



Kurt Holm

## Arbeiten mit dem Almo-Datenanalyse-System

Statistische Datenanalyse III  
Data-Mining III

Almo Statistik-System  
[www.almo-statistik.de](http://www.almo-statistik.de)  
[holm@almo-statistik.de](mailto:holm@almo-statistik.de)  
[kurt.holm@jku.at](mailto:kurt.holm@jku.at)

## Zum Begriff "Data-Mining"

Inzwischen sind wir mit dem Begriff "Data-Mining" nicht mehr sehr glücklich. Deswegen haben wir als 1. Titel für dieses Dokuments den Begriff "Statistische Datenanalyse" verwendet. Wir werden den Begriff "Data-Mining" jedoch beibehalten. Der Almo-Benutzer muss dabei akzeptieren, dass der Begriff in Almo eingengt wird auf die Auswertung von Daten, die die Form einer Datenmatrix besitzen. Siehe dazu Teil I.

## Weitere Almo-Dokumente

Die folgenden Dokumente können alle von der Handbuchseite in [www.almo-statistik.de](http://www.almo-statistik.de) herunter geladen werden. Die im folgenden Text häufig erwähnten Almo-Handbücher finden Sie ebenfalls unter dieser Adresse.

Das Almo-Dokument Nr. 0 "Arbeiten mit Almo" ist weitgehend identisch mit dem hier vorliegenden Dokument, dabei in mehreren Abschnitten ausführlicher. Lesen Sie auch dieses Dokument.

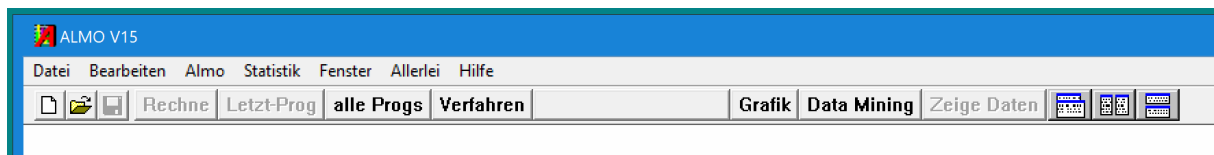
- 0. Arbeiten\_mit\_Almo.PDF (1 MB)
- 1a. Eindimensionale Tabellierung.PDF (1.8 MB)
- 1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
- 2. Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
- 3. Nicht-parametrische Verfahren.PDF (0.9 MB)
- 4. Kanonische Analysen.PDF (1.8 MB)  
Diskriminanzanalyse.PDF (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) (0,5 MB)
- 31. Pfadanalyse als wiederholte Regressionsanalyse (0,7 MB)
- 32. Datei-Operationen mit Almo (1,1 MB)
- 33. Wählerstromanalyse und Wahlhochrechnung (in Vorbereitung)

## Arbeiten mit dem Almo-Datenanalyse-System

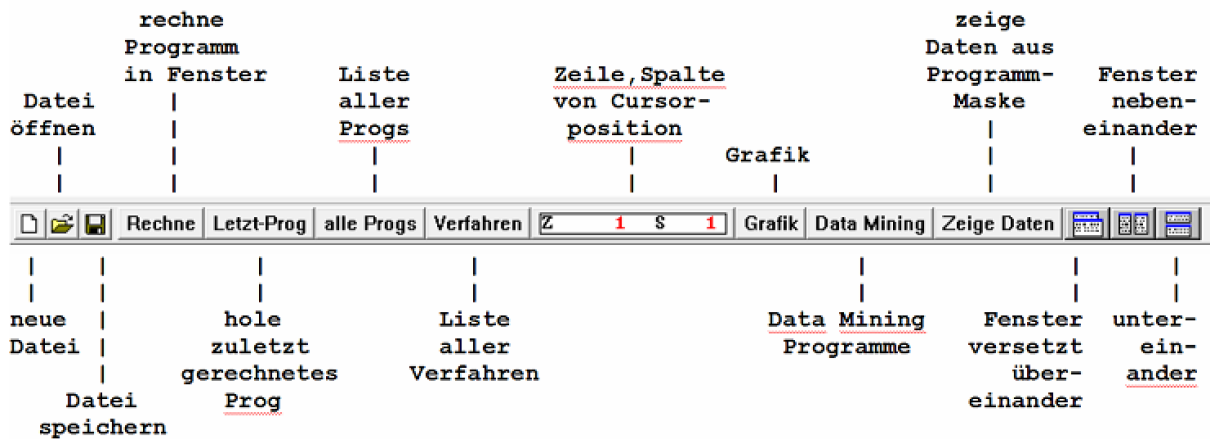
Wenn Sie mit dem Almo-Data-Mining-System arbeiten wollen, dann gehen Sie folgendermaßen vor:

Starten Sie Almo. Es erscheint das noch leere Almo-Fenster. Gleichzeitig erscheint eine Eingabe-Box mit mit den Knöpfen: (1) Hilfe und (2) Weiter. Klicken Sie auf "Weiter" und dann in der Knopfleiste (unterhalb der Menüzeile) auf den Knopf "Data Mining".



Wenn Sie mit den übrigen in Almo enthaltenen Verfahren arbeiten wollen, dann klicken Sie auf den Knopf „Verfahren“. Siehe die ausführliche Darstellung im Almo-Dokument Nr. 0 "Arbeiten mit Almo".

a. Die einzelnen Knöpfe in der "Knopfleiste" haben folgende Bedeutung:



**NEUE DATEI – SYMBOL** Die Datei-Auswahlbox von Windows erscheint. Wählen Sie ein Verzeichnis, bzw. Unterverzeichnis aus und geben Sie der Datei einen Namen. Almo erzeugt ein leeres Fenster

**DATEI ÖFFNEN - SYMBOL** Die Datei-Auswahlbox erscheint. Sie können jetzt eine Datei in den Almo-Editor laden

**DATEI SPEICHERN - SYMBOL** Die Datei im aktiven Fenster wird gespeichert.  
Beachte: Eine schon auf dem Speichermedium vorhandene Datei mit gleichem Namen wird über-schrieben

**RECHNE** Das Almo-Programm im aktiven Fenster wird ausgeführt

**LETZT-PROG** Das zuletzt gerechnete Programm wird wieder geladen

**alle PROGS** Angezeigt wird eine alphabetisch gereihete Liste aller in Almo vorhandener Programm-Masken und Syntaxprogrammen. Durch Mausklick auf ein Prog wird dieses geladen

**VERFAHREN** Eine Übersicht mit den in ALMO vorhandene Verfahren wird präsentiert. Durch Klick auf ein Verfahren laden Sie die zu diesem Verfahren vorhandenen Programm-Masken

**ZEILEN- UND SPALTEN-NUMMER**

Die Cursorposition im aktiven Fenster wird angegeben (sofern ein Fenster vorhanden ist)

**GRAFIK** Almo erzeugt bei einer Reihe von Programmen selbständig und automatisch Grafiken. Damit Sie aber auch für beliebige Daten Grafiken herstellen können, haben wir diesen Grafik-Knopf eingeführt. Wenn Sie auf ihn klicken, dann wird Ihnen eine Auswahl von Grafik-Typen angeboten. . Klicken Sie auf einen Grafiktyp, etwa "3D-Balken, eine Reihe" (geeignet für eindimensionale Häufigkeitsverteilungen). Almo lädt dann eine Grafik-Eingabemaske. In diese können Sie Ihre Daten eintragen und sich dann von Almo eine Grafik zeichnen lassen, indem Sie auf den großen Knopf "Grafik" klicken.

**DATA MINING** Sie erhalten eine Übersicht über die in Almo enthaltenen Data-Mining-Programme

**ZEIGE DATEN** Wenn sich im aktiven Fenster eine Programm-Maske befindet, dann werden die Daten, die in diesem Programm analysiert werden sollen, in einem zusätzlichen Fenster angezeigt

**SYMBOL-KNÖPFE** rechts neben Grafikknopf

Die 3 Knöpfe mit den Fenster-Symbolen bewirken, dass - sofern sich mehrere Fenster im Almo-Editor befinden - diese in der symbolisierten Weise übereinander, nebeneinander, untereinander gestellt werden

Nachdem Sie auf den Knopf „Data Mining“ geklickt haben, präsentiert Almo nachfolgende Übersicht über die Programme des Data-Mining-Systems.

## Programm auswählen

Für das gewählte Verfahren sind folgende Programme vorhanden  
Klicken Sie in die Zeile, die mit 'Prog \_\_ ' beginnt

### Data-Mining [ Sonderprogramme. Nicht in Standard-Almo ] =====

#### Kapitel 1: Eine Almo-Arbeitsdatei erstellen

- Prog45md: Schritt 1a: Daten aus Excel nach Almo übertragen  
und daraus eine Almo-Arbeitsdatei erstellen  
Prog45mh: Schritt 1b: Daten, die im Format FREI oder FIX vorliegen  
in eine Almo-Arbeitsdatei schreiben

#### Kapitel 2: Daten kennenlernen

- Prog45mg: Schritt 2: Daten anschauen  
Prog45ml: Schritt 3: Kennwerte der Variablen anschauen  
Prog45m4: Schritt 4: Variable auszählen

#### Kapitel 3: Daten bereinigen

- Prog45mo: Schritt 5a: Mittelwerte für fehlende Werte einsetzen  
Prog45mm: Schritt 5b: Prognosewerte für fehlende Werte einsetzen (mit ALM)  
Prog45mz: Schritt 5c: Prognosewerte für fehlende Werte einsetzen (Logitanalyse)  
Prog00ml: Schritt 5d: Rechenhilfe für multiple Imputation

#### Kapitel 4: Dateien vereinen

- Prog45mw: Schritt 6a: Dateien fusionieren (mit ALM)  
Prog45my: Schritt 6b: Dateien fusionieren (mit Logitanalyse)  
Prog45mx: Schritt 6c: Fusionierte Dateien vereinen

#### Kapitel 5: Mehrere Variable zu einer Messung kombinieren

- Prog45mv: Schritt 7a: Aus mehreren Variablenwerten einen Gesamtpunktwert bilden  
Prog45ms: Schritt 7b: Mit Faktorenanalyse einen gewichteten Gesamtpunktwert bilden  
Prog45mr: Schritt 7c: Mit Rasch-Skalierungsverfahren aus mehreren Variablen  
einen Messwert bilden

#### Kapitel 6: Zusammenhänge "blind" suchen

- Prog45m6: Schritt 8: Variable miteinander korrelieren

#### Kapitel 7: Einzelne Zusammenhänge genauer untersuchen

Prog45mb: Schritt 9: Variable tabellieren  
Prog45m7: Schritt 10: Streudiagramm für 2 oder 3 Variable

#### **Kapitel 8: Mehrfach-Zusammenhänge untersuchen**

Prog45mf: Schritt 11a: Ursachen für die Zielvariable: Allgemeines Lineares Modell  
Prog45gw: Schritt 11b: Alternative wenn Zielvariable polytom: Gewichtetes ALM  
Prog45m9: Schritt 11c: Alternative wenn Zielvariable nominal (dichotom, polytom):  
Logit-Analyse

#### **Kapitel 9: Prognose leisten**

Prog45mp: Schritt 12a: Werte für Zielvariable mit ALM prognostizieren  
Prog45mt: Schritt 12b: Werte für Zielvariable mit Logitanalyse prognostizieren

#### **Kapitel 10: Zusammengehörigkeiten zwischen Objekten suchen**

Prog45mn: Schritt 13: Cluster von Objekten bilden: Clusteranalyse

#### **Kapitel 11: Ursachen für die Clusterzugehörigkeit**

Prog45mq: Schritt 14a: Ursachen für die Clusterzugehörigkeit: ALM  
Prog45mu: Schritt 14b: Ursachen für die Clusterzugehörigkeit: Logit-Analyse

Abbruch

Lade selektiertes Programm

Gebe Info

Selektieren Sie nun durch Einmal-Klick ein Programm, z.B. „Prog45m4 Schritt 4: Variable auszählen“. Wenn Sie auf den Knopf "Gebe Info" klicken, dann erhalten Sie eine kurze Übersicht darüber, was das Programm leistet.



Wenn Sie in der großen Übersichts-Box auf den Knopf „Lade selektiertes Programm“ klicken (oder gleich auf den Programm-Namen doppelklicken), dann öffnet Almo ein sogenanntes Maskenprogramm. Es setzt sich aus hintereinander stehenden Dialogboxen zusammen. Sie sehen folgendes:

Prog45m4.Msk  
Variable auszählen

Das Programm erstellt Häufigkeitsverteilungen folgender Art

Kauf von	Faelle	%
1 Kleidung	204	20.40
2 Möbel	387	38.70
3 Technik	409	40.90

Dabei besteht die Möglichkeit die Häufigkeitsverteilungen für eine "Gruppierungsvariable" wiederholt zu erzeugen. Beispiel: Die Gruppierungsvariable sei "Geschlecht" (männlich, weiblich) Dann können die Häufigkeitsverteilungen separat ausgezählt werden für Männer und Frauen - und dies in einem Programm-durchlauf

Grafik: Balkendiagramme  
Handbuch P45 Data-Mining, Abschnitt P45.5

Programm-Bedienung --->

Vereinbare Variable=

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert

Datei der Variablennamen

"C:\Almo15\Testdat\DatMin.nam"

     zeige = Namensdatei in Output zeigen  
leer = nicht zeigen

Freie Namensfelder

Leere alle Eingabefelder dieser Sub-Box

erzeuge zusätzliche Namensfelder

Variablennamen in Datei speichern

Eingabefeld leer = nicht speichern



Datei aus der gelesen wird

"C:\Almo15\Testdat\DatMin.dir"

die auszuzählenden Variablen

Wohnort, Beruf, Einkommen, Hausbesitz, Produkt, Rueckzahl

Option: Gruppierungsvariable

Option: Ein- und Ausschliessen von Untersuchungseinheiten

Loesche wieder diese Box

Umkodierungen   
Kein\_Wert-Angabe

Einkommen (0 Schritt 10000 bis 60000=I)

erzeuge zusätzliche Felder für Umkodierungen / Kein\_Wert-Angaben

---

Kontrollieren, ob Umkodierung so erfolgt wie gewünscht

diese Variablen ...

Einkommen

1:20

... aus diesen Datensätzen  
vor und nach der Umkodierung  
zur Kontrolle anzeigen

Option: Untersuchungseinheiten gewichten

Grafik-Optionen

weitere Grafik-Optionen

zeige = Almo-Programmtext zeigen  
Editfeld leer = nicht zeigen

Das Maskenprogramm besteht aus einer einleitenden Beschreibung (darüber was es leistet) und mehreren grau schattierten Dialogboxen.

Die Dialogboxen enthalten rot eingefärbte Eingabefelder, in die der Benutzer seine Eingaben einschreibt. In vielen Fällen befindet sich links vor dem Eingabefeld ein

Knopf, mit dem die Benutzereingabe in gewisser Weise automatisiert werden kann (siehe dazu Almo-Handbuch, Teil 1, Bedienungsanleitung, Abschnitt 4b).

In den Dialogboxen befindet sich auch sehr häufig ein Knopf „Hilfe“. Nach Klick auf diesen Knopf erhält der Benutzer entsprechende Erläuterungen.

## Die Dialogboxen eines Maskenprogramms

Für alle in Almo vorhandenen Verfahren existieren derartige Maskenprogramme. Sie sind weitgehend gleich. Wir wollen deswegen im folgenden, die Dialogboxen, die in fast allen Almo-Verfahren auftreten, darstellen und erklären.

### P0.1 Eingabe-Box: Speicher für x Variable



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. Siehe Almo-Handbuch, Teil 1, Abschnitt 35.

### P0.2 Optionsbox: Weitere Vereinbarungen



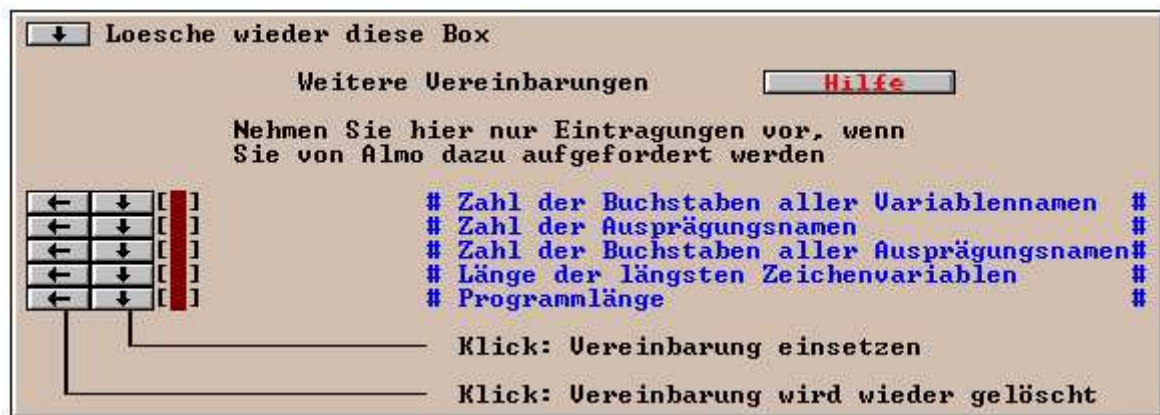
In Almo sind viele derartige Optionsboxen enthalten. Über „Optionen“ werden Voreinstellungen verändert oder zusätzliche Programmleistungen aufgerufen.

Wenn Sie diese Optionsbox durch Klick auf den Knopf mit nach unten weisenden Pfeil öffnen, dann erscheint folgende Mitteilung.



Es wird Ihnen also gesagt, daß die nun folgende Optionsbox bereits Eintragungen enthält, die Ihnen beispielhaft verdeutlichen soll, wie die Optionsbox auszufüllen ist. Sie müssen also alle Eingabefelder durchgehen und Ihre Daten und Analyseabsichten entsprechend ausfüllen. Das gilt nicht immer. Gelegentlich sind

die Eingabefelder der Optionsbox noch leer. Nach Klick auf „OK“ wird Ihnen dann die eigentliche Optionsbox präsentiert.



Am oberen Rand der Eingabe-Box sehen Sie den Knopf „Loesche wieder diese Eingabe-Box“. Nach Klick auf diesen Knopf wird der ursprüngliche Zustand wieder hergestellt. Almo arbeitet wieder mit den Standard-Voreinstellungen.

Almo benötigt Speicherplatz für die Variablen- und Ausprägungsnamen, für den Programmtext und für die Zeichenvariable (die aus den Daten eingelesen werden). Almo reserviert sich diesen Speicherplatz selbstständig, und zwar in Abhängigkeit von der Zahl der vom Benutzer vereinbarten Variablen.

Wenn Sie z.B. ungewöhnlich viele und sehr lange Ausprägungsnamen haben, dann reicht dieser Speicherplatz eventuell nicht aus. In diesem Fall müssen Sie selbst Speicher vereinbaren.

In der Eingabe-Box "Weitere Vereinbarungen" können Sie eine oder mehrere der folgende Vereinbarungen vornehmen. Klicken Sie dazu auf den nach unten weisenden Pfeil. Mit dem nach links weisenden Pfeil löschen Sie die Vereinbarung wieder.

```
V_Namen_Laenge=x; # Zahl der Buchstaben aller Variablennamen #
A_Namen          =x; # Zahl der Ausprägungsnamen #
A_Namen_Laenge=x; # Zahl der Buchstaben aller Ausprägungsnamen#
ZV_Laenge        =x; # Länge der längsten Zeichenvariablen #
ProgrammLaenge=x; # Programmlänge #
```

Setzen Sie für "x" eine Zahl ein - aber nur, wenn Sie durch eine Fehlermeldung von Almo dazu aufgefordert werden.

## P0.3 Eingabe-Box: Datei der Variablennamen

<b>Variablennamen</b>	
Datei der Variablennamen	<input type="button" value="Hilfe"/>
<input type="button" value="↔"/> <input type="button" value="📁"/>	<input c:\almo15\testdat\datmin.nam\""="" type="text" value="\"/>
<input type="button" value="↔"/> <input type="button" value="↓"/>	<input type="text" value="zeige"/>
	zeige = Namensdatei in Output zeigen leer = nicht zeigen
<b>Freie Namensfelder</b>	
	<input type="button" value="Hilfe"/>
<input type="button" value="↔"/>	Leere alle Eingabefelder dieser Sub-Box
<input type="button" value="↔"/>	<input type="text" value="Name 1=Wohnort:Stadt, Land"/>
<input type="button" value="↔"/>	<input type="text" value="Name 2=Geschlecht:m,w"/>
<input type="button" value="⋮"/>	erzeuge zusätzliche Namensfelder
<b>Variablennamen in Datei speichern</b>	
	<input type="button" value="Hilfe"/>
	Eingabefeld leer = nicht speichern
<input type="button" value="↔"/> <input type="button" value="📁"/>	<input type="text"/>

Die Box besteht aus 3 Sub-Boxen

1. Datei der Variablennamen
2. Freie Namensfelder
3. Variablennamen in Datei speichern

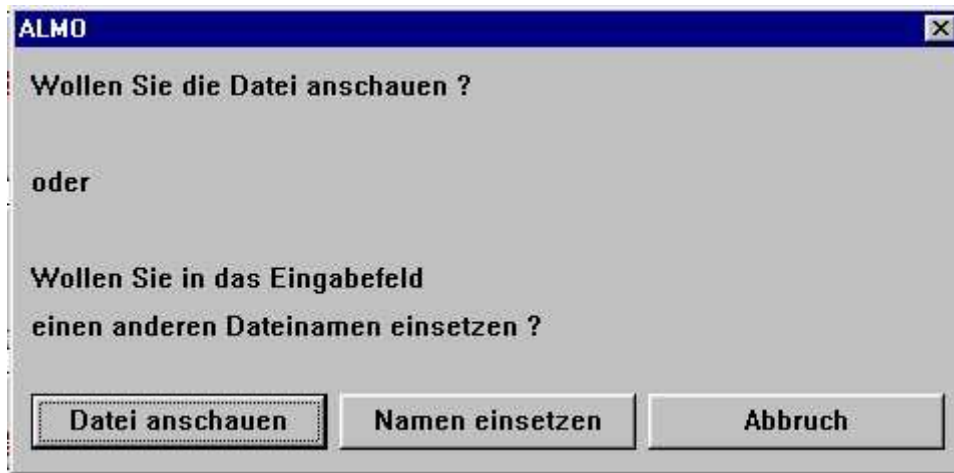
Sie müssen keine Variablennamen verwenden. Sie können auch nur Variablen-Nummern (mit einem V.. davor) verwenden - also z.B. V5, V48 ..... etc.

Dabei bezieht sich normalerweise die Nummer hinter V.. auf die Stelle der Variablen im eingelesenen Datensatz. V5 bezeichnet also die 5. Variable im eingelesenen Datensatz.

Es ist jedoch besser Variablennamen zu verwenden.

Sie können die Variablennamen aus einer Datei einlesen. In obiger 1. Eingabe-Box müssen Sie dann den Dateinamen angeben. Schreiben Sie den Dateinamen in das Eingabefeld oder noch einfacher: Klicken Sie zu diesem Zweck auf den Knopf mit dem Diskettensymbol.

Es erscheint dann eine Dialogbox, in der Sie gefragt werden, ob Sie die Datei anschauen wollen oder ob Sie in das Eingabefeld einen (neuen) Namen einsetzen wollen.



Klicken Sie auf "Namen einsetzen". Es erscheint dann die gewohnte Datei-Auswahlbox, in der Sie die Datei suchen, in der sich die Variablennamen befinden.

Im 2. Eingabefeld der 1. Eingabe-Box kann durch Klick auf den Knopf mit dem nach unten weisenden Pfeil das Wort „zeige“ eingesetzt werden. Das bewirkt, daß in der Ergebnisausgabe Ihres Programms die Datei der Variablennamen auch ausgegeben wird. Falls Variablennamen nicht Almo-gerecht geschrieben wurden, erscheint eine Fehlermeldung. Erst wenn Sie die Sicherheit haben, daß die Datei der Variablennamen fehlerfrei ist, können Sie das Wort „zeige“ aus dem Eingabefeld löschen.

Wenn Sie keine Datei der Variablennamen angelegt haben, dann müssen Sie selbstverständlich den Dateinamen im Eingabefeld löschen – durch Klick auf den doppelköpfigen „Rein-Raus-Knopf“. Durch ein nochmaliges Klicken auf den Knopf holen Sie den Text im Eingabefeld wieder zurück. Deswegen heißt der Knopf auch 'Raus-Rein-Knopf'.

In der 2. Sub-Box „Freie Namensfelder“ können Sie

- (1) wenn Sie keine Datei der Variablennamen angelegt haben, Variablennamen für Ihre Analysevariablen angeben, oder
- (2) zusätzlich zu den Variablennamen in der Datei der Variablennamen weitere angeben und/oder
- (3) Variablennamen in der Datei der Variablennamen überschreiben.

In der 3. Sub-Box "Variablennamen in eine Datei schreiben können Sie die Variablen- und Ausprägungsnamen in eine Datei speichern. Geben Sie einen neuen noch nicht vorhandenen Dateinamen an. Es muss der volle Pfad- und Dateinamen eingesetzt werden. Gespeichert werden die aktuellen Namen. Das sind die, die in der eingelesenen Datei enthalten sind sowie die zusätzlichen und überschriebenen Namen.

Von einem "Variablennamen" sprechen wir in Almo, wenn der Benutzer extra einen Namen definiert, z.B. so:

Name 5 =Geschlecht;

Dies ist dann der Variablenname von V5

Wenn Sie Variablennamen verwenden wollen, dann gibt es dafür 3 Möglichkeiten:

1. *Möglichkeit*: Sie legen eine Datei der Variablennamen an.

Schauen Sie sich als Beispiel die Datei '..\Almo\Testdat\DatMin.nam' an. Machen Sie einen Doppelklick auf den Dateinamen. Dann wird die Namensdatei geladen und Sie können sie sich anschauen. Sie sehen folgendes:

```
Name1=Wohnort:Stadt,Land;  
Name2=Geschlecht:m,w;  
Name3=Beruf:Selbst,Unselbst;  
Name4=Einkommen:bis 10,10-20,20 bis 30,30 bis 40,40-50;  
Name5=Kinderzahl;  
Name6=Hausbesitz:kein Haus,hat Haus;  
Name7=Rueckrate;  
Name8=Laufzeit;  
Name9=Produkt:Kleidung,Möbel,Technik;  
Name10=Rueckzahl:nein,ja;
```

Eine neue Namensdatei legen Sie folgendermaßen an: Klicken Sie auf das Menü DATEI und dann auf den Eintrag 'Neue Datei anlegen'. Wenn dann ein leeres Fenster erscheint, dann können Sie in dieses die Variablennamen schreiben.

Sie können es sich bequem machen und nur für einige Variable Namen in die Namensdatei schreiben; im Almo-Programm verwenden Sie dann für die andere Variablen deren Nummern (mit V.. davor, also z.B. V5). Sie können also problemlos Variablen-Namen und -Nummern mischen. Von Analyse zu Analyse können Sie dann in der Namensdatei Namen dazuschreiben und so allmählich eine vollständige Namensdatei entwickeln.

Um Ihre Namensdatei in das Eingabefeld einzutragen, klicken Sie auf den Knopf (mit dem Diskettensymbol) vor dem Dateinamen. Es erscheint dann eine Dialogbox und dann die Datei-Auswahl-Box, in der Sie Ihre Namensdatei suchen und anklicken können. Deren Name wird dann von Almo in das Eingabefeld gesetzt.

2. *Möglichkeit*: Sie verwenden keine Datei der Variablennamen, sondern schreiben die Namen, die Sie für Ihre Analyse benötigen, in die Eingabe-Box "Freie Namensfelder". Das Eingabefeld mit dem Dateinamen löschen Sie, indem Sie auf den Rein-Raus-Knopf vorne links klicken.

3. *Möglichkeit*: Sie können eine Datei der Variablennamen verwenden und zusätzlich dazu in die Eingabe-Box "Freie Namensfelder" Namen einschreiben.

Dabei ist es dann möglich, Variable zu überschreiben, für die schon in der Namensdatei ein Name steht. Man wird sich diese Möglichkeit zunutze machen, wenn man z.B. Variable für eine Analyse umkodiert.

Beispiel: In der Namensdatei steht

```
Name 3=Beruf:Arbeiter,Angestellter,Handwerker,Unternehmer;
```

Sie wollen die 4 Ausprägungen dieser Variablen auf 2 zusammenfassen und folgenden Namen geben:

```
Name 3=Berufsgruppe:Arbeitnehmer,Selbständige;
```

Diesen Namen schreiben Sie in die Eingabe-Box "Freie Namensfelder". In der Namensdatei brauchen Sie nichts zu ändern.

In der Eingabe-Box „Kein-Wert-Angabe und Umkodierungen“, weiter unten im Maskenprogramm schreiben Sie

```
Berufsgruppe(1,2=1; 3,4=2).
```

Wenn ein Variablen- und Ausprägungsnamen sehr lang ist, dann können Sie ihn auch in 2 oder mehr Eingabefelder schreiben. Beispiel:



Die Zeichenfolge --> (2 Minuszeichen, 1 Größer-Zeichen) am Ende des oberen Eingabefeldes signalisieren Almo, dass im nächsten Eingabefeld weitere Ausprägungsnamen zu derselbe Variablen folgen.

Sie können auch über die Grenze der Dialog-box hinaus nach rechts weiterschreiben. Das Eingabefeld verlängert sich automatisch über die rechte Dialogbox-Grenze hinaus.



**Wie man Variablen-Namen schreibt** (siehe Almo-Handbuch, Teil 2, Abschnitt 17)

Beispiele für Namensgebung:

```
Name 5 =Geschlecht:männlich,weiblich;
```

Variable 5 erhält den Namen 'Geschlecht'. Ihre Ausprägungen werden mit 'männlich' und 'weiblich' bezeichnet. Wir bezeichnen "Geschlecht" als Variablennamen und "männlich" und "weiblich" als Ausprägungsnamen.

'Name' kann mit 'N' abgekürzt werden. Man kann also auch schreiben:

```
N 5 =Geschlecht:männlich,weiblich;
```

Es ist auch möglich, die Codeziffern von männlich und weiblich vor die Ausprägungsnamen zu setzen. Man kann also schreiben:

```
N 5 =Geschlecht:(1)männlich,(2)weiblich;
```

Wir werden im nächsten Abschnitt ausführlich auf diese Schreibweise eingehen. Die Bestandteile der Namensgebung sind also folgende:

1. Das Almo-Schlüsselwort 'Name' oder kurz 'N'
2. die Variablennummer
3. das Gleichheitszeichen
4. der Variablenname

5. ein Doppelpunkt
6. eine Codeziffer in runden Klammern (nicht obligatorisch)
7. die Ausprägungsnamen; diese werden durch Beistriche voneinander getrennt
8. zum Schluss ein Semikolon

Almo unterscheidet zwischen Variablennamen und Ausprägungsnamen.

```
Name 6 =Lebensalter;
```

Hier wird nur der Variablennamen angegeben

```
Name 7 =:ja,nein;
```

Hier werden nur die Ausprägungsnamen verwendet (was eher selten sein wird). Beachte den Doppelpunkt. Almo wird damit signalisiert, daß jetzt Ausprägungsnamen folgen.

```
Name 20:30,45,60:65 =:ja,nein,weiß nicht;
```

Die Variablen 20 bis 30,45,60 bis 65 erhalten die 3 Ausprägungsnamen 'ja', 'nein' und 'weiß nicht'. Auf diese Art und Weise kann man Schreibarbeit sparen.

### Wie man Ausprägungsnamen schreibt

Siehe dazu auch die sehr ausführliche Erläuterung in Handbuch, Teil 3, P0.3.

Betrachten wir einige Beispiele

```
Name 3 = Beruf: Arbeiter, Angestellte, Sonstiger;
```

Variable 3 erhält den Variablennamen "Beruf". Die Variable besitzt 3 Ausprägungen, die die Ausprägungsnamen "Arbeiter", "Angestellter" und "Sonstiger" erhalten. Wurden diese 3 Berufe mit 1, 2, 3 kodiert, dann werden die Ausprägungsnamen von Almo korrekt in der Ergebnisliste verwendet. Es wird als in folgender Weise zugeordnet:

Codeziffer	Ausprägungsnamen
1	Arbeiter
2	Angestellter
3	Sonstiger

Keine Probleme treten auf, wenn die Werte-Untergrenze nicht 1 ist, sondern z.B. 4 und wenn die Schrittweite mit der die Codeziffer aufeinanderfolgen, nicht 1 ist sondern z.B. 5 oder sogar ungleich ist.

Wir können folgende Regel formulieren:

1. Regel: Almo ordnet die Ausprägungsnamen nacheinander den wachsenden Codeziffern zu – egal mit welcher Zahl diese beginnen und egal mit welcher Schrittweite sie aufeinander folgen. Die Codeziffern dürfen auch Dezimalzahlen sein.

Betrachten wir ein weiteres Beispiel. Die Variable Beruf besitzt 6 Ausprägungen, die mit 1, 2, 3, 4, 5, 6 kodiert sind. Wir wollen nur den Berufen mit der Codezahl 3 und 5 einen Namen zuordnen.

```
Name 13 = Beruf: (3) Schlosser, (5) Maurer;
```



Almo liefert z.B. das Programm zur Häufigkeitsverteilung folgenden Output:

Variable 13 Beruf

Wert	Faelle	%	% kumuliert
1	15	24.59	24.59
2	28	45.90	70.49
3 Schlosser	10	16.39	86.89
4	4	6.56	93.44
5 Maurer	3	4.92	98.36
6	1	1.64	100.00
Summe	61	100	

Wir können folgende Regel formulieren:

2. Regel: Werden Ausprägungsnamen eine Ziffer in runden Klammern vorausgestellt, dann werden diese Ausprägungsnamen den entsprechenden Codezahlen zugeordnet. Codezahlen, für die es keinen in dieser Weise gekennzeichneten Ausprägungsnamen gibt, erhalten keinen Namen. Die Codezahlen und die vorausgestellten Ausprägungsziffern dürfen auch negativ sein (mit Minuszeichen) und sie dürfen auch Dezimalzahlen sein.

## P0.4 Eingabe-Box: Datei aus der gelesen wird



In vielen Data-Mining-Programmen müssen sich die Daten in Format „direkt“ befinden. 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 mit 1 bis 8 Byte (je nach Größe des Variablenwertes).

Siehe die ausführliche Darstellung dieses Formats im Almo-Handbuch, Teil 2, Abschnitt 14.

Das Almo-Data-Mining-System enthält die 2 Programme, Prog45md und Prog45mh, mit deren Hilfe der Benutzer seine Daten in das direkte Format übertragen kann. Diese beiden Programm-Masken wurden im Teil 1 des "statistischen Datenanalyse-Systems" Abschnitt P45.1 und P45.2 ausführlich erläutert.

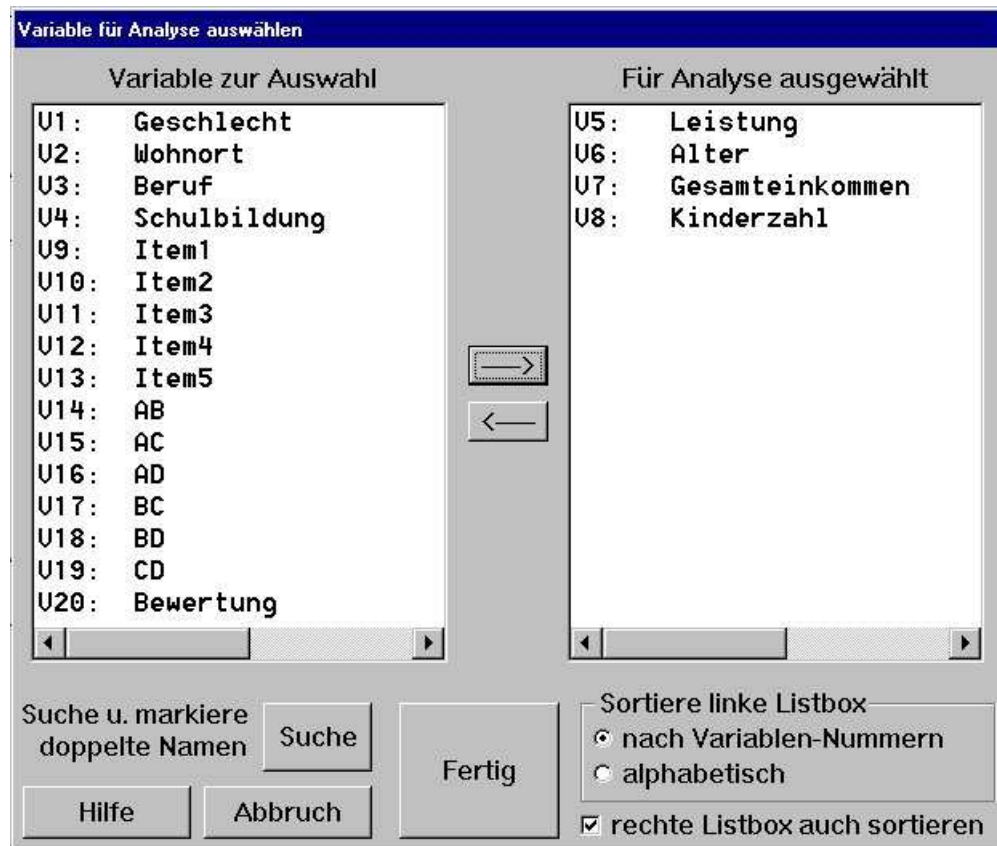
## P0.5 Eingabe-Box: Analysevariable, die auszuzählenden Variablen



Geben Sie hier die Variablen an, die Sie analysieren wollen. Sie können die Variablen unmittelbar in das Eingabefeld schreiben.

#### *Eingabe der zu analysierenden Variablen durch Mausklick*

Wenn Sie auf den Knopf mit den 2 Fenstersymbolen klicken, dann wird die Eingabe-Box "Variablen für Analyse auswählen" geöffnet. In Ihr geben Sie an, welche Variable analysiert werden sollen.



#### *Wie man die Dialogbox "Variablen für Analyse auswählen" bedient*

Klicken Sie auf eine Variable in der linken Listbox „Variable zur Auswahl“. Dann klicken Sie auf den Pfeilknopf. Die Variable wird dann in die rechte Listbox "Für Analyse ausgewählt" transportiert. Der „Transport“ kann auch in der umgekehrten Richtung erfolgen.

Mit gedrückter Strg-Taste und Klick auf einen Variablennamen markieren Sie diesen zusätzlich, ohne daß schon markierte Namen 'entmarkiert' werden.

Mit gedrückter Umschalt-Taste und Klick auf den oberen Variablennamen und dann auf den unteren Variablennamen markieren Sie einen ganzen Bereich.

Mit Doppelklick befördern Sie sofort einen Variablennamen von der einen Listbox in die andere.

Die Knöpfe am unteren Rand der Dialogbox haben folgende Bedeutung:

**SORTIERE** linke Listbox nach Variablennummern

Die Variablen in der linken Listbox werden nach aufsteigenden Nummern hintereinander gestellt.

SORTIERE linke Listbox alphabetisch

Die Variablen in der linken Listbox werden alphabetisch hintereinander gestellt. Variable, die keine Namen besitzen werden an das Ende gestellt.

Rechte Listbox auch sortieren

Die Variablen in der rechten Listbox werden nach der gleichen Sortiermethode wie in der linken Listbox hintereinander gestellt.

Knopf FERTIG

Wenn Sie abschließend auf den Knopf FERTIG klicken, dann werden die Variablen in der rechten Listboxen in das Eingabefeld des Maskenprogramms eingesetzt. Wenn die hintereinander gestellten Variablennamen zu lang würden, dann verwendet Almo automatisch Variablennummern.

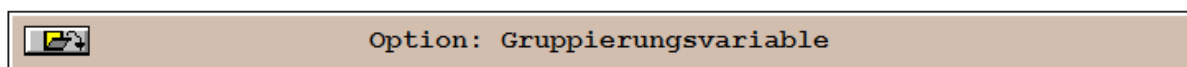
Knopf SUCHE

Variablennamen müssen eindeutig sein. Sie dürfen nicht doppelt vorhanden sein. Mit Klick auf den Knopf SUCHE prüft Almo, ob Namen doppelt oder sogar mehrfach vorkommen. Diese Variablennamen werden dann durch 2 vorausgehende Unterstriche markiert, z.B. so:

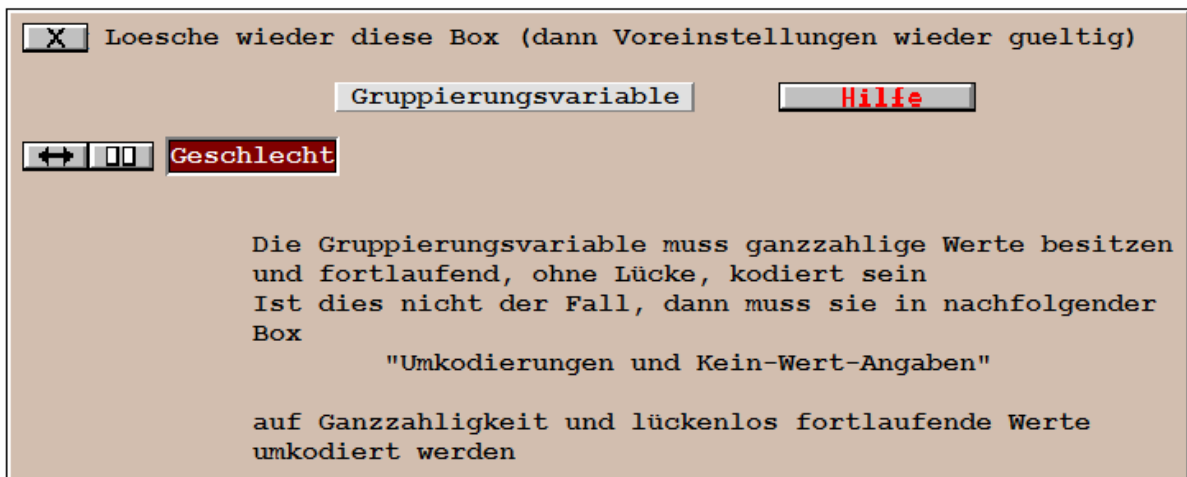
V25: \_\_Geschlecht

Diese Variablennamen dürfen dann nicht für die Analyse ausgewählt werden.

## P0.6 Optionsbox: Gruppierungsvariable



Wird die Optionsbox geöffnet, dann kann man beispielsweise folgendes eingeben



Es besteht die Möglichkeit die Analyse für eine "Gruppierungsvariable" wiederholt durchzuführen

Beispiel:

Die Gruppierungsvariable sei "Schulabschluß" mit den Ausprägungen: Volksschule, Gymnasium, Universität.

Dann kann die Analyse für die Analyse-Variablen separat durchgeführt werden für Personen mit

- und dies in einem Programmdurchlauf

Die Gruppierungsvariable muss ganzzahlige Werte besitzen und fortlaufend, ohne Lücke, kodiert sein. Ist dies nicht der Fall, dann muss sie nachfolgend in der Box

"Umkodierungen und Kein-Wert-Angaben"

auf Ganzzahligkeit und lückenlos fortlaufende Werte umkodiert werden

## P0.7 Optionsbox: Ein- und Ausschließen von Untersuchungseinheiten



Wenn Sie die Optionsbox öffnen, dann können Sie z.B. folgende Eingabe vornehmen:

The screenshot shows a dialog box titled "Option: Ein- und Ausschliessen von Untersuchungseinheiten". At the top, there are two buttons: "X" (Close) and "↔" (Reset). Below this, there are several sections, each with a "↔" button and a "Hilfe" (Help) button:

- Ein- und Ausschliessen von Untersuchungseinheiten**: A section header with a "Hilfe" button.
- nur diese Datensätzen verwenden wenn Eingabefeld leer, dann alle**: A section with a "↔" button and a "Hilfe" button.
- Untersuchungseinheiten in Analyse nur einschließen wenn .....**: A section with a "↔" button, a text input field containing "V8 kleiner 9", and a "Hilfe" button.
- Untersuchungseinheiten aus Analyse ausschließen wenn .....**: A section with a "↔" button, a text input field containing "V6 gleich 1.1", and a "Hilfe" button.
- nur eine Zufallsstichprobe von x % der Datensätze für Analyse verwenden**: A section with a "↔" button and a "Hilfe" button.
- Startwert für Zufallsgenerator**: A section with a "↔" button and a "Hilfe" button.

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

V8 kleiner 9

Das bedeutet: In die Analyse 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 aus der Analyse ausgeschlossen. Alle anderen werden analysiert.  
Ein drittes Beispiel: In beide Eingabefelder wird ein Eintrag gesetzt

Im Eingabefeld "nur Einschließen" steht V8 kleiner 9  
Im Eingabefeld "Ausschließen" steht V6 gleich 1.1

Das bedeutet: In die Analyse 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 Eingabefelder 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 Editierfeld 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 kleinergleich 2

Auch UND sowie ODER sind möglich. Beispiel

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

Siehe Abschnitt 29 - 32.

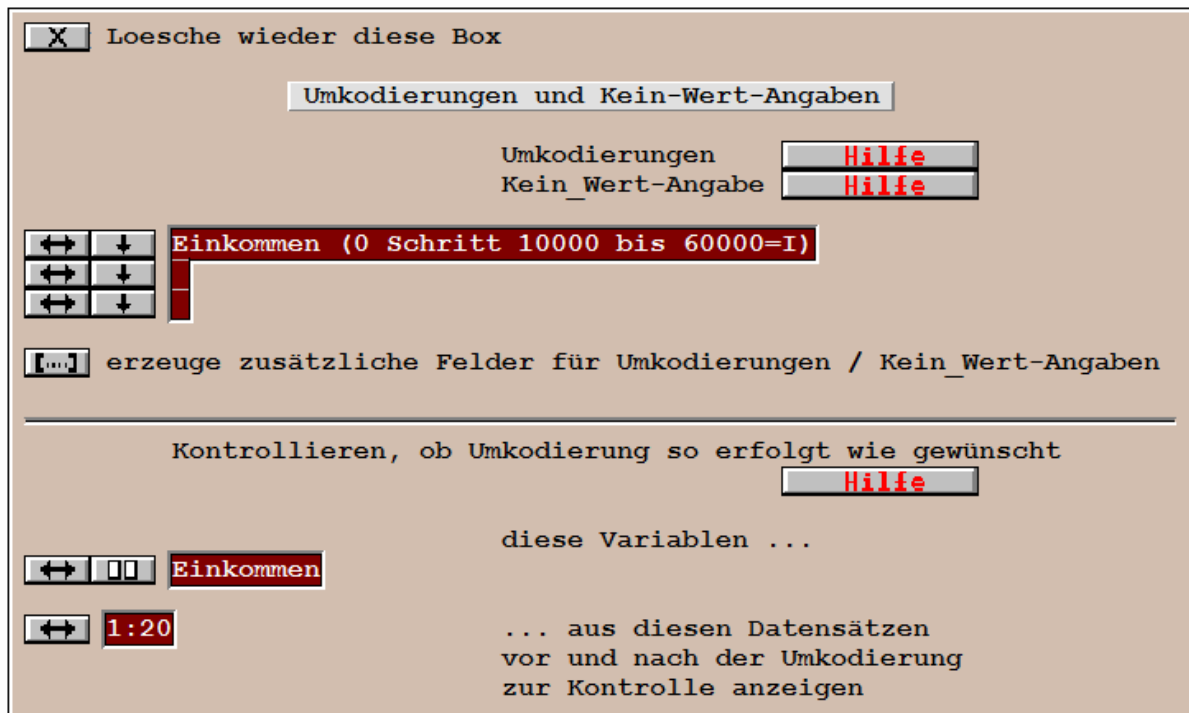
Anstelle von V.. können Sie selbstverständlich auch einen Variablennamen schreiben, z.B. Einkommen groesser 4000

Sie können auch durch bestimmte Umkodierungsanweisungen einen Ausschluß von Untersuchungseinheiten bewirken. Das ist u.U. sogar schneller. Wenn Sie z.B. in die Eingabe-Box für "Kein-Wert-Angabe und Umkodierungen" schreiben

V12(5,6 = Kein\_Wert)

dann werden die Werte 5 und 6 auf 'Kein\_Wert' gesetzt. Das hat folgende Konsequenz: Untersuchungseinheiten, die den Wert 5 oder 6 besitzen werden nicht ausgewertet. Sie werden aus der Analyse ausgeschlossen. Siehe dazu auch die Hilfe zum Umkodieren.

## P0.8 Optionsbox: Umkodierungen und Kein\_Wert-Angaben



Sehr häufig müssen Variable umkodiert werden. Betrachten wir ein Beispiel:  
Die Variable V3 Beruf hat die 5 Ausprägungen

- 1 Ungelernte Arbeiter
- 2 Facharbeiter
- 3 Angestellter
- 4 Beamter
- 5 Selbständiger

Sie wollen die Ungelernten und die Facharbeiter zusammenfassen, ebenso die Angestellte und die Beamten.

aus 1 und 2 soll also 1 werden (=Arbeiter)  
aus 3 und 4 soll 2 werden (=Angestellte/Beamte)  
aus 5 muß dann 3 werden (=Selbständige)

Die Almo-Umkodierungsanweisung für diesen Fall lautet

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

Wurde für V3 Beruf ein fehlender Wert mit 0 kodiert, so muß dies Almo mitgeteilt werden durch die Anweisung

```
Beruf( 0 = Kein_Wert )
```

Die beiden Anweisungen können zusammengefasst werden

```
Beruf( 0 = Kein_Wert; 1,2=1; 3,4=2; 5=3 )
```

Die Kein-Wert-Angabe ist also ebenfalls eine Umkodierung. Das Wort "Kein\_Wert" ist ein Schlüsselwort das von Almo in entsprechender Weise verstanden wird.

Betrachten wir dazu ein Beispiel:

Beruf soll mit Geschlecht tabelliert werden. Wird ein Datensatz eingelesen, der in Beruf den Wert 0 hat, dann wird Almo den betreffenden Datensatz nicht verwenden.

Umkodierungen und Kein-Wert-Angaben können vom Benutzer auf zwei Arten erzeugt werden:

1. Er schreibt die Anweisung selbst in das Eingabefeld
2. Mit einigen Mausklicks in einer Dialogbox bewirkt er, daß Also diese Anweisungen selbst erzeugt. Zu diesem Zweck klicken Sie in oben abgebildeter Eingabe-Box zuerst auf den Knopf mit dem nach unten weisenden Pfeil. Also präsentiert Ihnen dann eine Dialogbox, in der Sie mit einigen Klicks die Kein-Wert-Angabe vornehmen können.

### ***P0.5.1 Die "selbst geschriebene" Umkodierungs- und Kein-Wert-Anweisung***

Die "selbst geschriebene" Anweisung ist flexibler und bietet mehr Möglichkeiten. So können auch Gleichungen und "Wenn ... Dann..."-Anweisungen in das Eingabefeld geschrieben werden.

#### **P0.5.1.1 Die Kein\_Wert-Angabe**

Beispiel:

```
V3( 0 = Kein_Wert )  
Beruf( 0 = Kein_Wert )
```

Die Kein\_Wert-Angabe besteht aus

einer Variablennummer (im Beispiel: V3)  
oder einem Variablennamen (im Beispiel: Beruf)  
und einem Ausdruck in runden Klammern,  
bei dem rechts des Gleichheitszeichens "Kein\_Wert"  
steht.

Weitere Möglichkeiten für die Keinwert-Kodierung sind etwa

```
V6 ( -9 bis 0 = Kein_Wert )
```

Alle Werte von V6 zwischen -9 und 0 werden als Keinwert-Code begriffen

```
V10:27 ( 0,9 = Kein_Wert )
```

Für V10 bis V27 werden die Werte 0 und 9 als Keinwert-Code begriffen

Die Kein\_Wert-Angabe kann mit eventuellen Umkodierungen kombiniert werden.

Beispiel:

```
V5 ( 0 = Kein_Wert; 1:3 = 1; 3:10 = 2 )
```

Der Wert 0 wird auf Kein\_Wert gesetzt. 1 bis 3 wird zu 1. Und 3 bis 10 wird zu 2. Siehe dazu auch die Hilfe zum Umkodieren

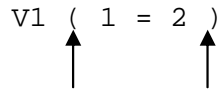
**BEACHTEN:** Alle diese Kein-Wert-Angaben ändern selbstverständlich nur die Variablenwerte während des Rechengangs, aber nicht die Variablenwerte in der Ursprungs-Datei (auf Platte oder sonstigem Datenträger).

### P0.5.1.2 Umkodierungen

Die Zahl der Umkodierungs-Möglichkeiten von Variablen in Almo ist sehr groß. Siehe dazu Handbuch, Teil 2, Abschnitt 16.

Beispiele:

V1 ( 1 = 2 )



öffnende schließende runde Klammer

Für Variable V1 soll der Werte 1 auf 2 umkodiert werden.

V1 ( 1,3,7 = 2 )

Für Variable V1 sollen die Werte 1,3,7 auf 2 umkodiert werden.

Die Umkodierungsanweisung besteht also aus

einer Variablennummer (im Beispiel: V1)  
und einem Ausdruck in runden Klammern

Anstelle der Variablennummer kann auch eine Variablenname stehen:

Beruf ( 1,3,7 = 2 )

Weitere Möglichkeiten:

V2 ( 1 bis 5 = 1; 5 bis 10 = 2; 10 bis 15=3; 15 bis 20 = 4 )

Für Variable V2 werden die Zahlen 1 bis 20 in 4 Gruppen zusammengefaßt. Liegt ein Wert genau auf der Grenze, dann wird er der unteren Gruppe zugeordnet. Das Almo-Wort 'bis' kann durch ':' (Doppelpunkt) abgekürzt werden.

V2:10 ( 1 bis 5 = 1; 5 bis 10 = 2; 10 bis 15=3; 15 bis 20 = 4 )

Diese Umkodierung wird für die Variablen V2 bis V10 durchgeführt.

V20 ( 0:2:4:6:8 = I) # I = großes i #

Bei V20 wird 0 bis 2 zu 1  
2 bis 4 zu 2  
4 bis 6 zu 3  
6 bis 8 zu 4

Lebensalter ( 1 Schritt 10 bis 100 = I )

Die Variable wird in 10-er Stufen zusammengefasst. Aus 1 bis 10 wird 1, aus 10 bis 20 wird 2, aus 20 bis 30 wird 3, ..... etc. Liegt ein Wert genau auf der Grenze, dann wird er der unteren Gruppe zugeordnet.

V8 ( 1:9 = Umdrehen )

Die Variable wird 'umgedreht'. Aus 9 wird 1  
aus 8 wird 2  
aus 7 wird 3  
.  
.  
aus 1 wird 9

V12 ( 1:3 = 1; 4,5 = 2; 6 = Kein\_Wert )



Die Werte 1 bis 3 werden zu 1. Die Werte 4 und 5 werden zu 2. Der Wert 6 wird auf 'Kein\_Wert' gesetzt. Das hat folgende Konsequenz: Untersuchungseinheiten, die den Wert 6 besitzen werden nicht ausgewertet. Sie werden aus der Analyse ausgeschlossen

```
Beruf( 1,2=1; 3,4=2; SONST=3 )
```

Durch SONST=3 werden die in der Umkodierungsanweisung nicht angegebenen Werte auf den Zahlenwert 3 umkodiert. D.h. aus den Werten 5, 6, .... etc. wird 3. SONST ist ein Almo-Schlüsselwort. Es darf auch in kleinen Buchstaben geschrieben werden.

### P0.5.1.3 Umkodieren von Zeichenvariable

Zeichenvariable können in Almo nicht ausgewertet werden. Sie müssen in Zahlenvariable umkodiert werden. Beispiel:

```
Beruf( 'Arbeiter' = 1;  
      'Angestellter' = 2;  
      'Selbstaendiger' = 3;  
      SONST=4 )
```

Eine Umkodierungsanweisung wird man normalerweise in eine Zeile schreiben. Wenn sie sehr lang ist dann kann man sie, wie man an diesem Beispiel erkennt, auch über mehrere Zeilen schreiben

Die Variable "Beruf" besitzt die Ausprägungen:

Arbeiter, Angestellter, Selbstaendiger, Beamter, Bauer

Die Ausprägungen der Zeichenvariablen müssen in obiger Umkodierungsanweisung zwischen Apostrophe geschrieben werden.

'Arbeiter' wird nun durch obige Anweisung auf den Zahlenwert 1 umkodiert, 'Angestellter' auf den Zahlenwert 2, etc.

Durch SONST=4 werden die in der Umkodierungsanweisung nicht genannten Berufe des Beamten, des Bauern etc. auf den Zahlenwert 4 umkodiert.

Beachte: SONST ist ein Almo-Schlüsselwort. Es darf nicht zwischen Apostrophe geschrieben werden.

BEACHTE: Alle diese Umkodierungen ändern selbstverständlich nur die Variablenwerte während des Rechengangs, aber nicht die Variablenwerte in der Ursprungs-Datei (auf Platte, Diskette oder sonstigem Medium).

### P0.5.1.4 Gleichungen

Sie können in das Eingabefeld für die Umkodierungen auch eine Gleichung eintragen. Betrachten wir einige Beispiele

GesamtEinkomm = MannEinkomm + FrauEinkomm ;



Semikolon nicht vergessen !

Vergessen Sie nicht das Semikolon als Abschluß der Gleichung.

```
V101 = ( V12 + V13 + V14 ) / 3 ;  
V102 = V7 ** 3 ;      # ** = bedeutet Potenzieren #
```

Siehe dazu Handbuch, Teil 2, Abschnitt 20 ff.

### P0.5.1.5 Wenn-Dann-Anweisung

Sie können in das Eingabefeld für die Umkodierungen auch eine "Wenn ... Dann ..." -Anweisung eintragen. Betrachten wir einige Beispiele

```
Wenn FamStand gleich 2 Dann
    DurchschnEinkomm = ( MannEinkomm + FrauEinkomm ) / 2 ;
EndeWenn
```

Die Wenn-Anweisung besitzt also folgende notwendige Schlüsselworte:

```
Wenn .... Dann .... EndeWenn
```

Hinter DANN steht die Anweisung, die ausgeführt werden soll.

Zwischen WENN und DANN steht ein logischer Ausdruck, üblicherweise ein Vergleich. Möglich sind z.B. folgende Vergleiche

```
Wenn      V5  gleich      7           Dann .... EndeWenn
Wenn      V6  groesser    8           Dann .... EndeWenn
Wenn      V7  kleiner     0.5         Dann .... EndeWenn

Wenn      V8  nichtgleich V9         Dann .... EndeWenn
Wenn      V10 groessergleich 2.5     Dann .... EndeWenn
Wenn      V11 kleinergleich 2        Dann .... EndeWenn
```

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

Anstelle der Worte "gleich", "groesser" usw. sind auch die üblichen mathematischen Symbole möglich, also:

```
=      >      <      ~=      >=      <=
          ↑
        nichtgleich
```

Auch UND sowie ODER sind möglich. Beispiel:

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

Die Wenn-Dann-Anweisung kann durch einen SONST-Teil erweitert werden. Beispiel:

```
Wenn V1 gleich 1 Dann V99=1; Sonst V99=2; EndeWenn
```

Wenn-Dann-Anweisungen können auch geschachtelt werden. Zur Wenn-Anweisung siehe Handbuch, Teil2, Abschnitt 29 ff.

### P0.5.2 Umkodierung und Kein-Wert-Angabe durch Mausclick

Klicken Sie zuerst in der Eingabe-Box „Kein\_Wert-Angabe und Umkodierungen“ auf den Knopf mit dem nach unten weisenden Pfeil. Almo präsentiert Ihnen dann die folgende Dialogbox, in der Sie mit einigen Klicks Umkodierungen und Kein-Wert-Angaben vornehmen können.

Betrachten wir ein Beispiel:

Die Variable V3 Beruf hat die 5 Ausprägungen

- 1 Ungelernte Arbeiter
- 2 Facharbeiter
- 3 Angestellter
- 4 Beamter
- 5 Selbständiger

Sie wollen die Ungelernten und die Facharbeiter zusammenfassen, ebenso die Angestellte und die Beamten.

Aus 1 und 2 soll also 1 werden.

Aus 3 und 4 soll 2 werden.

Aus 5 muß dann 3 werden.

Es soll also die Almo-Umkodierungsanweisung

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

entstehen

Die Vorgehensweise ist folgende:

1. Zuerst müssen Sie auf der linken Seite der obigen Dialogbox in der Listbox "umzukodierende Variable" die Variable V3 Beruf anklicken.
2. Danach klicken Sie auf den Knopf mit dem nach rechts gerichteten Pfeil. Dadurch wird die Variable in das Feld "Umkodierungs-Anweisung" transportiert. In diesem steht nun:

3. Unter der Überschrift "Alter Wert" finden Sie das Eingabefeld "Einzelwerte". In dieses schreiben Sie: 1,2
4. Unter der Überschrift "Neuer Wert" finden Sie in der gleichen Zeile ein Eingabefeld. In dieses schreiben Sie: 1

5. Dann klicken Sie auf den OK-Knopf in dieser Zeile. Daraufhin schreibt Almo in das Umkodierungs-Feld: 1,2=1  
Zusammen mit dem Variablennamen "Beruf" steht dort dann

```
Beruf(1,2=1)
```

6. In das Eingabefeld "Einzelwerte" schreiben Sie dann: 3,4 und in das Eingabefeld für den neuen Wert: 2 Wieder klicken Sie ok. Almo ersetzt dann im Feld "Umkodierungs-Anweisung" die schließende Klammer durch ein Semikolon (weil es erkennt, daß die Umkodierung fortgesetzt wird) und fügt dann hinzu: 3,4=2. Insgesamt steht dann in diesem Feld

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

7. Jetzt muß noch 5 in 3 umkodiert werden. Sie schreiben in "Einzelwerte" 5 und in das Eingabefeld für den neuen Wert 3. Dann Klick auf ok. Im Feld "Umkodierungs-Anweisung" wird dann hinzugefügt: 5=3. Insgesamt steht dann dort

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

8. Zum Abschluß klicken Sie auf den Knopf "Fertig". Almo verläßt die Dialogbox zur Umkodierung, kehrt in das Maskenprogramm zurück und schreibt dort in das Eingabefeld der Eingabe-Box „Kein\_Wert-Angabe und Umkodierungen“.

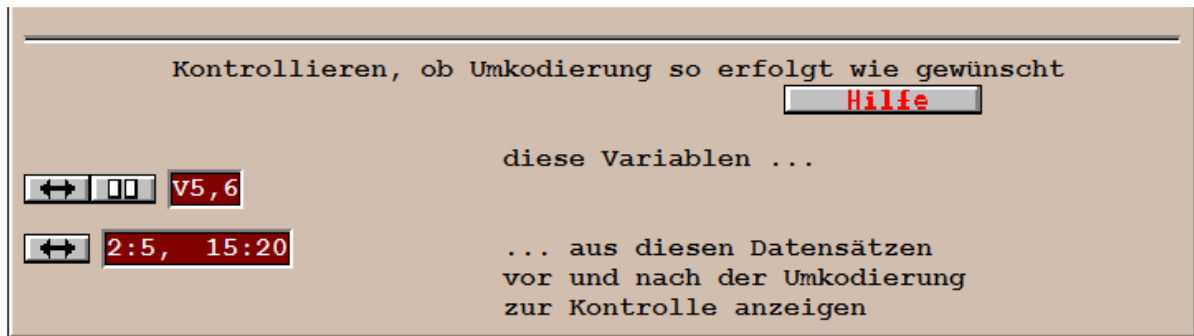
```
Beruf(1,2=1; 3,4=2; 5=3)
```

Auch das Feld "Umkodierungs-Anweisung" in der Dialogbox „Umkodieren“ ist ein EINGABE-Feld. Das bedeutet, daß Sie in diesem Feld korrigieren und löschen und schreiben können. Das ganze Feld löschen Sie, indem Sie auf den nach links weisenden Pfeilknopf klicken.

Die obige Dialogbox bietet vielfältige Umkodierungsmöglichkeiten. Sie werden ausführlich dargestellt im Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.5.5 Umkodierung und Kein-Wert-Angabe durch Mausclick.

## Sub-Box 2: Kontrollieren, ob Umkodierung so erfolgt wie gewünscht

In die 2. Sub-Box können Sie beispielsweise folgendes eingeben:



Werden Variable umkodiert, dann ist es für den Benutzer wichtig zu überprüfen, ob er die Umkodierungsanweisungen auch richtig geschrieben hat, ob also ALMO die Umkodierungen so durchgeführt hat, wie sie gewünscht worden sind.

*Eingabefeld 1:* Geben Sie in der Eingabe-Box zuerst die Variablen an, die kontrolliert werden sollen.

*Eingabefeld 2:* Dann geben Sie die Datensätze an, bei denen die Kontrolle erfolgen soll. In der Regel ist es nicht notwendig, alle Datensätze zu überprüfen.

Sie können z.B. durch die Anweisung

2:5, 15:20

eine Auswahl treffen. Der Doppelpunkt heißt "bis".

ALMO liefert dann folgende Ausgabe

Zur Kontrolle, ob Umkodierung so erfolgt wie gewünscht

-----  
Umkodierungsanweisungen:

V5,6(0=Kein\_Wert)  
V5(1:3=1; 3:10=2)  
V6(1:3=1; 3:6=2; 6:10=3)

Satz	2	:	V5,6	vor	Umkodierung=	5.1	3.1
Satz	2	:	V5,6	nach	Umkodierung=	2	2
Satz	3	:	V5,6	vor	Umkodierung=	4.2	2.2
Satz	3	:	V5,6	nach	Umkodierung=	2	1
Satz	4	:	V5,6	vor	Umkodierung=	2.1	1.2
Satz	4	:	V5,6	nach	Umkodierung=	1	1
Satz	5	:	V5,6	vor	Umkodierung=	4.2	3.2
Satz	5	:	V5,6	nach	Umkodierung=	2	2
Satz	15	:	V5,6	vor	Umkodierung=	4.1	3.2
Satz	15	:	V5,6	nach	Umkodierung=	2	2
Satz	16	:	V5,6	vor	Umkodierung=	3.2	1.2
Satz	16	:	V5,6	nach	Umkodierung=	2	1

```

Satz 17 : V5,6 vor Umkodierung= 4.2 2.2
Satz 17 : V5,6 nach Umkodierung= 2 1

Satz 18 : V5,6 vor Umkodierung= 3.2 3.2
Satz 18 : V5,6 nach Umkodierung= 2 2

Satz 19 : V5,6 vor Umkodierung= 1.1 4.1
Satz 19 : V5,6 nach Umkodierung= 1 2

Satz 20 : V5,6 vor Umkodierung= 3.1 2.2
Satz 20 : V5,6 nach Umkodierung= 2 1

```

## P0.9 Optionsbox: Untersuchungseinheiten gewichten



Wenn Sie die Optionsbox öffnen, dann können Sie beispielsweise folgendes eingeben:



Als Gewichtungvariable müssen Sie die Almo-Variable GEWICHT1 verwenden, beim Programm Prog45m4 ausnahmsweise GEWICHT0. GEWICHT0 und GEWICHT1 sind Systemvariable.

Wenn Sie z.B. in das Eingabefeld schreiben

```

Wenn V1 gleich 1 dann GEWICHT1 = 2.5; EndeWenn
Wenn V1 gleich 2 dann GEWICHT1 = 3.0; EndeWenn

```

dann wird eine Untersuchungseinheit, die in V1 den Wert 1 besitzt, so behandelt, wie wenn sie aus 2.5 Untersuchungseinheiten bestehen würde und eine Untersuchungseinheit, die in V1 den Wert 2 besitzt, so wie wenn sie aus 3 Untersuchungseinheiten bestehen würde.

Wenn Sie keine Gewichtung verwenden wollen, dann machen Sie das Eingabefeld leer. Klicken Sie dazu auf den Rein-Raus-Knopf vor dem Editierfeld

Wollen Sie jedoch eine Gewichtung vornehmen, dann können Sie alle Fälle gleich gewichten, was nicht viel Sinn macht. Sie schreiben z.B.

```
GEWICHT1 = 2.5;
```

Dadurch wird jede einzelne Untersuchungseinheit 2.5 Mal in die Analyse aufgenommen.

Sehr viel sinnvoller ist es, spezifische Untergruppen zu gewichten. Sie können z.B. alle Männer doppelt zählen (!!). Sie schreiben dann

```
Wenn Geschlecht gleich 1 dann GEWICHT1 = 2.0; EndeWenn
```

(Geschlecht=1 sei der Code für die Männer)

Oder ein anderes Beispiel:

```
Wenn Beruf gleich 1 dann GEWICHT1 = 0.5 ; EndeWenn
```

```
Wenn Beruf gleich 2 dann GEWICHT1 = 0.75; EndeWenn
```

```
Wenn Beruf gleich 3 dann GEWICHT1 = 1.25; EndeWenn
```

(wobei: 1=Arbeiter, 2=Angestellte, 3=Selbständige)

Sie müssen also zwischen WENN und DANN einen logischen Ausdruck schreiben. Möglich sind z.B. folgende Anweisungen

```
WENN V5 gleich 7 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V6 groesser 8 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V7 kleiner 0.5 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V8 nichtgleich V9 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V10 groessergleich 2.5 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V11 kleinergleich 2 DANN GEWICHT1 = 2.0; EndeWenn
```

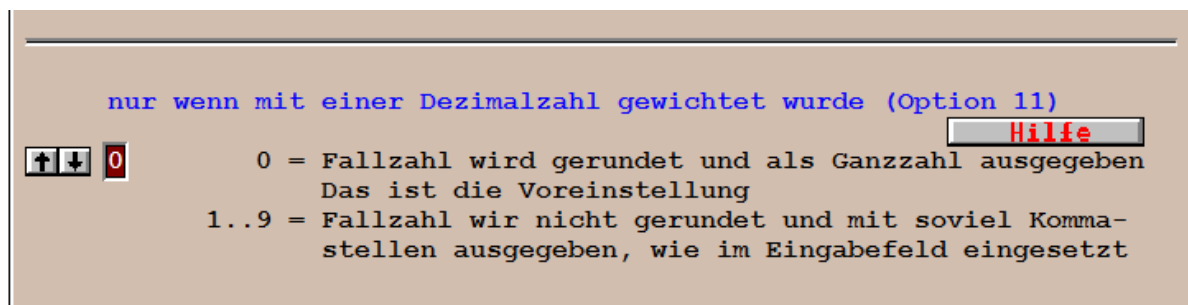
Auch UND sowie ODER sind möglich. Beispiel

```
WENN V5 gleich 7 UND V10 groesser V20 DANN GEWICHT1 = 2.0; EndeWenn
```

```
WENN V6 kleiner V5 ODER V12 nichtgleich 2 DANN GEWICHT1 = 2.0; EndeWenn
```

Zur WENN...DANN -Anweisung siehe Almo-Handbuch, Teil 1, Bedienungsanleitung, Abschnitt 29 - 31.

In einigen Masken-Programmen besitzt die Optionsbox folgende 2. Sub-Box



Wenn Sie mit einer Dezimalzahl gewichtet haben, dann können Häufigkeiten auftreten, die nicht ganzzahlig sind, sondern Dezimalstellen besitzen. Sie haben beispielsweise geschrieben

Gewicht1 = 1.25;

Dann können (gewichtete) Häufigkeiten auftreten, die Dezimalstellen besitzen, z.B. 452.75. Hier gibt es nun 2 Möglichkeiten zu verfahren

1. Möglichkeit: Wenn Sie

- a) mit einer Dezimalzahl gewichtet haben und
- b) in der 2. Sub-Box in das Eingabefeld eine 0 eingesetzt haben

dann werden die (gewichteten) Häufigkeiten auf Ganzzahlen auf- bzw. abgerundet. Ebenso werden alle sonstigen Koeffizienten aus den gerundeten ganzzahligen Häufigkeiten ermittelt. Bedingt durch dieses Runden können für die Tabellen unterschiedliche Summen entstehen

2. Möglichkeit: Wenn Sie

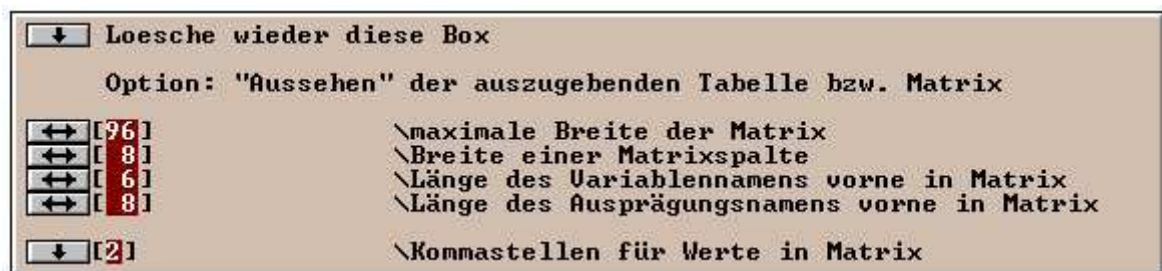
- a) mit einer Dezimalzahl gewichtet haben und
- b) in der 2. Sub-Box in das Eingabefeld eine Zahl grösser 0 (z.B. 2) eingesetzt haben

dann werden die Häufigkeiten nicht gerundet und mit (im Beispiel) 2 Kommastellen ausgegeben. Alle sonstigen Koeffizienten, die auf den Häufigkeiten beruhen, werden aus den nicht gerundeten Häufigkeiten ermittelt.

## P0.9 Optionsbox: Optionen: „Aussehen“ der auszugebenden Tabellen bzw. Matrix



Wenn Sie die Optionsbox öffnen, dann sehen Sie folgendes:



In vielen Programmen gibt Also eine Tabelle bzw. Matrix aus. Betrachten wir eine Beispieltabelle, die obige „Aussehensparameter“ verwendet:

	Produkt			Summe
	Kleidung 1	Möbel 2	Technik 3	
Geschl maennlic	91	209	213	513
weiblich	113	178	196	487



Summe	204	387	409	1000
-------	-----	-----	-----	------

Der Variablenname vorne in der Tabelle lautet nur "Geschlec", da er nur 6 Zeichen umfassen darf.

Die Ausprägungsnamen vorne dürfen nur 8 Buchstaben haben. Bei "maennlic" fehlt deswegen das abschließende "h".

Eine Tabellenspalte besitzt eine Breite von 8 Zeichen. Demzufolge besitzen auch die Ausprägungsnamen oben in der Tabelle nur maximal 8 Buchstaben. Ist ein Name länger, dann wird er abgeschnitten. Ist er kürzer, dann wird er rechtsbündig geschrieben und vorne mit Blanks aufgefüllt. Hinter jede Spalte setzt Almo eine zusätzliche Leer-Spalte, die aus einem Blank besteht.

Der Variablennamen oben in der Tabelle kann die ganze zur Verfügung stehende Breite einnehmen.

Auf einem 17-Zoll-Bildschirm kann eine Tabelle mit einer Breite von maximal 96 Zeichen noch "in einem Stück" dargestellt werden. Soll sie ausgedruckt werden, muss man als Zeichengröße 8 angeben. Ist die Tabelle insgesamt breiter als der im 1. Eingabefeld "maximale Breite ..." angegebene Wert, dann wird die Tabelle in folgender Weise umgebrochen:

	Produkt	
	Kleidung 1	Möbel 2
Geschl maennlic	91	209
weiblich	113	178
Summe	204	387

	Produkt	Summe
	Technik 3	
Geschl maennlic	213	513
weiblich	196	487
Summe	409	1000

Die vorgegebene Breite der gesamten Tabellenausgabe beträgt 96 Spalten. Unsere Beispieltabelle ist nicht so breit. Ist eine Tabelle breiter, dann wird sie senkrecht geteilt. Die Teile werden dann untereinander gestellt.

Wird die Optionsbox **nicht geöffnet**, dann verwendet Almo bestimmte Voreinstellungen, die bei den verschiedenen Maskenprogrammen verschieden sein können. In der Regel werden Variablen- und Ausprägungsnamen vorne und oben in der Tabelle mit 8 Zeichen geschrieben.

## P0.10 Eingabe-Box: Grafik-Optionen



Wenn Sie die Optionsbox öffnen, dann sehen Sie folgendes:



*Eingabefeld 1:*

Zur Erzeugung von Grafiken bietet Ihnen Almo drei Möglichkeiten an:

1. Möglichkeit

Grafik = Almo;

Es wird eine hochauflösende Grafik erzeugt. In der Ergebnisliste finden Sie einen großen Knopf mit der Aufschrift "Grafik". Wenn Sie auf diesen klicken, dann öffnet Almo das Grafikfenster und erzeugt eine Grafik.

2. und 3. Möglichkeit

Grafik = Excel;

Grafik = Stanford;

ALMO erzeugt Grafik-Tabellen, die Sie in Excel bzw. Stanford-Graphics einlesen können und dort unmittelbar in eine Graphik Ihrer Wahl wandeln können.

Bei den meisten Maskenprogrammen ist die 2. und 3. Möglichkeit nicht gegeben. Andererseits können bei einigen Maskenprogrammen einfache Grafiken im Textmodus zusätzlich angefordert werden. Diese können wie Text ausgedruckt werden.

Die Excel- und Stanford-Grafik-Tabellen werden in Dateien gespeichert. Für Excel erhalten die Dateien die Erweiterung ...XLS für Stanford-Graphics ...DAT. Der Dateiname wird Ihnen in der Ergebnisliste mitgeteilt.

*Eingabefeld 2:*

Wenn Sie '1' eingeben, dann werden die Almo-Grafiken direkt in die Ergebnisliste eingesetzt. Wenn Sie durch die Ergebnisliste blättern oder scrollen, dann werden Ihnen (anschliessend an Tabellen und Matrizen) auch die Almo-Grafiken gezeigt.

Wenn Sie '0' eingeben, dann können die Almo-Grafiken nur im Grafik-Editor angeschaut werden. In der Ergebnisliste ist dann (anschliessend an Tabellen und Matrizen) nur ein Grafikknopf enthalten. Durch Klick auf diesen Knopf gelangen Sie in den Grafik-Editor, wo Ihnen die Grafik gezeigt wird.

Der Grafikknopf ist auch vorhanden, wenn Sie '1' eingeben, wenn also die Grafiken in die Ergebnisliste eingesetzt werden. Durch Klick auf den Grafikknopf in der Ergebnisliste können die Grafiken dann im Grafik-Editor bearbeitet werden und von dort durch Klick auf den Knopf "Einsetzen" in der veränderten Form wieder in die Ergebnisliste übergeben werden. Eine Bearbeitung der Grafik wird häufig notwendig sein. Man möchte beispielsweise die Balken in einem Balkendiagramm schlanker abgebildet haben, als dies Almo standardmäßig tut. Oder man möchte mehr Perspektive in die Grafik bringen. Oder man möchte noch zusätzliche Beschriftungen einfügen etc.

Durch Klick auf den Knopf 'Grafik löschen' kann die Grafik gelöscht werden. Man wird dies tun, wenn man einen kurzen und kompakten Output haben möchte. Man kann die Grafik aber wieder zurückholen, wenn man wieder auf den Grafikknopf klickt.

Sie können die Ergebnisliste (insgesamt oder in Blöcken) inklusive der eingelagerten Almo-Grafiken ausdrucken oder in eine andere Datei speichern. Einen Block erzeugen Sie durch "Strg+Mausklick" auf die obere und auf die untere Blockzeile. Siehe Menü "Hilfe / 5.Teil".

Auch können Sie die Ergebnisliste (insgesamt oder in Blöcken) inklusive der eingelagerten Almo-Grafiken exportieren. Klicken Sie dazu im Menü 'Datei' auf den Eintrag '*Export : Speichern für Textverarbeitungsprogramm*'. Almo erzeugt dann ein Dokument im rtf-Format, das in MS Word (inklusive der Grafiken) geladen werden kann. Das funktioniert aber nur, wenn Sie bei der Installation von WORD den Grafikfilter für Metafiles miteinbezogen haben. Das ist der Normalfall.

Technische Anmerkung: Die Almo-Grafiken werden als "enhanced metafiles" in die Ergebnisliste eingesetzt und von dort aus gedruckt und exportiert. Damit wird die bestmögliche Bildqualität erreicht.