1 | 1 | 1 | 3 | 1 | 5 | 1 | 1 | 2 | 2 | 4 |
| 2 | 7 | 9 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 6 | 1 | 1 | 3 | 1 | 3 |
| 2 | 5 | 10 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 7 | 1 | 2 | 3 | 1 | 5 |
| 8 | 4 | 11 | 0 | 1 | 1 | 1 | 0 | 1 | 3 | 12 | 1 | 1 | 3 | 2 | 1 |
| 5 | 2 | 12 | 1 | 0 | 0 | 1 | 1 | 3 | 3 | 13 | 1 | 1 | 3 | 2 | 2 |
| 3 | 7 | 13 | 1 | 1 | 0 | 1 | 0 | 3 | 3 | 2 | 1 | 2 | 1 | 1 | 16 |
| 1 | 9 | 14 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | | | | | | |
| 2 | 1 | 15 | 1 | 1 | 1 | 0 | 0 | 3 | 3 | | | | | | |
| | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |
| | | V3 | | | | | | | | V1 | | | | | |
| | | Verbindungsvariable | | | | | | | | Verbindungsvariable | | | | | |

V3 in der Stammdatei ist die Personen-Nr.

V1 in der Anfügedatei ist ebenfalls die Personen-Nr.

Die beiden Dateien können also über V3 bzw. V1 als **Verbindungsvariable** zusammengefügt werden.

BEACHTEN:

Für die Personen-Nr. 14 und 15 sind in der Anfügedatei keine Daten vorhanden. In der neuen gemeinsamen Datei sind dementsprechend die angefügten Variablen für diese Personen auf Kein-Wert "kw" gesetzt.

Nachdem die Programm-Maske Prog00mf gerechnet wurde liegt folgende neue *gemeinsame Datei* vor:

| neue gemeinsame Datei | | | | | | | | | | | | | | | |
|-----------------------|---|----|---|---|---|---|---|---|----|----|---|---|---|---|----|
| 2 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 16 |
| 1 | 6 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | kw | 2 | 1 | 2 | 1 | 1 | 16 |
| 3 | 1 | 3 | 1 | 1 | 1 | 0 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 1 | 5 |
| kw | 2 | 4 | 1 | 1 | 1 | 1 | 0 | 1 | 3 | 4 | 1 | 1 | 1 | 2 | 4 |
| 1 | 8 | 5 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 5 | 1 | 1 | 2 | 2 | 4 |
| 1 | 9 | 6 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 6 | 1 | 1 | 3 | 1 | 3 |
| 3 | 3 | 7 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 7 | 1 | 2 | 3 | 1 | 5 |
| 4 | 1 | 8 | 1 | 0 | 1 | 1 | 1 | 3 | 1 | 8 | 1 | 1 | 1 | 2 | 2 |
| 2 | 7 | 9 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 9 | 1 | 1 | 2 | 1 | kw |
| 2 | 5 | 10 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 10 | 1 | 1 | 2 | 1 | 4 |
| 8 | 4 | 11 | 0 | 1 | 1 | 1 | 0 | 1 | 3 | 11 | 1 | 1 | 2 | 2 | 2 |

```

5  2  12  1  0  0  1  1  3  3    12  1  1  3  2  1
3  7  13  1  1  0  1  0  3  3    13  1  1  3  2  2
1  9  14  1  1  1  1  1  2  1    kw  kw  kw  kw  kw  kw
2  1  15  1  1  1  0  0  3  3    kw  kw  kw  kw  kw  kw

|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
V1 V2  V3 V4  ..... V10  V11 V12  ..... V16

```

Die Variablen V1 bis V10 entstammen der Stammdatei. Die Variablen V11 bis V16 wurden aus der Anfügedatei übernommen.

Das Programm fügt auch die beiden Dateien der Variablen-Namen zusammen. In unserem Beispiel sind die Namensdateien der Stamm- und der Anfügedatei folgende

Variablenamen Stammdatei

```

Name 1=Stamm_Einkommen;
Name 2=Stamm_Kinderzahl;
Name 3=Stamm_Person;
Name 4=Stamm_Item1:(0)klein,(1)gross;
Name 5=Stamm_Item2;
Name 6=Stamm_Item3;
Name 7=Stamm_Item7;
Name 8=Stamm_Item8;
Name 9=Stamm_AB:aa,bb,cc;
Name10=Stamm_AC;

```

Variablenamen Anfügedatei

```

Name 1=Anf_Person;
Name 2=Anf_Geschlecht:männlich,weiblich;
Name 3=Anf_Wohnort:Stadt,Land;
Name 4=Anf_Beruf:Arbeit,Ang,Beamt;
Name 5=Anf_Schulbildung:niedrig,hoch;
Name 6=Anf_Leistung;

```

Die beiden Namensdateien werden zur *neuen gemeinsamen* Namensdatei zusammengefügt

```

N1=Stamm_Einkommen;
N2=Stamm_Kinderzahl;
N3=Stamm_Person;
N4=Stamm_Item1:(0)klein,(1)gross;
N5=Stamm_Item2;
N6=Stamm_Item3;
N7=Stamm_Item7;
N8=Stamm_Item8;
N9=Stamm_AB:aa,bb,cc;
N10=Stamm_AC;
N11=Anf_Person;
N12=Anf_Geschlecht:männlich,weiblich;
N13=Anf_Wohnort:Stadt,Land;
N14=Anf_Beruf:Arbeit,Ang,Beamt;
N15=Anf_Schulbildung:niedrig,hoch;
N16=Anf_Leistung;

```

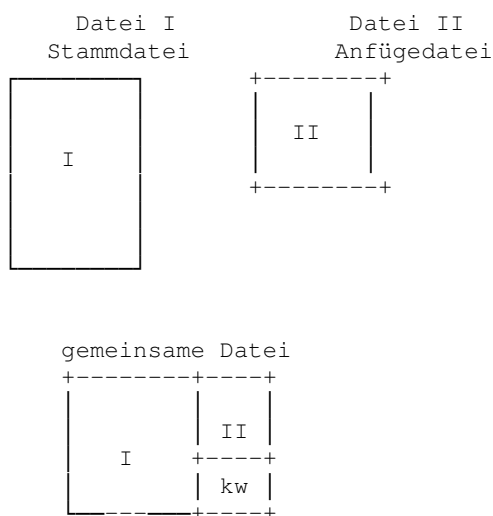
Hier kann ein Problem auftreten. In der Stammdatei kann ein Variablenamen enthalten sein, der auch in der Anfügedatei verwendet wird. Nach dem Zusammenfügen sind dann zwei Variable vorhanden, die denselben Variablenamen besitzen. Das ist nicht zulässig.

In unserem Beispiel würde dieser Fall auftreten, wenn die Personen-Nummer, die als *Verbindungsvariable* verwendet wird, in beiden Dateien den Namen "Person" besitzen würde. Wir haben dies dadurch vermieden, dass wir diese Variable einmal "Stamm_Person" und einmal "Anf_Person" genannt haben.

Almo überprüft in den 4 Programmen des Zusammenfügens jedoch nicht, ob dieser Fall aufgetreten ist. Das muss der Benutzer nachträglich mit dem Programm *PruefNamen.Alm* tun. Das Programm findet man durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Hauptfensters. Bei Ausprägungsnamen stellt sich dieses Problem nicht. Diese dürfen bei verschiedenen Variablen gleich sein.

7.3 Regeln für das Zusammenfügen

Beim Zusammenfügen über eine Verbindungsvariable entsteht eine "gemeinsame Datei" dadurch, dass zwei Dateien in einer gewissen Weise hintereinander gestellt werden. Datei I nennen wir "Stammdatei" und Datei II "Anfügedatei". Bildlich dargestellt:



Für das Zusammenfügen zweier Dateien gelten folgende Regeln

1. Bestehen Stamm- und Anfügedatei aus numerischen Variablen dann wird Prog00mf oder Prog00mg verwendet. Besitzt eine oder besitzen alle beiden auch Zeichenvariable (Worte), dann muss Prog00mu oder Prog00mv verwendet werden
2. Die Daten der Stammdatei und der Anfügedatei sollten im Format FREI vorliegen. Format FIX ist nicht erlaubt. Bei Format DIREKT können Probleme auftreten. Siehe dazu nachfolgenden Abschnitt.
3. Die beiden Dateien müssen eine "Verbindungs-Variable" besitzen. Das kann z.B. eine Personen-Nummer in beiden Dateien sein. Siehe das Beispiel
4. Die "Verbindungsvariable" kann eine numerische Variable oder eine Zeichenvariable sein.
5. Für einen Datensatz in der "Stammdatei" sucht das Programm einen "Bruder" (oder eine Schwester) in der "Anfügedatei". Er (oder sie) wird gefunden, wenn die "Verbindungsvariable" für beide denselben Wert besitzt. Die Suche wird dann beendet.
6. Befindet sich in der Anfügedatei ein weiterer Bruder, dann wird dieser nicht berücksichtigt. D.h. sind in der Anfügedatei mehrere Datensätze mit demselben Wert in der Verbindungs-

variablen, dann wird in der Reihenfolge der erste zum "Bruder". Die nachfolgenden werden nicht berücksichtigt. Sie gehen nicht in die "gemeinsame Datei" ein.

7. Befinden sich in der Anfügedatei Datensätze, die nicht als Bruder gebraucht werden (deren Verbindungsvariable einen Wert besitzt, der nicht in der Stammdatei vorkommt), dann gehen sie verloren. Sie erscheinen nicht in der "gemeinsamen Datei".

8. Im umgekehrten Fall gilt dies jedoch nicht: Wird zu einem Datensatz der Stammdatei kein Bruder in der Anfügedatei gefunden, dann geht dieser Datensatz trotzdem in die gemeinsame Datei ein. Die anzufügenden Variablen werden für ihn auf kw (Kein-Wert) gesetzt.

Prinzip:

Die Stammdatei sucht in der Anfügedatei nach Daten und schöpft diese ab.

Anmerkung:

Almo verfügt auch über Programm-Masken, die eine vollständige "Verschmelzung" zweier Dateien bewirken. Im nachfolgenden Abschnitt wird das "Verschmelzen" behandelt, Siehe dazu die Programm-Masken Prog00mx und Prog00my, sowie das Syntax-Programm "Datban11.Alm". Zu finden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters.

9. Die Anfügedatei wird durch das Zusammenfügen nicht verändert oder sogar gelöscht. Nach dem Anfügen ist sie unbeschädigt noch vorhanden; ebenso die Stammdatei.

10. Die gemeinsame Datei kann auch nur aus einigen ausgewählten Variablen der Stamm- und der Anfügedatei gebildet werden. Dafür sollten die Programm-Masken Prog00mg oder Prog00mv verwendet werden.

11. Zusammengefügt werden nicht nur die beiden Daten-Dateien sondern auch, sofern vorhanden, die beiden Namens-Dateien (d.h. die Dateien der Variablen- und Ausprägungsnamen)

12. Sind keine Namens-Dateien vorhanden, dann müssen selbstverständlich in der Programm-Maske die beiden entsprechenden Eingabefelder leer gemacht werden.

13. Ist nur eine Namens-Datei vorhanden, so geht nur diese in die neue Datei ein

7.4 Zulässige Datenformate für das Zusammenfügen

Format FREI

Liegen die Daten der Stamm- und Anfügedatei im Format FREI vor, dann können keine Probleme auftreten. Sind Zeichenvariable (z.B. Worte) vorhanden, dann muss der Benutzer die sandfarbene Optionsbox (nach der Box "Stamm-Datei" bzw. "Anfüge"-Datei) öffnen und die Zeichenvariablen angeben.

Format FIX

Dieses Format ist nicht zulässig.

Format DIREKT:

Sind die Daten in der Stamm- und der Anfügedatei so nummeriert, dass sie mit V1 beginnen und dann kontinuierlich aufeinander folgen - dann dürfen die Stamm- und/oder Anfüge-Datei

das Format DIREKT besitzen.

Hingegen: Die Daten der Stammdatei und der Anfügedatei dürfen nicht im Format DIREKT vorliegen - sofern die Variablen-Nummern nicht mit V1 beginnen und kontinuierlich aufeinander folgen.

Almo bringt in diesem Fall eine Warnung, bricht aber nicht ab. Dem Benutzer wird es dadurch ermöglicht, in der entstandenen Datei nachzuschauen, welche Variable an welche Stelle in der Aufeinanderfolge verschoben wurde. Die Warnung ist folgende:

```
***** MITTEILUNG
Folgende Variable werden als
ALLE_V aus Datei 1 'C:\Almo15\Progs\Stammdat3.dir' gelesen
V1 9 2:8 10
***** WARNUNG
BEACHTTE: Die Variablen-Nummern folgen nicht kontinuierlich aufeinander ??????????
*****
```

Beim Format DIREKT (aber nur bei diesem) können Variablen-Nummer und Stelle in der Aufeinanderfolge der Variablen nicht übereinstimmen. Es ist möglich, dass z.B. folgende Zuordnung besteht:

```
Aufeinanderfolge: 1 2 3 4 5 6 7
Variablennummer : V3 V1 V2 V4 V5 V7 V6
```

Das ist zulässig. Deswegen bringt Almo auch keine Fehlermeldung wenn dieser Fall auftritt. Reihenfolge und Variablen-Nummer können eigentlich nur auseinander fallen, wenn der Benutzer beim Erzeugen der Direktdatei die Variablen umgeordnet hat, z.B. mit folgenden Anweisungen in der Almo-Programmiersprache

```
Lese V1:7 .....
in Format FREI

Schreibe V3,1,2,4,5,7,6
in Format DIREKT
```

Oder wenn im Verlauf eines Almo-Programms die Variablen um-nummeriert und neu gespeichert wurden.

Wird eine so erzeugte Datei in einem Almo-Programm (mit der Angabe "alle_V") eingelesen, dann wird korrekt (in unserem Beispiel) der 1. eingelesene Wert an die Variable V3 zugewiesen usw. Es entsteht also kein Fehler. Beim Direkt-Format merkt sich Almo zu jedem Variablenwert dessen Variablen-Nummer.

Will der Benutzer eine derartige Direktdatei umordnen, so dass sie mit V1 beginnend fortlaufend nummeriert ist, dann kann er dies mit dem eigens dafür geschaffenen Prog00mw tun. Das Programm wird gefunden durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Das Programm ordnet automatisch die Variablen nach ihrer Variablennummer hintereinander an. Reihenfolge und Nummer stimmen dann wieder überein.

7.5 So funktioniert das Zusammenfügen

Beim Zusammenfügen über eine Verbindungsvariable entsteht eine "gemeinsame Datei" dadurch, dass zwei Dateien in einer gewissen Weise hintereinander gestellt werden. Die gemeinsame Datei kann aber auch nur aus einigen ausgewählten Variablen der Stamm- und der Anfügedatei gebildet werden. Der Vorgang des Zusammenfügens ist, an einem Beispiel gezeigt, folgender:

a. Ein Datensatz aus der Stammdatei kann bildlich so veranschaulicht werden

```
|-----|
V1          ...          V10
```

BEACHTTE: Was wie hier mit V1, V2, V.. bezeichnen kann eine numerische Variable aber auch eine Zeichenvariable sein

Ein Datensatz aus der Anfügedatei ist beispielsweise

```
|-----|---|---|
V1          ...          V5 V6
```

Wir wollen beobachten, was mit einer Variable der Anfügedatei, z.B. V5 (dies sei z.B. die Variable der Schulbildung) im Verlauf des Zusammenfügens geschieht.

b. Nachdem in der Anfügedatei der passende Datensatz gefunden wurde wird dieser zunächst an das Ende seines "Bruders" aus der Stammdatei gestellt

```
          Datensatz aus Stammdatei          | Datensatz aus Anfügedatei
|-----|-----|-----|-----|
V1          ...          V10 V11          ...          V15 V16
```

Ein Datensatz der Stammdatei umfasst die Variablen V1 bis V10 Ein Datensatz der Anfügedatei umfasst die Variablen V1 bis V6. Letztere werden nach dem Anfügen zu V11 bis V16.

Aus der "unter unserer Beobachtung stehenden" Variablen V5 (der Schulbildung) ist jetzt V15 geworden.

Auch die Variablen-Namen der Anfügedatei (sofern vorhanden), die ursprünglich die Namens-Nummern Name1 bis Name5 hatten werden um-nummeriert auf Name11 bis Name15.

Im Programm Prog00mf ist damit das Zusammenfügen beendet. Das gilt auch für Prog00mu bei dem numerische und Zeichenvariable zusammengefügt werden.

Dieser verlängerte Datensatz kann nun im Programm Prog00mg (mit Optionen) in die neue "gemeinsame Daten-Datei" eingegeben werde. In der letzten Eingabebox vom Prog00mg im letzten Eingabefeld ist dann einzusetzen:

V1:10,11:16 oder kurz: V1:16

7.5.1 Umgeordneter Datensatz bei Prog00mg und Prog00mv

Der Benutzer kann aber auch aus dem neuen verlängerten Datensatz beliebige Variable herausgreifen und in die neue "gemeinsame Datei" geben. Unsere folgenden Ausführungen gelten entsprechend auch für Prog00mv (für numerische und Zeichenvariable).

Z.B soll V15 (das ehemalige V5 - Schulbildung) nach vorne an die 2. Stelle im Datensatz vorgerückt werden. Dann schreibt der Benutzer in Prog00mg in der letzten Eingabebox im letzten Eingabefeld

V1, 15, 2:14,16

Almo ordnet den Datensatz dann in folgender Weise um

```
|---|---|-----|-----|-----|---|
V1 V15 V2          ...          V10 V11          ..          V14 V16
```

Der Benutzer kann auch nur einige ausgewählte Variable in die gemeinsame Datei geben. Er

will z.B. V11 und V13 nicht in die gemeinsame Datei übernehmen. Der Datensatz ist dann folgender:

```
|--|---|-----|-----|---|---|---|
V1 V15 V2          ...          V10 V12 V14 V16
```

Die gemeinsame Datei umfasst jetzt nur noch 14 und nicht mehr 16 Variable

Die Daten werden nun in dieser Reihenfolge in die gemeinsame Daten-Datei im Format FREI und im Format DIREKT geschrieben. Dabei werden die Variablen kontinuierlich um-nummeriert auf V1, V2, V3, ... fortlaufend bis V14. Der Datensatz ist also folgender:

```
|--|---|-----|-----|---|---|---|
V1 V15 V2          ...          V10 V12 V14 V16  <--- seitherige Var.-Nummern

V1  V2 V3          ...          V11 V12 V13 V14  <--- neue Var.-Nummern
```

Die "unter unserer Beobachtung stehenden" ursprüngliche Variablen V5 (Schulbildung) ist nun schlussendlich zu V2 geworden.

Variablen- und Ausprägungs-Namen

Dasselbe gilt auch für den Variablen_Namen. Diese werden auch neu durchnummeriert
Für das ehemalige V5 war der Name: Name 5 = Schulbildung:niedrig, hoch;
Jetzt ist er Name 2 = Schulbildung:niedrig, hoch;
Mit dieser neuen Namens-Nummer geht er in die neue "gemeinsame Namens-Datei" ein

Folgendes Problem kann entstehen: In der Stammdatei kann ein Variablenamen enthalten sein, der auch in der Anfügedatei verwendet wird. Nach dem Zusammenfügen sind dann zwei Variable vorhanden, die denselben Variablenamen besitzen. Das ist nicht zulässig. Almo überprüft in den 4 Programmen des Zusammenfügens jedoch nicht, ob dieser Fall aufgetreten ist. Das muss der Benutzer nachträglich mit dem Programm "PruefNamen.Alm" tun. Das Programm findet man durch Klick auf den Knopf "alle Progs" am Oberrand des Almo-Hauptfensters. Bei Ausprägungsnamen stellt sich dieses Problem nicht. Diese dürfen bei verschiedenen Variablen gleich sein.

Neue Datei lesen

Soll die gemeinsame Datei anschliessend in einem anderen Programm eingelesen, werden, dann muss der Benutzer sie

1. im Format FREI in der Programm-Maske
angegeben mit den Variablen: V1:14
2. im Format Direkt muss er sie in der Programm-Maske
angegeben mit der Bezeichnung: alle_V
wobei dann Almo die Variablen eine nach der anderen
als V1 bis V14 einliest
und der Benutzer mit diesen Nummern auf die Variablen
zugreifen kann

Was hier kompliziert erscheint, ist tatsächlich für den Benutzer sehr einfach: Almo liefert ihm eine Daten-Datei und eine genau dazu passende Datei der Variablen-Namen. Wenn der Benutzer den Platz einer Variablen im Datensatz nicht kennt, dann muss er nur in der Namens-Datei nach dieser Variablen nachschauen.

Wenn jedoch keine Namensdatei existiert und der Benutzer, wie hier vorgeführt, ausgewählte

und umgeordnete Variable in die gemeinsame Datei schreibt, dann kann es sehr wohl kompliziert werden, wenn er eine bestimmte Variable sucht.

7.6 Eingabe in Prog00mf

Für das Zusammenfügen zweier Dateien bietet Almo die folgenden 4 Masken-Programme an:

- Prog00mf für numerische Variable
- Prog00mg für numerische Variable mit Umordnung des Datensatzes
- Prog00mu für numerische und Zeichen-Variable
- Prog00mv für numerische und Zeichen-Variable mit Umordnung des Datensatzes

Die Programm-Masken unterscheiden sich nur geringfügig. Es soll deswegen nur Prog00mf abgebildet und erläutert werden.

Prog00mf.Msk

"einfaches" Zusammenfügen
von 2 Dateien über eine Verbindungsvariable
Die Dateien enthalten nur numerische Variable

ein Beispiel

Regeln für das Zusammenfügens

Ablauf des Zusammenfügens

genauer:
An eine Stamm-Datei soll eine Anfüge-Datei angefügt werden

Bedingungen

1. Die Dateien enthalten nur numerische Variable
2. Die Daten der Stamm- und Anfüge-Datei müssen im Format FREI oder DIREKT vorliegen

Format FIX ist nicht erlaubt
|
lesen

Beispiel:

| | |
|--|--|
| <p>Die 1.Datei sei "Stammdat2.fre"</p> <pre> 2 4 1 1 1 1 1 1 2 3 1 6 2 1 1 1 1 0 1 kw 3 1 3 1 1 1 0 1 3 3 kw 2 4 1 1 1 1 0 1 3 1 8 5 1 1 1 1 1 3 3 1 9 6 1 1 1 1 1 3 1 3 3 7 1 1 1 1 1 1 3 4 1 8 1 0 1 1 1 3 1 2 7 9 1 1 1 1 1 3 1 2 5 10 1 1 1 1 1 3 1 8 4 11 0 1 1 1 0 1 3 5 2 12 1 0 0 1 1 3 3 3 7 13 1 1 0 1 0 3 3 1 9 14 1 1 1 1 1 2 1 2 1 15 1 1 1 0 0 3 3 </pre> | <p>Sie soll zusammengefügt werden mit der Datei "Anfuegdat2.fre"</p> <pre> 1 1 2 1 1 16 3 1 1 1 1 5 4 1 1 1 2 4 8 1 1 1 2 2 9 1 1 2 1 kw 10 1 1 2 1 4 11 1 1 2 2 2 5 1 1 2 2 4 6 1 1 3 1 3 7 1 2 3 1 5 12 1 1 3 2 1 13 1 1 3 2 2 2 1 2 1 1 16 </pre> |
|--|--|

V3 in "Stammdat2.fre" ist die Personen-Nr.
V1 in "Anfuegdat2.fre" ist ebenfalls die Personen-Nr.

Die beiden Dateien koennen also über V3 bzw. V1 als Verbindungsvariable zusammengefügt werden.

BEACHTE:

Für die Personen-Nr. 14 und 15 sind in "Anfuegdat2.fre"
keine Daten vorhanden. In der neuen gemeinsamen Datei
sind dementsprechend die angefügten Variablen für
diese Personen auf Kein_Wert gesetzt.

Die neue Datei "Zusammen.fre" ist folgende

| | | | | | | | | | | | | | | | |
|----|---|----|---|---|---|---|---|---|----|----|----|----|----|----|----|
| 2 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 16 |
| 1 | 6 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | kw | 2 | 1 | 2 | 1 | 1 | 16 |
| 3 | 1 | 3 | 1 | 1 | 1 | 0 | 1 | 3 | 3 | 3 | 1 | 1 | 1 | 1 | 5 |
| kw | 2 | 4 | 1 | 1 | 1 | 1 | 0 | 1 | 3 | 4 | 1 | 1 | 1 | 2 | 4 |
| 1 | 8 | 5 | 1 | 1 | 1 | 1 | 1 | 3 | 3 | 5 | 1 | 1 | 2 | 2 | 4 |
| 1 | 9 | 6 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 6 | 1 | 1 | 3 | 1 | 3 |
| 3 | 3 | 7 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 7 | 1 | 2 | 3 | 1 | 5 |
| 4 | 1 | 8 | 1 | 0 | 1 | 1 | 1 | 3 | 1 | 8 | 1 | 1 | 1 | 2 | 2 |
| 2 | 7 | 9 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 9 | 1 | 1 | 2 | 1 | kw |
| 2 | 5 | 10 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 10 | 1 | 1 | 2 | 1 | 4 |
| 8 | 4 | 11 | 0 | 1 | 1 | 1 | 0 | 1 | 3 | 11 | 1 | 1 | 2 | 2 | 2 |
| 5 | 2 | 12 | 1 | 0 | 0 | 1 | 1 | 3 | 3 | 12 | 1 | 1 | 3 | 2 | 1 |
| 3 | 7 | 13 | 1 | 1 | 0 | 1 | 0 | 3 | 3 | 13 | 1 | 1 | 3 | 2 | 2 |
| 1 | 9 | 14 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | kw | kw | kw | kw | kw | kw |
| 2 | 1 | 15 | 1 | 1 | 1 | 0 | 0 | 3 | 3 | kw | kw | kw | kw | kw | kw |

| | | | | | | | | | | | | | | |
V1 V2 V3 V4 V10 V11 V12 V16

Programm-Bedienung ---> [Hilfe](#)

[Speicher fuer x Variable](#) [Hilfe](#)

Vereinbare Variable= **200**

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Stamm-Datei

"C:\Almo15\TESTDAT\STAMMDAT2.FRE"

frei Format der Daten [Hilfe](#)


V1:10 der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V



V3 Verbindungsvariable aus Stamm-Datei



KeinWert-Angabe für Verbindungsvariable
[Hilfe](#)
[Hilfe](#)


weitere Angaben zur Stamm-Datei - nicht wenn Format DIREKT




Anfüge-Datei



 "C:\Almo15\TESTDAT\Anfuegdat2.fre"


 frei Format der Daten  Hilfe


  V1:6 der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_v



 v1 Verbindungsvariable aus Anfüge-Datei

   KeinWert-Angabe für Verbindungsvariable


 Hilfe
 Hilfe



 weitere Angaben zur Anfüge-Datei - nicht wenn Format DIREKT

Datei der Variablennamen der Stamm-Datei  Hilfe

  ".\TESTDAT\Stammdat2.nam"


Leeren Sie das Eingabefeld wenn keine Namensdatei vorhanden ist

Datei der Variablennamen der Anfüge-Datei  Hilfe

  ".\TESTDAT\ANFUEGdat2.NAM"

Leeren Sie das Eingabefeld wenn keine Namensdatei vorhanden ist

Neue zusammengefügte Datei, die erzeugt werden soll

 ".\Progs\Zusammen2"

Geben Sie einen neuen Dateinamen ohne Erweiterung an
Almo erzeugt 3 Dateien:

1. eine nicht lesbare Almo-Direktdatei mit der Erweiterung __.dir
2. eine anschaulbare Datei im freien Format mit der Erweiterung __.fre
3. eine Datei der Variablen- u. Ausprägungsnamen mit der Erweiterung __.nam

Erläuterungen zu den Dialogboxen

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten

Programmtext vorkommende Variablennummer. Normalerweise ist dies die Nummer der letzten Variable des eingelesenen Datensatzes. Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen. Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 an

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Stamm-Datei

The screenshot shows a dialog box titled "Stamm-Datei". It contains several input fields and buttons:

- A file path field: "C:\Almo15\TESTDAT\STAMMDAT2.FRE"
- A format dropdown menu: "frei" (with a "Hilfe" button next to it)
- A variable count field: "V1:10" (with a "Hilfe" button next to it)
- A section for "Verbindungsvariable aus Stamm-Datei" with a dropdown set to "V3" and a "Hilfe" button.
- A section for "KeinWert-Angabe für Verbindungsvariable" with a dropdown set to a red bar and a "Hilfe" button.

Im 1. Teil der Eingabebox wird der Name der Stammdatei, ihr Format und die Variablenzahl eines Datensatzes mitgeteilt. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4. Der Dateiname darf auch abgekürzt so geschrieben werden:

".\Testdat\Stammdat2.fre"

Der Pfad C:\Almo15 darf also durch einen Punkt repräsentiert werden.

Im 2. Teil muss die *Verbindungsvariable*, genauer deren Variablen-Nummer in der Stammdatei oder deren Variablen-Name in der Stammdatei, angegeben werden. Zum Begriff der Verbindungsvariablen siehe oben 2. Beispiel.

Die Verbindungsvariable kann den Wert *Kein-Wert* besitzen. Ist sie (beim Format FREI) in der Datei bereits mit "kw" kodiert, dann wird das Eingabefeld leer gelassen; ebenso wenn sie beim Format DIREKT bereits mit dem Kein-Wert-Code enthalten ist. Wenn aber beispielsweise für die Verbindungsvariable V3 der Wert -99 von Almo als Kein-Wert-Code interpretiert werden soll, dann muss hier eingegeben werden

V3 (-99 = KeinWert)

Optionsbox "weitere Angaben ..."

The screenshot shows a small dialog box with a title bar that reads "weitere Angaben zur Stamm-Datei - nicht wenn Format DIREKT".

Bei Diese Optionsbox wird nur geöffnet, wenn die Stammdatei im Format *erweitert-frei* vorliegt, d.h wenn sie Zeichenvariable enthält oder das Komma als Dezimalzeichen verwendet. Bei Prog00mf und Prog00mg dürfen Zeichenvariable nicht vorhanden sein. Dort kann also nur der Fall auftreten, dass das Komma als Dezimalzeichen verwendet wird.

Eingabebox: Anfüge-Datei

Optionsbox "weitere Angaben ..."

Die Eingabebox und die Optionsbox entsprechen den bereits oben erläuterten Boxen für die Stammdatei.

Eingabebox: Datei der Variablennamen der Stammdatei

Eingabebox: Datei der Variablennamen der Anfügedatei

Der Benutzer gibt hier den vollen Pfad- und Dateinamen an. Wobei der Dateiname auch abgekürzt so geschrieben werden darf:

```
".\Testdat\Stammdat2.nam"
```

Der Pfad C:\Almo15 darf also durch einen Punkt repräsentiert werden.

Eingabebox: Neue zusammengefügte Datei, die erzeugt werden soll

Der Benutzer muss für die neue gemeinsame Datei der Daten und der Variablennamen einen vollen Pfad- und Dateinamen angeben. Zulässig ist auch die oben beschriebene Kurzform. In unserem Beispiel in Prog00mf wird eingegeben ".\Progs\Zusammen2"

Zu beachten ist:

1. Es wird nur *ein* Name angegeben
2. Der Name wird ohne Erweiterung (wie etwa ".fre" oder ".nam") geschrieben.

Almo erzeugt 3 Dateien:

1. eine nicht lesbare Almo-Direktdatei mit der Erweiterung _____.dir
2. eine anschauliche Datei im freien Format mit der Erweiterung _____.fre
3. eine Datei der Variablen- u. Ausprägungsnamen mit der Erweiterung _____.nam

In unserem Beispiel in Prog00mf erzeugt Almo also folgende 3 Dateien:

Die Datendatei im Format DIREKT ".Progs\Zusammen2.dir"

Die Datendatei im Format FREI ".Progs\Zusammen2.fre"

Die Datei der Variablennamen ".Progs\Zusammen2.nam"

7.6.1 Datensatz bei Prog00mg und Prog00mv umgeordnet speichern

In Abschnitt 46.6.5.1 wurde gezeigt, wie in diesen beiden Programm-Masken der gemeinsame Datensatz umgeordnet in die gemeinsame Datei gegeben werden kann. In beiden Masken ist nachfolgende zusätzliche Eingabebox enthalten. In ihr wird so ausführlich erläutert, wie vorzugehen ist, dass das hier nicht nochmals vorgetragen werden muß.

Variable für neue zusammengefügte Datei


Ein Datensatz besteht nach dem Zusammenfügen nun (in unserem Beispiel) aus den Variablen der Stamm-Datei mit den Variablen-Nummern `V1 :10` und den Variablen der Anfügedatei mit den Variablen-Nummern `V11:16`

Auch die Variablennamen der Anfügedatei haben jetzt die neuen Nummern `Nome11 bis Name16`

Soll die neue gemeinsame Daten-Datei und Namens-Datei mit diesen Variablen- und Namens-Nummern gebildet werden, dann wird in das nachfolgendes Eingabefeld geschrieben `V1 :16`

Der Datensatz kann noch umsortiert werden, z.B soll V15 nach vorne an die 2. Stelle im Datensatz vorgerückt werden. Dann schreibt man in nachfolgendes Eingabefeld: `V1, 15, 2:14,16`

Man kann auch nur einige ausgewählte Variable in die gemeinsame Datei geben, z.B. V11 und V13 nicht in die gemeinsame Datei übernehmen wie nachfolgend:

 `V1,15,2:10,12,14,16`

Dieser so veränderte Datensatz wird von Almo neu nummeriert beginnend mit V1 und dann fortlaufend 2, 3, 4 Die Direktdatei (im Beispiel "Zusammen.dir") wird mit diesen Variablen-Nummern erzeugt und gespeichert. Auch die zusammengeführten und umsortierten Variablen-Namen werden mit diesen Nummern in die Namensdatei (im Beispiel "Zusammen.nam") gespeichert

8 Verschmelzen zweier Dateien

8.1 Beispiel

Im März werden 1000 Wähler zu ihrer Sympathie gegenüber 3 Kandidaten befragt. Es entsteht eine Datenmatrix mit 1000 Zeilen und 4 Spalten. Die 4. Spalte enthält die Wähler-Nummer. Die Matrix wird als Datei gespeichert.

Im April werden dieselben Wähler zu denselben Kandidaten wieder befragt, wobei dem Wähler i die gleiche Nummer zugewiesen wird, die er bei der 1. Befragung im März erhalten hatte. Von den 1000 Wählern konnten jedoch nur 900 angetroffen werden. Dafür werden 100 neue Wähler befragt, die dieselben demografischen Merkmale wie die ausgefallenen Wähler besitzen. Ihnen werden die Nummern 1001 bis 1100 zugeordnet. Es wird eine 2. Datei erstellt.

Die Variablen `A1, B1, C1` enthalten die Sympathie-Werte gegenüber den Kandidaten `A, B, C` vom März

Die Variablen `A2, B2, C2` enthalten die Sympathie-Werte gegenüber denselben Kandidaten vom April

| März-Datei | | | | April-Datei | | | | |
|------------|----|----|-----------|-------------|----|----|----|-----------------|
| A1 | B1 | C1 | Wähler-Nr | Wähler-Nr | A2 | B2 | C2 | |
| 4 | 8 | 1 | 1 | 400 | 4 | 6 | 1 | |
| 3 | 8 | 4 | 2 | 3 | 3 | 8 | 2 | die seitherigen |
| 3 | 6 | 3 | 3 | 1 | 4 | 9 | 1 | 1000 Wähler in |
| . | . | . | . | . | . | . | . | anderer Reihung |
| . | . | . | . | . | . | . | . | |
| 5 | 2 | 6 | 1000 | . | . | . | . | |
| | | | | ----- | | | | |
| | | | | 1001 | 3 | 7 | 4 | |
| | | | | 1002 | 4 | 2 | 1 | |
| | | | | . | . | . | . | die 100 neuen |
| | | | | . | . | . | . | Ersatz-Wähler |
| | | | | 1100 | 7 | 8 | 3 | |

Durch einen Trennstrich in der Aprildatei haben wir kenntlich gemacht, dass 100 neue Datensätze im April erstzweise dazu gekommen sind.

In einer realen 2-stufigen Befragung würden sicherlich auch noch demografische Variable (wie Geschlecht, Alter, Schulbildung etc.) erfragt und als Variable in die Dateien aufgenommen werden. Auch würden man versuchen, psychologische Einstellungs-Variable zu erheben.

Die beiden Dateien werden nun in zwei Schritten "verschmolzen"

- Die beiden Dateien werden wie oben in Abschnitt 7 beschrieben "zusammengefügt". Dabei wird die Märzdatei als Stammdatei und die Aprildatei als Anfügedatei betrachtet. Die Wähler-Nr. wird als Verbindungs-Variable verwendet. Dieselbe Wähler-Nummer in den beiden Dateien bezieht sich also auf dasselbe Individuum. Für den Wähler-Nr. i entsteht folgender gemeinsamer Datensatz

| A1 | B1 | C1 | Wähler-Nr | Wähler-Nr | A2 | B2 | C2 |
|----|----|----|-----------|-----------|----|----|----|
|----|----|----|-----------|-----------|----|----|----|

Die Wähler-Nummer ist zwei Mal im gemeinsamen Datensatz enthalten. Das muss in Kauf genommen werden. Für die 100 Wähler aus der März-Datei, die im April nicht angetroffen wurden, sind alle April-Variablen d.h. die Wähler-Nr, A2, B2, C2 gleich "kw" (=Kein-Wert).

- Die 100 neuen Wähler, die im April ersatzweise befragt wurden, werden am Schluss der gemeinsamen Datei "angehängt". Dabei erhalten sie in allen März-Variablen den Wert "kw" (Kein-Wert)

Die neue "verschmolzene" Datei, die Almo erzeugt, ist dann folgende:

| A1 | B1 | C1 | Wähler-Nr | Wähler-Nr | A2 | B2 | C2 | |
|----|----|----|-----------|-----------|----|----|----|---------------|
| 4 | 8 | 1 | 1 | 1 | 4 | 9 | 1 | |
| 3 | 8 | 4 | 2 | kw | kw | kw | kw | |
| 3 | 6 | 3 | 3 | 3 | 3 | 8 | 2 | |
| . | . | . | . | . | . | . | . | |
| . | . | . | . | . | . | . | . | |
| 5 | 2 | 6 | 1000 | 1000 | 7 | 6 | 3 | |
| | | | | ----- | | | | |
| kw | kw | kw | kw | 1001 | 3 | 7 | 4 | |
| kw | kw | kw | kw | 1002 | 4 | 2 | 1 | |
| kw | kw | kw | kw | . | . | . | . | die 100 neuen |
| kw | kw | kw | kw | . | . | . | . | Ersatz-Wähler |
| kw | kw | kw | kw | 1100 | 7 | 8 | 3 | |

In unserem Beispiel haben wir angenommen, das der Wähler Nr. 2 vom März einer von den 100 ist, die im April nicht angetroffen wurde. Er erhält deswegen in allen April-Variablen den Wert "kw" (=Kein-Wert) zugewiesen.

Ein 2. Beispiel

Dieses Beispiel wird in der Programm-Maske Prog00mx verwendet. In der Datei "Mai_Kauf" sind die Kunden enthalten, die im Mai gekauft haben, in der Datei "Junikauf" diejenigen, die im Juni gekauft haben Ein Teil der Kunden hat sowohl im Mai als auch im Juni gekauft. Diese Kunden sind also in beiden Dateien enthalten. Die Zahl der Datensätze und der Variablen in den beiden Dateien dürfen verschieden sein

| Datei 1 Maikauf | | | | Datei 2 Junikauf | | |
|--------------------|-------|-------|------------|---------------------|-------------|-------|
| Artikel-Nr. | Preis | Menge | Kunden-Nr. | Kunden-Nr. | Artikel-Nr. | Menge |
| 64 | 23 | 14 | 7 | 1 | 47 | 89 |
| 46 | 32 | 60 | 1 | 2 | 31 | 59 |
| 8 | 37 | 85 | 4 | 4 | 24 | 53 |
| 47 | 34 | 69 | 5 | 10 | 17 | 69 |
| 37 | 81 | 9 | 11 | 5 | 32 | 20 |
| 44 | 54 | 17 | 12 | | | |

In Schritt 1 des Verschmelzungsprozesses werden Mai- und Juni-Datei "zusammengefügt", so wie in Abschnitt 7 beschrieben. Die Kunden-Nr. wird als Verbindungsvariable verwendet. Die Kunden in der Mai-Datei mit den Kunden-Nummern 7, 11 und 12 haben im Juni nicht gekauft. Sie erhalten in der gemeinsamen Datei in den Juni-Variablen den Wert "kw" (Kein-Wert).

Im 2. Schritt werden die Kunden aus der Juni-Datei, die nicht für das Zusammenfügen gebraucht wurden, am Ende der gemeinsamen Datei "angehängt" Dies sind die beiden Kunden 2 und 10, die im Mai nicht gekauft haben. Sie erhalten in den Mai-Variablen den Wert "kw" (Kein-Wert).

Das Programm verschmilzt die Dateien zu folgender gemeinsamer Datei

| Mai | | | | Juni | | |
|-------------|-------|-------|------------|------------|-------------|-------|
| Artikel-Nr. | Preis | Menge | Kunden-Nr. | Kunden-Nr. | Artikel-Nr. | Menge |
| 64 | 23 | 14 | 7 | kw | kw | kw |
| 46 | 32 | 60 | 1 | 1 | 47 | 89 |
| 8 | 37 | 85 | 4 | 4 | 24 | 53 |
| 47 | 34 | 69 | 5 | 5 | 32 | 20 |
| 37 | 81 | 9 | 11 | kw | kw | kw |
| 44 | 54 | 17 | 12 | kw | kw | kw |
| kw | kw | kw | kw | 2 | 31 | 59 |
| kw | kw | kw | kw | 10 | 17 | 69 |

8.2 Eingabe in Prog00mx und Prog00my

Prog00mx "verschmilzt" zwei Dateien, wenn alle Variable numerisch sind. Prog00my leistet dasselbe, wenn die Variablen numerisch und/oder Zeichenvariable sind. Die beiden Programm-Masken sind identisch mit Prog00mf und Prog00mf. Nur eine Eingabebox, die Box "Speicher-Vereinbarung", ist anders. Nur diese wird deswegen abgebildet und erläutert.

Eingabebox: Speicher-Vereinbarungen

Speicher-Vereinbarungen

Vereinbare

Variable= ; Zahl der Variablen

VA = ; mindestens Zahl der Datensätze in Datei 2

1. Eingabefeld:

Die Zahl der vereinbarten Variablen muß mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Man kann die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen.

Beispiel: Datei 1 umfasse 20 Variable, Datei 2 umfasse 40 Variable. Dann gibt man mindestens 40 an. Die Zahl der Fälle spielt keine Rolle.

2. Eingabefeld:

Werden eine Datei 1 (Stammdatei) und eine Datei 2 (Anfügedatei) verschmolzen, dann muss die Zahl der Datensätze in der Datei 2 (der Anfügedatei) dem Programm mitgeteilt werden. Die eingegebene Zahl darf grösser aber nicht kleiner sein

9 Transponieren einer Datenmatrix

Gelegentlich werden Daten "verkehrt herum" geschrieben. Die Untersuchungsobjekte (z.B. Personen) wurden als Spalten und die Variablen (z.B. deren Merkmale) als Zeilen geschrieben. Eine solche Datenmatrix kann durch die meisten Almo-Maskenprogramme (auch die meisten anderen Statistikprogramme) nicht ausgewertet werden. Sie muss zuvor transponiert werden.

Es wurde beispielsweise folgende Datenmatrix geschrieben

| Variab | Untersuchungsobjekt | | | | | | | | | |
|--------|---------------------|--------|---------|-------|------|------|-------|------|-------|------|
| V1 | 4.2 | 3.2 | 1.1 | 3.1 | 2.1 | 2.2 | 1.1 | 1.1 | 7.1 | 5.1 |
| V2 | 2.2 | 3.2 | 4.1 | 2.2 | 6.1 | 7.1 | 3.2 | 4.1 | 4.2 | 7.1 |
| V3 | Meier | Müller | Hofer | Taler | Maas | Krug | Klaus | Bott | Fuchs | Holm |
| V4 | 3.1 | 7.1 | 4.1 | 1.1 | 3.2 | 2.1 | 9.1 | 8.2 | 6.1 | 1.1 |
| V5 | 5.1 | 6.1 | 4.34567 | 3.1 | 6.1 | 7.1 | 9.1 | 1.2 | 5.1 | 8.2 |

Die Untersuchungsobjekte bilden die Spalten, die Variablen die Zeilen. In unserem Beispiel enthält V3 in der 3. Zeile die Namen der Untersuchungspersonen. Dies ist also eine Zeichenvariable. Die anderen Variablen sind numerische Variable.

Almo enthält die beiden Programm-Masken Prog01m2 (nur für numerische Werte) und Prog01m3 für den allgemeinen Fall, dass auch Zeichenvariable in der Datenmatrix enthalten sind.

Bei Prog01m3 liest Almo alle Daten (auch die numerischen) als Zeichenvariable in eine interne Datenmatrix ein (was für den Benutzer belanglos ist), transponiert diese und schreibt sie dann im Format FREI in eine neue Datei.

Prog01m3 liefert als Ergebnis diese transponierte Datenmatrix

| Untersuchungsobjekt | Variable | | | | |
|---------------------|----------|-----|--------|-----|---------|
| | V1 | V2 | V3 | V4 | V5 |
| 1 | 4.2 | 2.2 | Meier | 3.1 | 5.1 |
| 2 | 3.2 | 3.2 | Müller | 7.1 | 6.1 |
| 3 | 1.1 | 4.1 | Hofer | 4.1 | 4.34567 |
| 4 | 3.1 | 2.2 | Taler | 1.1 | 3.1 |
| 5 | 2.1 | 6.1 | Maas | 3.2 | 6.1 |
| 6 | 2.2 | 7.1 | Krug | 2.1 | 7.1 |
| 7 | 1.1 | 3.2 | Klaus | 9.1 | 9.1 |
| 8 | 1.1 | 4.1 | Bott | 8.2 | 1.2 |
| 9 | 7.1 | 4.2 | Fuchs | 6.1 | 5.1 |
| 10 | 5.1 | 7.1 | Holm | 1.1 | 8.2 |

Im folgenden soll das allgemeinere Prog01m3 abgebildet und kurz erläutert werden.

Prog01m3.Msk

Daten-Matrix transponiert speichern
für numerische Variable und Zeichenvariable

Datensätze aus einer Datei einlesen
in eine Datenmatrix geben
diese dann transponiert in Datei speichern

Bedingung:
Einzulesende Datei nicht in Format DIREKT erlaubt

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare Variable= 20

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Datei aus der gelesen wird Hilfe

bei Datei-Problemen

"C:\Almo15\Testdat\verkehrt2.fre"

frei Format der Daten (nur frei oder fix) Hilfe

10 Zahl der Spalten der einzulesenden Daten-Matrix

wenn Dateiformat erweitert-FREI oder Fix

Datei für transponierte Daten-Matrix
die Daten werden im Format frei gespeichert

"..\Progs\Transdat.fre"

Programmende

Die einzulesende Datei muss sich im Format "frei" oder "fix" befinden. Das Format "direkt" ist nicht erlaubt. Diese drei Format wurden in Abschnitt 1 ausführlich dargestellt. Liegt die zu transponierende Datei im direkten Format vor, dann kann sie mit Prog00mp in das freie Format kopiert werden.

Erläuterungen zu den Dialogboxen

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Für Almo ist normalerweise die Zahl der Variablen gleich der Zahl der Spalten der Datenmatrix. Hier tritt aber nun eine kleine Komplikation auf: Durch das Transponieren werden die Zeilen der eingelesenen Matrix zu Spalten. Als Zahl der "vereinbarten Variablen" muss deswegen (mindestens) die größere der beiden - Spalten oder Zeilen der einzulesenden Datenmatrix - eingesetzt werden.

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Datei aus der gelesen wird

In der Eingabebox werden der Dateiname, das Format und die Spaltenzahl der einzulesenden Datenmatrix angegeben. Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

Der *Dateiname* darf auch abgekürzt so geschrieben werden:

```
"..\Testdat\verkehrt2.fre"
```

Der Pfad C:\Almo15 darf also durch einen Punkt repräsentiert werden.

Das *Format* in unserem Beispiel für Prog01m3 ist FREI und zwar Standard-Frei. Als Trennzeichen zwischen den Werten wird das Blank verwendet, als Dezimalzeichen der Punkt. Siehe dazu folgenden Abschnitt 1.

Die Datei enthält numerische Variable und auch eine Zeichenvariable. Almo liest alle Variable (auch die numerischen) als Zeichenketten ein. Das ist eine Besonderheit, die in Almo nur einmal, hier in Prog01m3 verwendet wird.

Optionsbox: Wenn Dateiformat erweitert-Frei oder FIX

Befinden sich die Daten im freien Format, dann kann die Optionsbox geschlossen bleiben, wenn

- (1) als Trennzeichen zwischen den Werten das Blank verwendet wird und
- (2) als Dezimalzeichen der Punkt.

Befinden sich die Daten im fixen Format, dann muss die Optionsbox geöffnet werden, weil zumindest die Feldbreite der Variablen angegeben werden muss.

Eingabebox: Datei für transponierte Dateimatrix

Der Benutzer muss hier einen neuen Dateinamen einsetzen. Almo legt die neue Datei an und speichert in sie die transponierte Datenmatrix.

10 Sortieren einer Datei

Almo enthält zwei Programm-Masken zum Sortieren von Dateien. Prog07m1 sortiert eine Datei, die nur aus numerischen Variablen besteht. Das allgemeine Prog07m2 sortiert eine Datei, die numerische Variable und/oder Zeichenvariable enthält. Dabei können bis zu 10 numerische Variable und/oder Zeichenvariable hintereinander "geschachtelt" als Sortiervariable eingesetzt werden. Im folgenden werden wir nur das allgemeinere Prog07m2 beschreiben.

10.1 Ein Beispiel

Betrachten wir eine Beispiel: Folgende Datei wird eingelesen

| V1 | V2 | V3 | V4 | V5 | V6 | V7 |
|---------|----|------|----|------|--------------|-----|
| Römer | 21 | nein | 31 | ja | Beamter | 1.5 |
| Hofer | 23 | ja | kw | nein | ' ' | 2 |
| Holm | 22 | nein | 32 | nein | Angestellter | kw |
| Bauer | 18 | ja | 33 | ja | Arbeiter | 4 |
| bauer | 19 | ja | 33 | ja | arbeiter | 4 |
| Ofner | 26 | ' ' | 34 | ja | Bauer | 5 |
| Schwarz | 31 | ja | 35 | nein | Arbeiter | 2.5 |
| Kahl | 22 | nein | 36 | ja | Beamter | 2 |
| Herr | 21 | nein | 37 | ja | beamter | 6 |
| See | 23 | ja | 38 | nein | angestellter | 7 |
| Bauer | 34 | nein | 39 | nein | ' ' | 8 |
| Reh | 22 | ja | 40 | ja | Arbeiter | 9 |
| Sieger | 22 | ' ' | 41 | ja | Bauer | 10 |

Die Datei wird nach V1 (Familiennamen) sortiert und innerhalb des Familiennamens nach V2 (dem Alter). Die Sortiervariablen werden also "geschachtelt". Insgesamt sind 10 Sortiervariable möglich. Die Programm-Optionen wurden im Beispiel so eingestellt, dass aufsteigend sortiert wird und zwischen Groß- und Kleinbuchstaben nicht unterschieden wird.

Die sortierte Datei ist folgende:

| | | | | | | |
|---------|----|------|----|------|--------------|-----|
| Bauer | 18 | ja | 33 | ja | Arbeiter | 4 |
| bauer | 19 | ja | 33 | ja | arbeiter | 4 |
| Bauer | 34 | nein | 39 | nein | ' ' | 8 |
| Herr | 21 | nein | 37 | ja | beamter | 6 |
| Hofer | 23 | ja | kw | nein | ' ' | 2 |
| Holm | 22 | nein | 32 | nein | Angestellter | kw |
| Kahl | 22 | nein | 36 | ja | Beamter | 2 |
| Ofner | 26 | ' ' | 34 | ja | Bauer | 5 |
| Reh | 22 | ja | 40 | ja | Arbeiter | 9 |
| Römer | 21 | nein | 31 | ja | Beamter | 1.5 |
| Schwarz | 31 | ja | 35 | nein | Arbeiter | 2.5 |
| See | 23 | ja | 38 | nein | angestellter | 7 |
| Sieger | 22 | ' ' | 41 | ja | Bauer | 10 |

Der Name "Bauer" tritt 3 mal auf. Ein Mal wird er klein geschrieben. Da aber "nicht unterscheiden" eingestellt wurde, spielt das keine Rolle. Innerhalb der 3 Fälle mit Namen "Bauer" wird nun nach der numerischen Variablen "Alter" aufsteigend sortiert.





10.2 Eingabe in Prog07m2





Prog07m2.Msk
Datensätze mit numerischen und Zeichenvariablen sortieren
mit Optionen





Programm-Bedienung ---> [Hilfe](#)

Speicher fuer x Variable [Hilfe](#)
Vereinbare Variable= **20**

 Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Datei der Variablennamen [Hilfe](#)
  **".\Testdat\Bestand2.nam"**
  **zeige** zeige = Namensdatei in Output zeigen
leer = nicht

Freie Namensfelder [Hilfe](#)
  
 **erzeuge zusätzliche Namensfelder**

Datei aus der gelesen wird [Hilfe](#) bei Datei-Problemen
 **".\TESTDAT\Bestand2.fre"**
 **frei** Format der Daten [Hilfe](#)
  **V1:7** der Datensatz enthält diese Variablen
Bei Format DIREKT schreiben Sie: alle_V

Loesche wieder diese Box

Wenn Dateiformat FIX oder nicht Standard-FREI

Fam.name, Hausbesitz, arbeitslos, Beruf

Zeichenvariable
 Geben Sie hier die Variablen an, die
 Zeichenvariable (also Worte) sind

Punkt
 Dezimalzeichen
 Punkt= Dezimalwerte sind mit dem Punkt
 als Dezimalzeichen geschrieben
 Komma= Dezimalwerte sind mit dem Komma
 als Dezimalzeichen geschrieben



wenn die Daten spaltengebunden sind (= Format FIX)
 schreiben Sie hier die Feldanweisung

wenn die Daten durch Trennzeichen getrent sind (= Format FREI)


Eingabefeld leer = jeder Datensatz ist vom
 nachfolgenden durch einen Zeilenumbruch
 getrennt
 ohne Zeilenumbruch = mehrere Datensätze
 stehen hintereinander in einer Zeile

Blank
 Nur bei Zeichenvariablen und Format FREI
 Trennzeichen hinter Zeichenvariablen
 Hinter den Zeichenvariablen steht als Trenn-
 zeichen ein Blank oder Komma oder Tabulator


die zu sortierenden Variablen Hilfe

  **Fam.name,Alter**


im Beispiel wird zuerst nach dem Fam.name und bei gleichem Fam.namen nach dem Alter sortiert.
Es können beliebig viele geschachtelte Sortiervariable angegeben werden

 **1** Hilfe


1 = aufsteigend sortieren
0 = absteigend sortieren

 **0**


bei numerischen Variablen
0 = Vorzeichen beachten
1 = nicht beachten (Absolutwert sortieren)

 **1**


bei Zeichen-Variablen
0 = Groß- und Kleinbuchstaben unterscheiden
1 = nicht unterscheiden

 **1**



bei Zeichen-Variablen
1 = deutsche Umlaute in Vokale wandeln
Ä Ö Ü ä ö ü ß --> A O U a o u s
0 = nach ASCII-Wert sortieren

 Option: Ein- und Ausschliessen von Untersuchungseinheiten

Die Variablen werden mit ihren umkodierten Werten sortiert
Variable mit KeinWert werden nach vorne (bei aufsteigendem
Sortieren) bzw. nach hinten (bei absteigendem Sortieren)
gestellt

 Option: Umkodierungen und Kein-Wert-Angaben

Datei, in welche die sortierten Daten geschrieben werden

  **".\PROGS\Sortiert"**

Geben Sie einen neuen Dateinamen ohne Erweiterung an
Almo erzeugt 2 Dateien:

1. eine nicht lesbare Datei im Direkt-Format mit der Erweiterung __.dir
2. eine anschaulbare Datei im freien Format mit der Erweiterung __.fre

Erläuterungen zu den Boxen:

Eingabebox: Speicher für x Variable

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.1. Die Zahl der vereinbarten Variablen muss mindestens so hoch sein, wie die höchste im gesamten Programmtext vorkommende Variablennummer. Normalerweise ist dies die Nummer der

letzten Variable des eingelesenen Datensatzes. Sie können die Zahl der vereinbarten Variablen aus Sicherheitsgründen auch höher setzen. Beispiel: Ein Datensatz aus Ihrer Datei umfasst 40 Variable. Dann geben Sie mindestens 40 oder mehr (z.B. 100) an. Sie dürfen großzügig sein.

Eingabebox: Weitere Vereinbarungen

Siehe dazu dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.2.

Eingabebox: Datei der Variablennamen

Siehe Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.3.

Eingabebox: Freie Namensfelder

Siehe P0.3.

Eingabebox: Datei aus der gelesen wird

In der Box werden der Dateiname, das Format und die einzulesenden Variablen angegeben.

Siehe dazu Almo-Dokument Nr. 0 "Arbeiten mit Almo", Abschnitt P0.4.

In unserem Beispiel werden V1,3,5,6 als Zeichenvariable (=Worte) angegeben.

Optionsbox: Wenn Dateiformat FIX oder nicht Standard-Frei

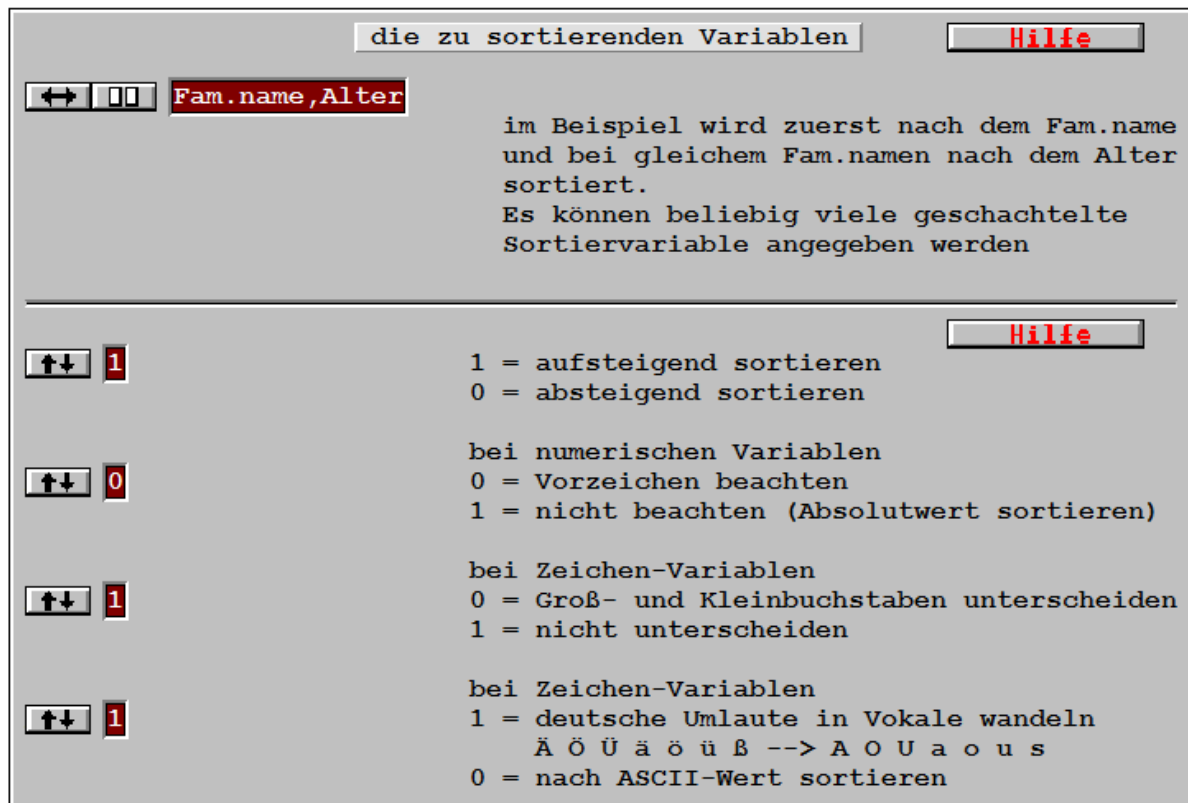
Befinden sich die Daten im freien Format, dann kann die Optionsbox geschlossen bleiben, wenn

- (1) alle Variable numerisch sind
- (2) als Trennzeichen zwischen den Werten das Blank oder Komma verwendet wird und
- (3) als Dezimalzeichen der Punkt.

Befinden sich die Daten im fixen Format, dann muss die Optionsbox geöffnet werden, weil zumindest die Feldbreite der Variablen angegeben werden muss.

In unserem Beispiel sind V1,3,5,6 Zeichenvariable. Die Optionsbox muss also geöffnet werden.

Eingabebox: Die zu sortierenden Variablen



Bevor diese Eingabebox erläutert werden kann, muss dargestellt werden, wie in Almo Zeichenvariable sortiert werden

10.2.1 Regeln für das Sortieren von Zeichenvariable.

Auf- oder absteigend sortieren

Aufsteigend sortieren bei numerischen Variablen: Die größere Zahl kommt nach hinten

Absteigend sortieren bei numerischen Variablen: Die kleinere Zahl kommt nach hinten

Zeichenvariable werden nach dem Zahlenwert ihrer ASCII-Kodierung sortiert. Probleme entstehen bei den deutschen Umlauten. Siehe nachfolgend.

Vorzeichen beachten oder Absolutwert

Numerische Variable können nach ihrem Absolutwert sortiert werden oder unter Beachtung ihres Vorzeichens.

Behandlung von Kein-Wert

Bei numerischen Variablen wird KeinWert (kw) als kleinst mögliche Zahl verstanden.

Bei aufsteigendem Sortieren werden also numerische Variable mit KeinWert nach vorne und bei absteigendem Sortieren nach hinten gestellt

Bei Zeichenvariablen wird Kein-Wert in der Datenmatrix durch ein Blank zwischen Einfach-Apostrophen gekennzeichnet, also so: ' '

Der ASCII-Wert des Blanks ist 32. Das ist das kleinste "sichtbare" Zeichen.

Bei aufsteigendem Sortieren von Zeichenvariablen werden deswegen Variable mit KeinWert (wie bei numerischen Variablen) nach vorne und bei absteigendem Sortieren nach hinten gestellt

Behandlung von Zeichenvariablen

Die Zeichenvariablen werden nach dem ASCII-Wert ihrer Zeichen sortiert.

| | | | | | | | | | | | |
|-------------|----|----|-----|----|----|--|----|----|-----|-----|-----|
| Zeichen: | A | B | ... | Y | Z | | a | b | ... | y | z |
| ASCII-Wert: | 65 | 66 | ... | 89 | 90 | | 97 | 98 | ... | 121 | 122 |

Klein-Buchstaben haben höhere ASCII-Werte als Großbuchstaben.
Der "Klammeraffe" @ hat den ASCII-Wert 64, steht also vor A.

Ziffern in Zeichenvariable (z.B. in dem Wort 'Platz 1') haben folgende Ascii-Werte

| | | | | | | |
|-------------|----|----|----|----|-------|----|
| Zeichen: | 0 | 1 | 2 | 3 | | 9 |
| ASCII-Wert: | 48 | 49 | 50 | 51 | | 57 |

Bei aufsteigender Sortierung stehen Ziffern also vor Buchstaben

Groß- und Kleinbuchstaben unterscheiden

Soll zwischen Groß- oder Kleinbuchstaben beim Sortieren unterschieden werden und wird aufsteigend sortiert, dann kommen zuerst die Großbuchstaben und dann die Kleinbuchstaben. Wird absteigend sortiert, dann umgekehrt.
Wird nicht unterschieden, dann werden alle Buchstaben wie Kleinbuchstaben behandelt.

Behandlung der deutschen Umlaute

Die ASCII-Kodierung der deutschen Umlaute ist nicht einheitlich. Siehe dazu den Wikipedia-Artikel zu ASCII. Almo übernimmt die in MS-DOS übliche Kodierung. Dies ist die vorherrschende Kodierung.

| | | | | | | | |
|-------------|-----|-----|-----|-----|-----|-----|-----|
| Zeichen: | ä | ö | ü | Ä | Ö | Ü | ß |
| ASCII-Wert: | 132 | 148 | 129 | 142 | 153 | 154 | 225 |

Diese Kodierung folgt keiner erkennbaren Logik. Im Almo-Sortierprogramm werden deswegen 2 Wege angeboten deutsche Umlaute zu sortieren.

1. Möglichkeit: Die obige Ascii-Kodierung wird verwendet.

Das führt dazu, dass die Umlaute bei aufsteigender Sortierung zum Schluss kommen - da ihre ASCII-Werte höher sind als die der Buchstaben und Ziffern.

2. Möglichkeit: Die deutschen Umlaute werden als Vokale behandelt.

Dann wird nach dem ASCII-Wert sortiert

| | | | |
|-----|---|------|---|
| Aus | ä | wird | a |
| | ö | | o |
| | ü | | u |
| | Ä | | A |
| | Ö | | O |
| | Ü | | U |
| aus | ß | wird | s |

Einige häufige Sonderzeichen

| | | | | | | | | | | | | | |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Zeichen: | ! | " | # | \$ | % | & | (|) | = | ? | + | - | @ |
| ASCII-Wert: | 33 | 34 | 35 | 36 | 37 | 38 | 40 | 41 | 61 | 63 | 43 | 45 | 64 |

Das Blank (Leerzeichen) besitzt den ASCII-Wert 32

Die vollständige ASCII-Tabelle mit den Werten 1 bis 255 findet man in Almo unter "Almo_Rsc\ASCII-Tabelle.doc". Die Tabelle muss mit Word (oder einem kompatiblen Textprogramm) gelesen werden. Siehe auch Wikipedia im Internet.

Eingabefeld im oberen Abschnitt der Eingabebox:

Sortiert wird nach V1 (Fam.name) und bei gleichen Werten von V1 nach V2 (Alter).

Ein anderes Beispiel. Im Eingabefeld steht: V1,3,5

Sortiert wird nach V1, bei gleichen Werten von V1 nach V3 und bei gleichen Werten von V1 und V3 nach V5.

Es können maximal 10 Sortiervariable angegeben werden.

In unserem Beispiel ist die erste Sortiervariable eine Zeichenvariable und die zweite eine numerische Variable. Also: Zeichen- und numerische Variable als Sortiervariable können beliebig gemischt werden.

Unterer Abschnitt der Eingabebox

Eingabefeld 1: Sortierfolge

Wird 1 eingegeben, dann wird aufsteigend sortiert.

Beispiel für numerische Variable: Die Zahlen 99,5,22,1 werden sortiert zu 1,5,22,99.

Wird 0 eingegeben, dann wird absteigend sortiert.

Beispiel für numerische Variable: In obigem Beispiel wäre die sortierte Zahlenfolge dann: 99,22,5,1

Bei Zeichenvariablen: Siehe obige Ausführungen in 10.2.1

Eingabefeld 2: Vorzeichen bei numerischen Variablen

0 = Vorzeichen beachten. Negative Zahlen stehen also bei aufsteigendem Sortieren vor positiven

1 = Vorzeichen nicht beachten. Die Zahlen werden nach ihrem Absolutwert sortiert

Eingabefeld 3: Groß- und Kleinschreibung bei Zeichenvariablen

0 = unterscheiden. Siehe oben 10.2.1

1 = nicht unterscheiden. Großbuchstaben werden wie Kleinbuchstaben behandelt.

Eingabefeld 4: Deutsche Umlaute

0 = nach ihrem ASCII-Wert sortieren. Siehe oben 10.2.1

1 = wie die entsprechenden Vokale behandeln. ß wie s.

Eingabebox: Ein- und Ausschließen von Untersuchungseinheiten

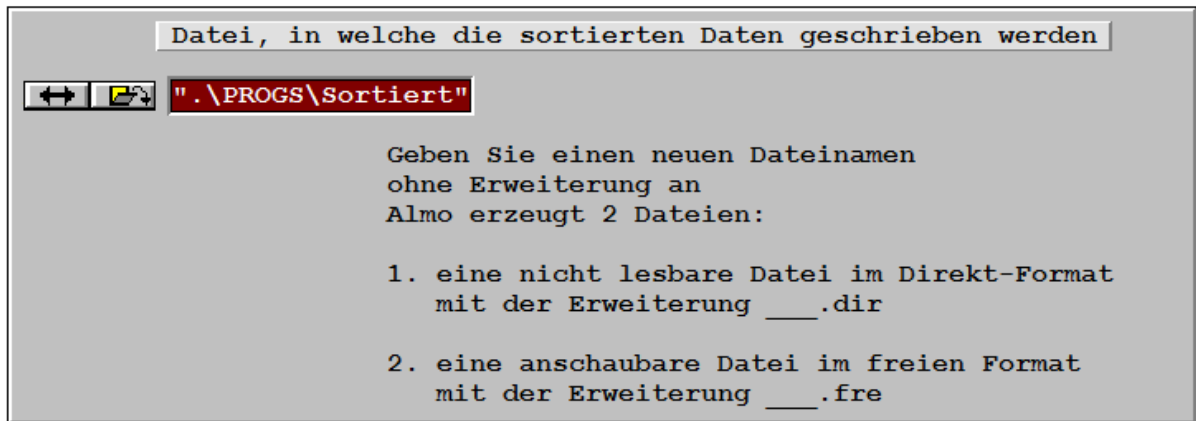
Siehe oben die ausführliche Beschreibung der Optionsbox in Abschnitt 3 oder Almo-Dokument Nr. 0 "Arbeiten mit Almo", P0.7.

Eingabebox: Kein_Wert-Angabe und Umkodierungen

Siehe Almo-Dokument Nr. 0 "Arbeiten mit Almo", P0.5

Werden Variable umkodiert, dann werden sie mit ihren umkodierten Werten sortiert. Variable mit KeinWert werden nach vorne (bei aufsteigendem Sortieren) bzw. nach hinten (bei absteigendem Sortieren) gestellt.

Eingabebox: Datei, in welche die sortierten Daten geschrieben werden



Geben Sie den Dateinamen (ohne Erweiterung) an.