ALMO
Statistiksystem

Pfadanalyse
als wiederholte Regressionsanalyse

P25

Kurt Holm

www.almo-statistik.de
holm@almo-statistik.de
kurt.holm@jku.at

2015

Weitere Almo-Dokumente
Die folgenden Dokumente können alle kostenlos 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 Mehrfach-antworten (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

Inhaltsverzeichnis

P25 Pfadanalyse

P25.1 Einführung

Eine Pfadanalyse kann (1) als wiederholt angewandte Regressionsanalyse oder (2) als Strukturgleichungs-Modell gerechnet werden. Almo enthält ein Programm zur 1. Form der Pfadanalyse. Diese wird im folgenden Text dargestellt. Zur 2. Form siehe etwa Reinecke (2014). Die 1. Form der Pfadanalyse wird in der Literatur auch *explorative Pfadanalyse* und die 2. Form *konfirmatorische Pfadanalyse* genannt.

Das Almo-Pfadanalyse-Programm wurde von Kurt Holm geschrieben, der auch diesen Text verfasst hat.

Wissenschaftliche Erklärung bedarf der "Ursache-Wirkungs-Aussage". Es genügt nicht festzustellen, dass zwischen zwei Variablen ein Zusammenhang besteht. Es muss auch die Richtung des Zusammenhanges angegeben werden. Bei der nicht-experimentellen Forschung werden zwar Zusammenhänge entdeckt, deren Ursache-Wirkungs-Richtung bleibt jedoch oft unbekannt. Der Forscher muss diese Richtung postulieren. Das geschieht in der sogenannten Pfadanalyse durch eine bestimmte Anordnung der Variablen.

Wir wollen annehmen, zwischen 3 Variablen V1, V2 und V3 bestünde folgender Kausalzusammenhang:

$$V1 \longrightarrow V2 \longrightarrow V3$$

V1 bewirkt ursächlich V2 und V2 bewirkt ursächlich V3. Die Pfeile symbolisieren eine Ursache-Wirkungs-Relation. Die Korrelationskoeffizienten könnten dann etwa folgende sein:

	V1	V2	V3
V1		0.5	0.3
V2			0.6
V3			

Wir können diese Korrelationsmatrix auch graphisch darstellen. Die gestrichelten Linien symbolisieren nun keine kausalen, sondern nur korrelative Verknüpfungen.

$$\begin{tikzcd} V1 \arrow[r, ".5"] \arrow[rr, dash, dotted, bend right, ".3"'] & V2 \arrow[r, ".6"] & V3 \end{tikzcd}$$

Wenn V1 nicht selbst direkt auf V3 einwirkt - und das haben wir ja ausgeschlossen - wird es trotzdem mit V3 korrelieren und zwar (etwas verkürzt dargestellt) mit dem Produkt 0.5*0.6=0.30. Wenn nun die empirisch ermittelte Korrelation zwischen V1 und V3 =0.5 ist, dann ist dieser Koeffizient sozusagen um 0.2 zu groß. In diesem Falle muss ein direkter Ursache-Wirkungszusammenhang von V1 nach V3 bestehen. Das Kausalmodell ist dann:

$$\begin{tikzcd} V1 \arrow[r] \arrow[rr, bend right] & V2 \arrow[r] & V3 \end{tikzcd}$$

Der Korrelationskoeffizient von .5 entsteht also, weil V1 mit V3 indirekt (über V2) und zusätzlich direkt verbunden ist. Ein Korrelationskoeffizient, das haben wir jetzt gelernt, gibt also keine sichere Auskunft darüber, wie der direkte Zusammenhang beschaffen ist. Einmal lag dem Korrelationskoeffizienten zwischen V1 und V3 in der Größe von .3 überhaupt kein direkter Zusammenhang zugrunde. Ein anderes Mal bestand der V1-V3-Koeffizient von .5 aus 2 Komponenten, einer direkten und einer indirekten.

Wir wollen Kausalzusammenhänge entdecken. D.h. wir wollen direkte Zusammenhänge feststellen. Und wir wollen wissen, wie stark die jeweiligen direkten Zusammenhänge sind.
Die Pfadanalyse ist eine statistische Methode, die uns diese Information liefert. Dabei müssen wir allerdings eine bestimmte Anordnung der Variablen postulieren, d.h. wir müssen bestimmte Ursache-Wirkungs-Relationen behaupten.

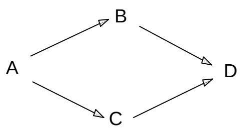
Die Koeffizienten, die uns die Stärke des direkten Zusammenhanges angeben, werden Pfadkoeffizienten genannt. Wir können abschließend folgenden Zusammenhang zwischen Korrelations- und Pfadkoeffizienten feststellen.

$$(1)\ r=p+i$$

r = Korrelationskoeffizient
p = Pfadkoeffizient
i = indirekter Effekt

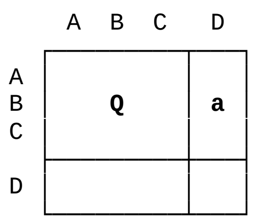
Der Algorithmus der Pfadanalyse
Wir wollen hier den Rechengang der Pfadanalyse zunächst nur überblicksweise darstellen. Anschließend werden wir ihn in aller Ausführlichkeit vortragen. Wir wollen annehmen, unser Untersuchungsgegenstand sei in 4 quantitativ messbare Variable A, B, C und D aufgegliedert. Diese 4 Variablen liegen in standardisierten Werten vor.

1) Zunächst muss entschieden werden, in welcher kausalen Folge die Variablen hintereinander gereiht werden sollen. Wir wollen annehmen, folgendes Kausalmodell sei plausibel.

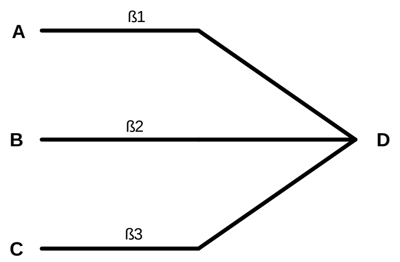


Wir entschließen uns also für die Reihenfolge A, B, C, D. Da B und C auf gleicher Kausalstufe stehen, hätten wir auch die Reihe A, C, B, D bilden können. Wir müssen uns jedoch für eine der beiden Reihen (zunächst) entscheiden. Es ist durchaus möglich, später die Pfadanalyse mit der alternativen Reihung durchzuführen. (Wenn B und C tatsächlich auf der gleichen Kausalstufe stehen, dann dürfen die beiden Reihungen keine signifikant verschiedenen Pfadkoeffizienten erbringen.)

2) Die 4 Variablen werden interkorreliert. Es entsteht die Korrelationsmatrix **R**, die wir uns untergliedert denken in eine Submatrix **Q** und einen Spaltenvektor **a**



3) Die Pfadanalyse wird nun als wiederholt angewandte, multiple Regression durchgeführt. Wir betrachten D als die abhängige Variable und A, B und C als die unabhängigen Variablen. Graphisch dargestellt

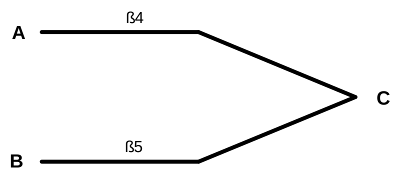


Die drei Regressionskoeffizienten (Pfadkoeffizienten) ß1 bis ß3 erhalten wir durch folgenden Kalkül

(2) **ß** = inv(**Q**)***a**

ß ist der Spaltenvektor der (drei) ß-Koeffizienten ß1,ß2,ß3. Der Spaltenvektor **a** enthält die Korrelationen aller unabhängigen mit der einen abhängigen Variablen. **a** ist die letzte Spalte der Korrelationsmatrix **R** ohne das letzte Spaltenelement.
Q ist die Sub-Korrelationsmatrix der unabhängigen Variablen. **Q** erhält man, indem man die letzte Zeile und Spalte aus **R** streicht. Die Matrix **Q** muss dann noch zur Matrix inv(**Q**) invertiert werden. Gemäß unserer Annahme über die kausalen Zusammenhänge müsste der Koeffizient ß1 für den direkten Zusammenhang A nach D =0 sein.

4) Derselbe Rechengang wird nun für C als abhängige und A und B als unabhängige Variable durchgeführt.



Die Koeffizienten werden wieder nach folgendem Kalkül berechnet

(3) **ß** = inv(**S**)***b**

ß ist der Spaltenvektor der (beiden) ß-Koeffizienten ß4 und ß5. **b** ist die letzte Spalte der **Q**-Matrix (die ihrerseits durch Streichung der letzten Spalte und Zeile aus der **R**-Matrix entstand) ohne das letzte Spaltenelement.
S ist die Submatrix der Korrelationen zwischen den unabhängigen Variablen A und B. Sie entsteht durch Streichung der letzten Spalte und Zeile der **Q**-Matrix. **S** muss noch zu inv(**S**) invertiert werden. Gemäß unserer Modellannahme müsste ß5, der Zusammenhang B nach C, gleich 0 sein.

5) Derselbe Rechengang wird nun entsprechend für B als abhängige und A als unabhängige Variable wiederholt.

6) Die Pfadanalyse kann schließlich für eine oder mehrere konkurrierende Reihungen der Variablen wiederholt werden.

P25.1.1 Definition der Pfadanalyse

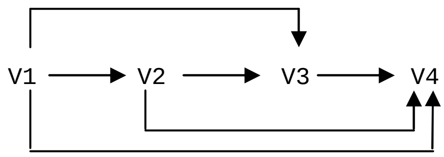
Die Pfadanalyse ist eine Methode, mit deren Hilfe eine Korrelationsmatrix von Variablen in ein rekursives Kausalmodell überführt wird. Wir wollen diese kurze Definition in einzelne Elemente auflösen und ergänzen.

1) Das Ausgangs-Datenmaterial ist eine Interkorrelationsmatrix verschiedener Variablen. Die Variablen können quantitativ und/oder ordinal und/oder nominal-dichotom sein. Ordinale Variable in die Pfadanalyse einzuführen ist nicht unproblematisch. Wir haben dies im Almo-Dokument 5 "Korrelation", Abschnitt 19.0.4 (von Hermann Denz) und im Anhang von Heinrich Potuschak "Der allgemeine Korrelationskoeffizient" diskutiert.
2) Es wird vermutet, dass diese Variablen kausal verknüpft sind.
3) Dabei unterstellen wir zunächst, dass die Variablen in einer Kette hintereinander gereiht sind und dass jede Variable von allen vor ihr stehenden Variablen kausal bestimmt wird.
4) Das Kalkül, der eine Interkorrelationsmatrix in ein volles rekursives System überführt, ist die "wiederholt angewandte multiple Regression".
5) Die Pfadkoeffizienten sind dann identisch mit dem standardisierten partiellen Regressionskoeffizienten.
6) Es ist nun möglich, dass verschiedene Regressionskoeffizienten (=Pfadkoeffizienten) nicht signifikant von 0 verschieden sind. Dadurch vereinfacht sich das kausale Modell der Variablen.
7) Der Kalkül der Pfadanalyse, d.h. der Kalkül der wiederholt angewandten multiplen Regression, kann mit verschiedenen konkurrierenden Variablen-Reihungen durchgeführt werden.

P25.1.2 Das "volle" rekursive, lineare System

Ein volles rekursives System liegt vor, wenn jede nachgeordnete Variable Pfeile von allen ihr vorgeordneten Variablen empfängt.

Betrachten wir ein Beispiel mit 4 Variablen:



Jede Variable erhält von jeder vorausgehenden Variablen einen Pfeil

P25.1.3 Die Pfadanalyse als wiederholt angewandte Regression

Die ALMO-Pfadanalyse ist eine auf ein volles rekursives Variablen-Modell wiederholt angewandte Regression. Betrachten wir ein Beispiel mit 6 Variablen V1,2,...6. Wir werden dieses Beispiel anschließend mit Prog25m2 rechnen und in Abschnitt P25.1.8 die Ergebnisse ausführlich besprechen. Das untere Dreieck der (symmetrischen) Interkorrelationsmatrix der 6 Variablen ist folgendes (ohne letzte Spalte für V6):

	V1	V2	V3	V4	V5
V1	1.00				
V2	.34	1.00			
V3	.13	.42	1.00		
V4	.15	.42	.59	1.00	
V5	.07	.24	.10	.05	1.00
V6	.01	.06	.07	.01	.41

Wir betrachten V6 zunächst als abhängige Variable. Die anderen vor ihr gereihten Variablen V1 bis V5 sind die unabhängigen Variablen. Für dieses Modell führen wir eine multiple Regression durch. Danach streichen wir die Variable V6 aus unserer Betrachtung. Wir betrachten nun V5 als die abhängige Variable und alle vor ihr angereihten Variablen als die unabhängigen Variablen. Die Streichung führen wir so durch, dass wir aus der Korrelationsmatrix R die letzte Spalte und Zeile streichen. Auf diese Weise wird die R-Matrix "abgearbeitet". Immer die letzte Zeile und Spalte wird gestrichen - bis wir schließlich nur noch eine 2x2-Matrix haben.

P25.1.4 Die Zahl der Input-Variablen

In der Pfadanalyse ist es üblich, jene Variablen, die am Anfang der Reihung stehen, ihrerseits also nicht mehr durch andere Modellvariable bestimmt sind, als input-Variable (oder exogene Variable) zu bezeichnen. Die übrigen Variablen sind dann endogene Variable. Wenn man - wie wir das in unserem 6-Variablen-Modell getan haben - die Korrelationsmatrix bis zu einer 2x2- Matrix "abarbeitet", dann gibt es natürlich nur eine input-Variable. Es ist aber nun zulässig, die "Abarbeitung" früher zu beenden.

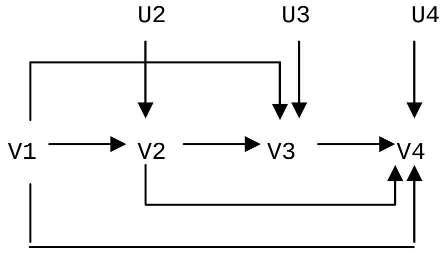
P25.1.5 Der Pfadkoeffizient der Restvariablen

Eine abhängige Variable ist durch die ihr vorgeordneten unabhängigen Variablen nicht voll determiniert. Der quadrierte multiple Korrelationskoeffizient R^2 gibt an, wieviel Prozent an Varianz in der abhängigen Variablen durch die unabhängigen Variablen erklärt werden.

Wir können nun unterstellen, dass es ein uns unbekanntes Variablenbündel U gibt, das noch auf die abhängige Variable einwirkt. Wir betrachten U wie eine einzige Variable und nennen sie "Restvariable". Der Koeffizient β_u , d.h. der Pfadkoeffizient dieser Restvariablen bezüglich der abhängigen Variablen ist $\beta_u = 1 - R^2$.

Für jede Variable, die im vollen rekursiven System den Status einer abhängigen Variablen einnimmt, und das sind alle bis auf die 1. Variable in der Reihenfolge der Variablen, kann nun ein solcher β_u - Koeffizient berechnet werden. Jede Variable, bis auf die 1., besitzt also eine Restvariable.

In unserem 4-Variablen-Modell würde das so aussehen:



P25.1.6 Eingabe in Maskenprogramm P25m1

Almo stellt 2 Maskenprogramme zur Verfügung, eines bei dem die Daten von Untersuchungseinheiten eingelesen werden und eines bei dem eine fertige, schon vorhandene Korrelationsmatrix eingelesen wird. Letzteres wird im nachfolgenden Abschnitt P25.1.7 dargestellt.

Prog25m1.Msk

Pfadanalyse

Auf ein rekursives Kausalmodell wird die Regressionsanalyse wiederholt angewendet

Beispiel: Es wird vermutet, dass die soziale Herkunft (V5) die Schulbildung (V6) determiniert, diese die berufliche Leistung (V3) und diese das Einkommen (V7).

Die vermutete Kausal-Reihung der Variablen ist also folgende

```
           V5    -->    V6    -->    V3     -->   V7
        sozHerk --> SchBild --> Leistg    --> Einkomm
```

Almo rechnet ein volles rekursives System: Für jede Variable werden alle in der Reihung vor ihr stehenden als unabhängige Variable eingesetzt. Almo rechnet folgende Regressionsanalysen:

	abhäng. Var	unabhäng. Var
1. Regress.analyse	V7	V5, V6, V3
2. Regress.analyse	V3	V5, V6
3. Regress.analyse	V6	V5

Es wird also folgendes Modell ermittelt

```
            ┌────────────────────────────────────────┐
            │┌──────────────────────────┐            │
            ││                          │            │
       sozHerk────────SchBild──────Leistg─────────Einkomm
                          │                          │
                          └──────────────────────────┘
```

Wenn Regressionskoeffizienten nicht signifikant oder zu klein sind dann können die entsprechenden Pfade aus dem Pfaddiagramm herausgelöscht werden. Das Modell vereinfacht sich dadurch.

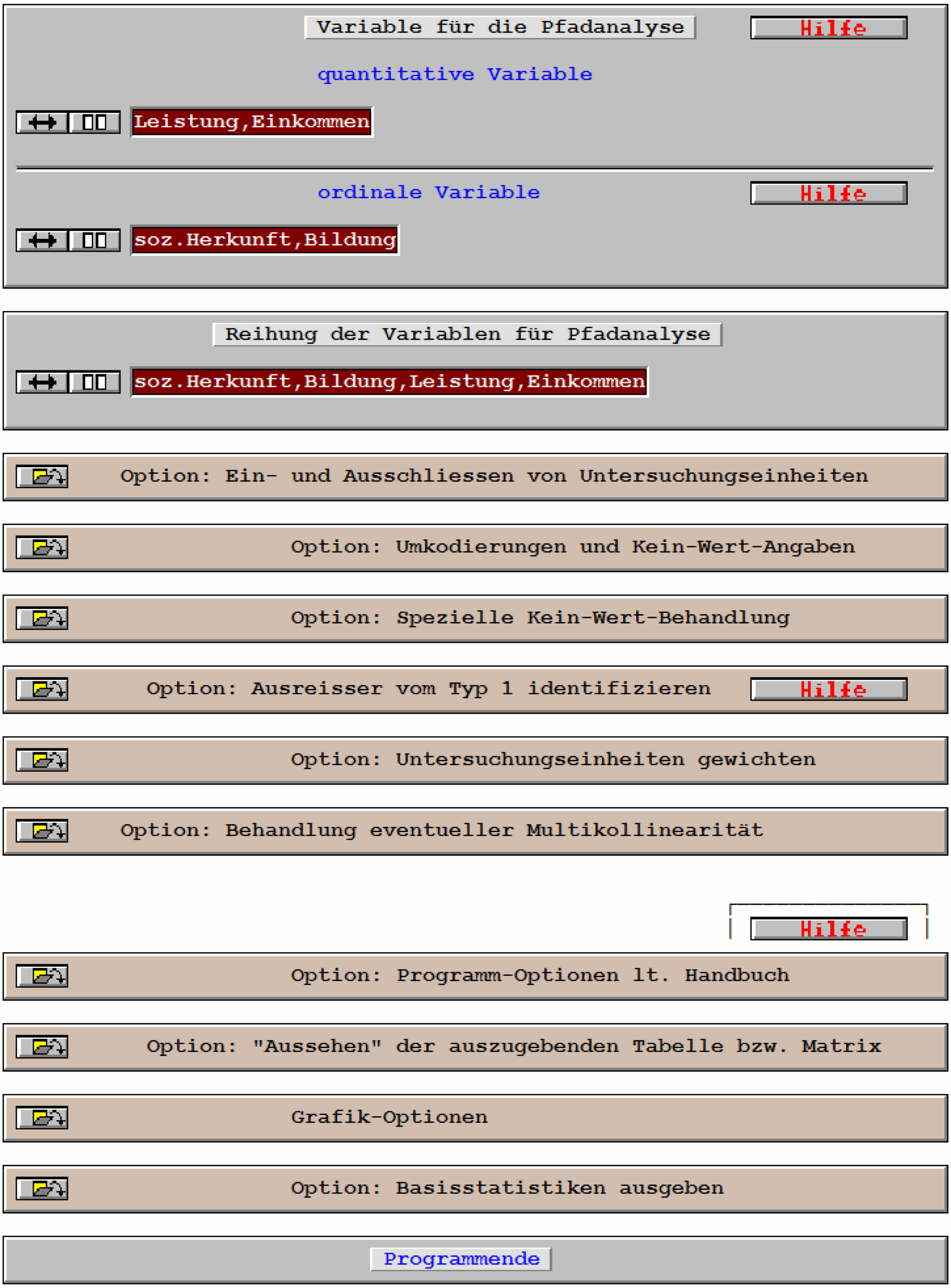
Siehe Handbuch, Teil 4, Abschnitt P25

Programm-Bedienung ---> Hilfe

Speicher fuer x Variable Hilfe

Vereinbare Variable= 20

Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert



P25.1.6.1 Erläuterungen zu den Boxen

Eingabe-Box: Vereinbare Variable
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.1.

Eingabe-Box: Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.2.

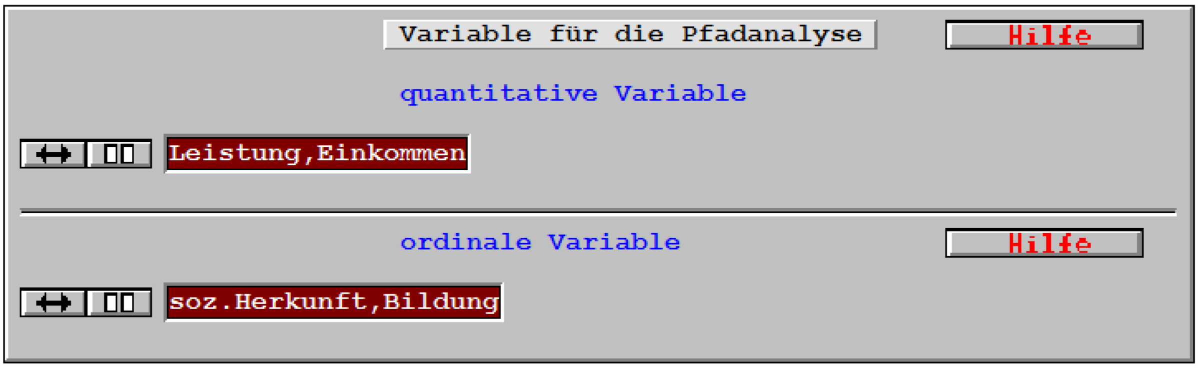
Eingabe-Box: Datei der Variablennamen
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.3.

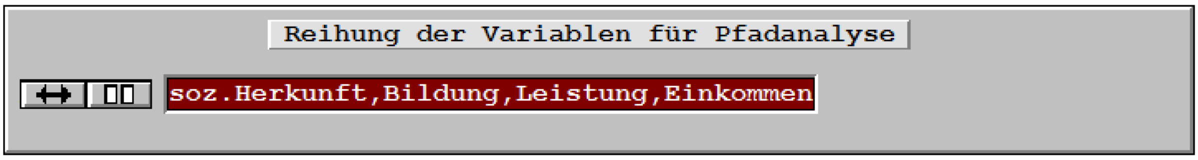
Eingabe-Box: Freie Namensfelder
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.3.

Eingabe-Box: Datei aus der gelesen wird
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.4.

Eingabe-Box: Wenn Dateiformat FIX oder nicht Standard-FREI
Siehe Almo-Dokument Nr.0 "Arbeiten mit Almo", Abschnitt P0.4.

Eingabe-Box: Variable für die Pfadanalyse





Geben Sie hier die Variablen an, für die Sie eine Pfadanalyse rechnen wollen; zuerst die quantitativen in die 1. Eingabebox und dann die ordinalen in die 2. Eingabebox. In der 3. Eingabebox muss die Reihenfolge eingegeben werden, in der die Variablen vermutlich kausal hintereinander stehen.

Befinden sich ordinale Variable in der Analyse, dann berechnet Almo die Korrelationsmatrix (auf die es dann den Kalkül der Pfadanalyse anwendet) nach dem „Groß-Gamma-Kalkül". Siehe dazu das Almo-Dokument 5 "Korrelation". Dabei werden dann auch Signifikanzen über den F- bzw. t-Test ermittelt. Für die ordinalen Variablen ist das sehr problematisch. Wir würden eher dazu raten, ordinale Variable – sofern sie genügend Ausprägungen besitzen – als quantitative zu behandeln.

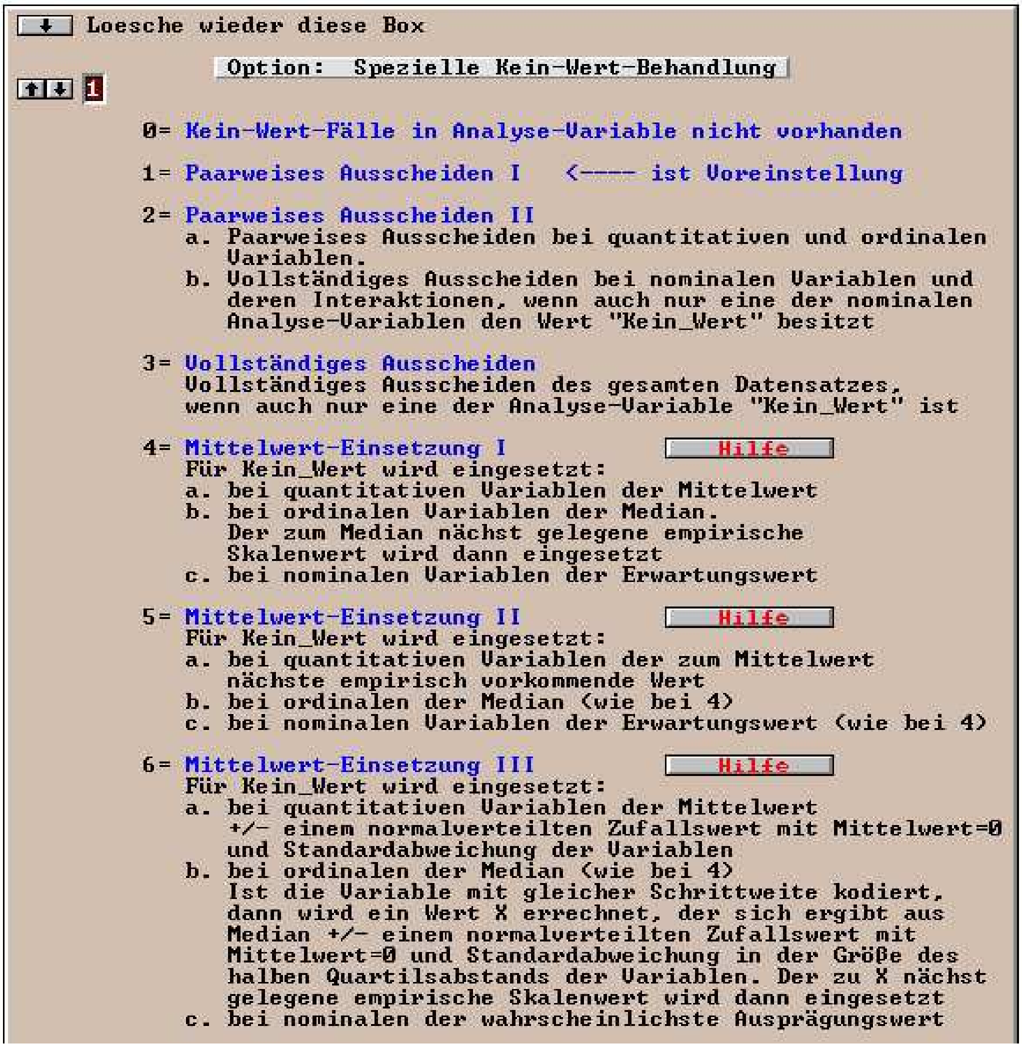
Nominal-dichotome Variable können wie quantitative behandelt werden. Selbstverständlich ist es möglich, polytome Variable durch Umkodierung zu dichotomisieren.

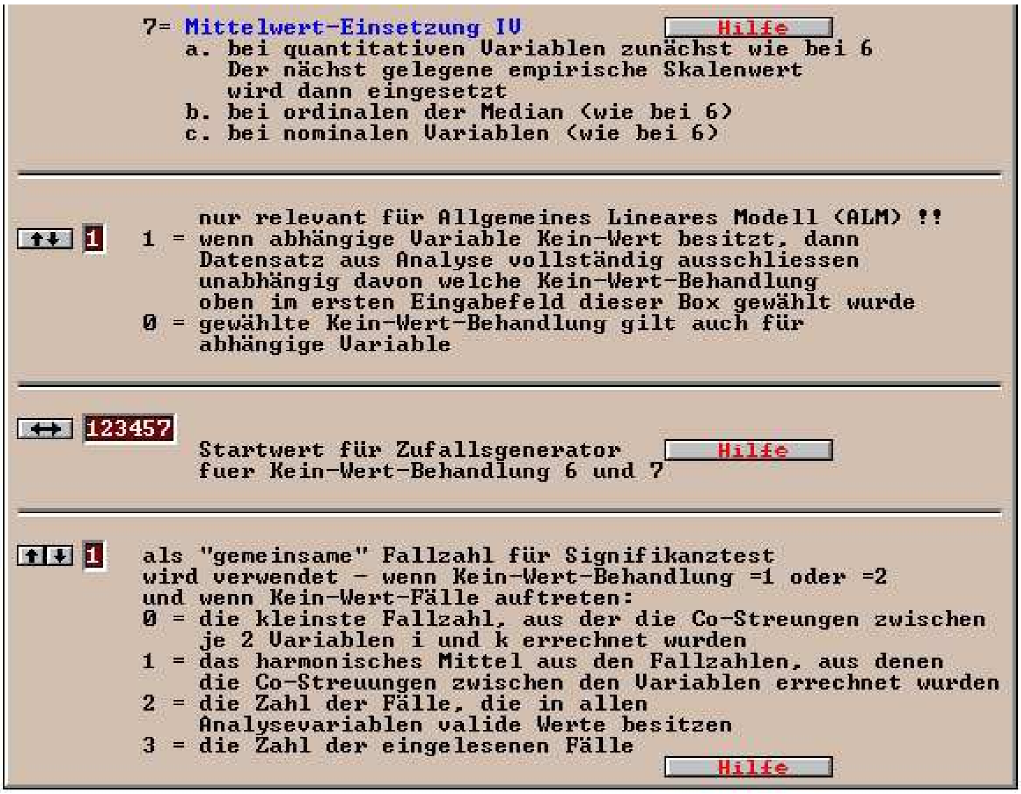
Eingabe-Box: Option: Ein- und Ausschliessen von Untersuchungseinheiten
Siehe P0.7.

Eingabe-Box: Kein_Wert-Angabe und Umkodierungen
Siehe P0.5.

Eingabe-Box: Option: Spezielle Kein-Wert-Behandlung

Besitzt eine oder mehrere Analysevariablen keinen Wert, dann verwendet Almo standardmäßig das "paarweise Ausscheiden". Der Benutzer hat die Möglichkeit eine von 7 Methoden zur Kein-Wert-Behandlung zu wählen. Dazu muss die Optionsbox geöffnet werden. Man sieht dann folgende große Box.





Beachte: In der Box wird auch die Kein-Wert-Behandlung nominaler Variablen angeboten. Diese ist hier irrelevant, da in der Pfadanalyse keine nominalen Variablen zugelassen sind. Zulässig sind nominale-dichotome Variable, die aber vom Programm wie quantitative behandelt werden.

Kein-Wert-Behandlung 1: "Paarweises Ausscheiden“
Wir werden dieses Verfahren im Handbuch P45 „Data Mining“, in Abschnitt P45.12.4 sehr ausführlich darstellen. Hier wollen wir es nur kurz beschreiben.

Betrachten wir die Korrelationsmatrix zwischen den 3 Variablen V1, V2 und V3.

	V1	V2	V3
V1	r_{11}	r_{12}	r_{13}
V2		r_{22}	r_{23}
V3			r_{33}

Jeder einzelne Korrelationskoeffizient r_{ik} für die beiden Variablen i und k wird nur aus den Untersuchungseinheiten errechnet, für die aus beiden Variablen i und k valide Werde vorhanden sind. In die Diagonale der Korrelationsmatrix wird 1.0 eingesetzt. Die Folge dieser Vorgehensweise ist, dass die verschiedenen Korrelationskoeffizienten aus verschiedenen Fallzahlen berechnet sind. Almo ermittelt standardmäßig das harmonische Mittel aus den verschiedenen Fallzahlen und verwendet dieses für Signifikanztests.

Kein-Wert-Behandlung 2: „Paarweises Ausscheiden II“
Wie „Paarweises Ausscheiden I“.

Kein-Wert-Behandlung 3: „Vollständiges Ausscheiden“
Vollständiges Ausscheiden des gesamten Datensatzes, wenn auch nur eine der ursächlichen Analyse-Variable "Kein_Wert" ist.

Kein-Wert-Behandlung 4: Mittelwert-Einsetzung I
Almo ermittelt zuerst Mittelwerte (für quantiative Variable), Median (für ordinale Variable) und den Erwartungswert (für nominale Variable).
Almo gibt diese Werte aus.

Für Kein_Wert wird eingesetzt:

a) bei quantitativen Variablen der Mittelwert

b) bei ordinalen Variablen der Median (=der mittlere Wert)
Liegt der Median nicht auf einem empirischen Wert, sondern zwischen 2 empirschen Werten, dann wird der nächst gelegene Nachbarwert als KW-Einsetzungswert verwendet.

Kein-Wert-Behandlung 5: Mittelwert-Einsetzung II
Für Kein_Wert wird eingesetzt:
a. bei quantitativen Variablen der zum Mittelwert nächste empirsch vorkommende Wert
b. bei ordinalen Variablen der Median wie bei Kein-Wert-Behandlung 4

Kein-Wert-Behandlung 6: Mittelwert-Einsetzung III
Für Kein_Wert wird eingesetzt:

a. bei quantitativen Variablen der Mittelwert +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung der Variablen.
Wir könnten auch formulieren: Es wird ein normalverteilter Zufallswert mit Mittelwert und Standardabweichung der Variablen eingesetzt.

b. bei ordinalen Variablen der Median.

Ist die Variable (was eher ungewöhnlich ist) mit ungleichen Schrittweiten kodiert (z.B. 1, 2, 5, 6, 23), dann wird der Median eingesetzt.
Liegt dieser zwischen zwei empirisch vorkommenden Werten, dann wird der zum Median nächst gelegene empirische Wert verwendet.

Ist die Variable mit gleicher Schrittweite kodiert, dann wird ein Wert X errechnet, der sich ergibt aus Median +/- einem normalverteilten Zufallswert mit Mittelwert=0 und Standardabweichung in der Größe des halben Quartilsabstands der Variablen. Der zu X nächst gelegene empirische Skalenwert wird dann eingesetzt.

Bei quantitativen und bei ordinalen Variablen wird also eine normalverteilte Zufallszahl mit Mittelwert=0 generiert.
Als Standardabweichung wird bei quantitativen Variablen die der jeweiligen Variablen verwendet. Bei ordinalen Variablen wird der halbe Quartilsabstand verwendet.

Betrachten wir ein Beispiel: Die quantitative Variable sei das Lebensalter. Almo errechnet für sie einen Mittelwert von 40 und eine Standardabweichung von 20. Dann wird eine normalverteilte Zufallszahl mit Mittelwert=0 und Standardabweichung=20 erzeugt. Nehmen wir an es entsteht der Zufallswert -15.25. Für den fehlenden Wert wird dann eingesetzt X = 40-15.25 = 24.75.

Bei einer ordinalen Variablen wird entsprechend verfahren. Als Standardabweichung für die Generierung der Zufallszahl wird der halbe

Quartilsabstand verwendet. Der ermittelte X-Wert wird bei der ordinalen Variablen aber noch nicht als KW-Einsetzungswert verwendet. Es wird nach dem empirisch vorkommenden Wert gesucht, der am dichtesten bei X liegt. Dieser wird als KW-Einsetzungswert verwendet. So wird verhindert, dass KW-Einsetzungswerte entstehen, die empirisch nicht vorkommen.

Kein-Wert-Behandlung 7: Mittelwert-Einsetzung IV
Für Kein_Wert wird eingesetzt:

a. Bei quantitativen Variablen:
   Es wird zunächst ein Wert X errechnet, der sich ergibt aus dem Mittelwert +/- einem normalverteilten Zufallswert mit Mittelwert=0 und der Standardabweichung der Variablen. Dann wird der zu X nächst gelegene empirische Skalenwert für Kein_Wert eingesetzt. So wird verhindert, dass KW-Einsetzungswerte entstehen, die empirisch nicht vorkommen.

b. bei ordinalen Variablen wie bei Kein-Wert-Behandlung 6

Kein-Wert-Behandlung 4 und 5 unterscheiden sich von 6 und 7 dadurch, dass bei 6 und 7 eine Zufallsvariation dem Mittelwert bzw. Median hinzugefügt wird.

Die Kein-Wert-Behandlung 4 unterscheiden sich von 5 nur dadurch dass für die quantitativen Variablen ein Mal der Mittelwert und das andere Mal der zum Mittelwert nächste empirisch vorkommende Wert als KW-Einsetzungswert verwendet wird.

Warum Zufallswert hinzufügen?

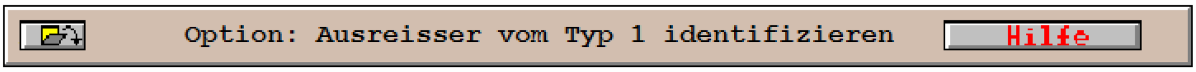
Es muss noch folgende Frage beantwortet werden: Warum wird der Mittelwert bzw. der Median bei Kein-Wert-Behandlung 6 und 7 durch einen Zufallswert überlagert?

Wird als KW-Einsetzungswert nur der Mittelwert (bzw. der Median) verwendet, dann wird die Varianz der Variablen verringert, weil für Kein-Wert immer derselbe Wert eingesetzt wird.

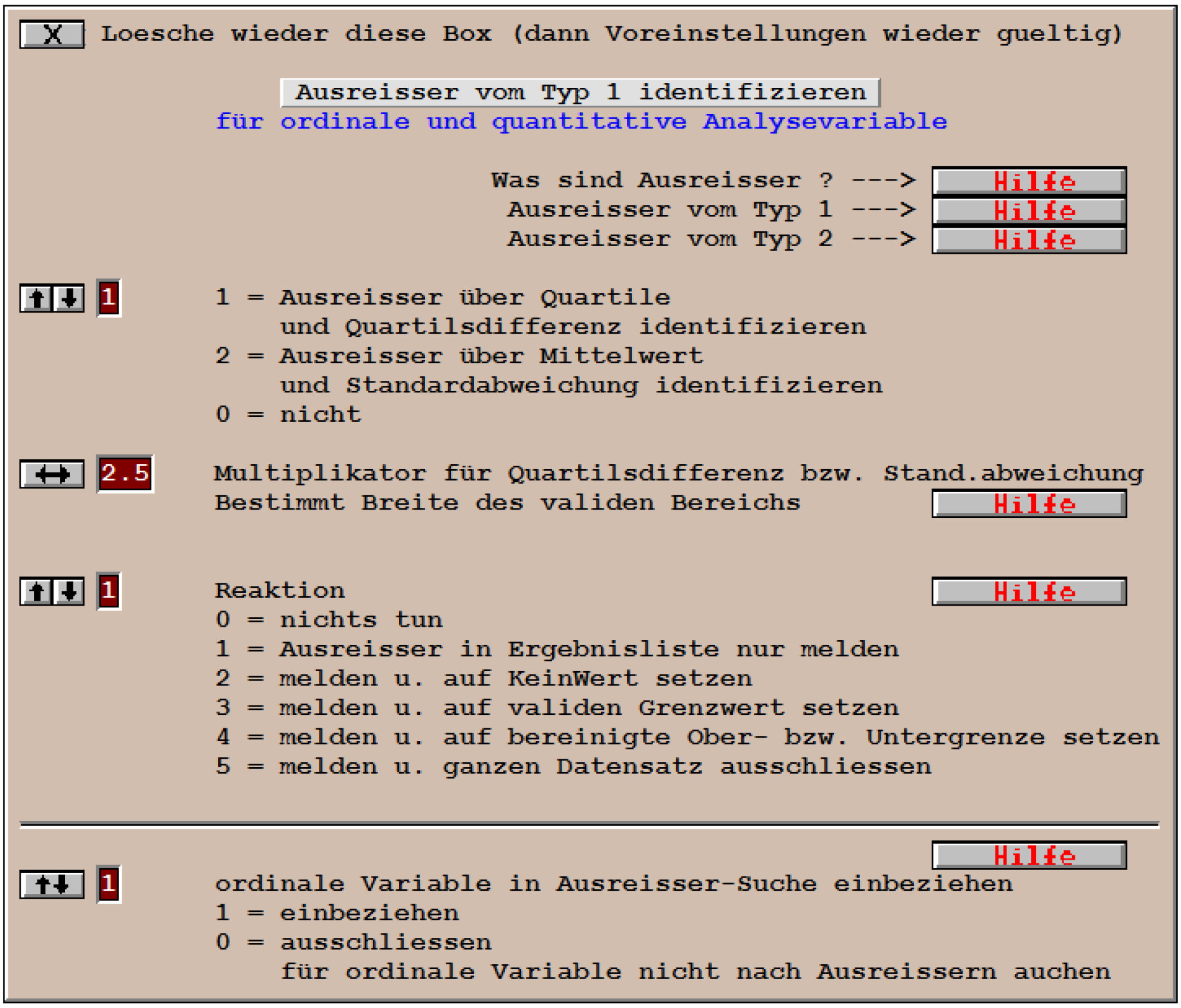
Werden mit den so erzeugten „vollständigen“ Daten beispielsweise Korrelationen errechnet, dann werden die Signifikanzen dieser Korrelationen überschätzt. Siehe dazu etwa R. J. A. Little/D. B. Rubin (1990, S. 381).

Die Überlagerung durch einen normalverteilten Zufallswert mit der Standardabweichung der Variablen bezweckt also, dass die Varianz der Variablen (fast) unverändert bleibt. Gleiches gilt auch für nominale Variable. Der Erwartungswert der Variablen ist immer derselbe. Dadurch wird die Varianz verringert. Durch den "wahrscheinlichsten" Wert bleibt die Streuung (fast) unverändert.

Eingabe-Box: Option: Ausreisser identifizieren



Wird die Optionsbox geöffnet, dann sieht man folgendes:

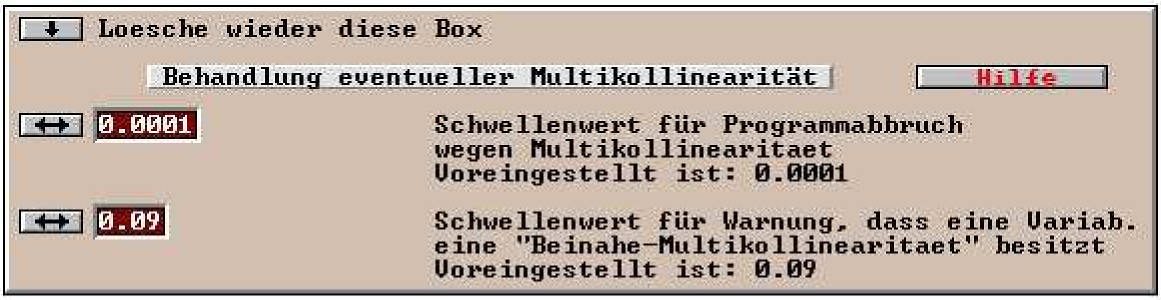


Ausreisser können die Ergebnisse erheblich beeinflussen. Siehe die ausführliche Darstellung im Almo-Dokument 23 "Ausreisser entdecken".

Eingabe-Box: Option: Untersuchungseinheiten gewichten
Siehe P0.8.

Eingabe-Box: Behandlung eventueller Multikollinearität

Optionsbox geöffnet:



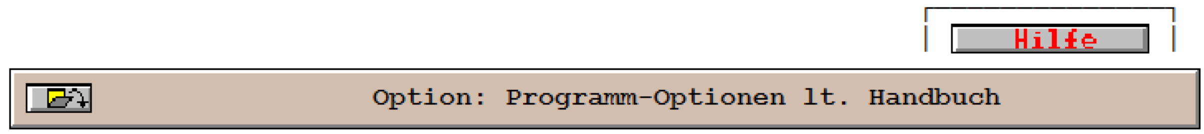
Wir empfehlen, die in den beiden Eingabefeldern enthaltenen Zahlenwerte zunächst nicht zu ändern.

Almo überprüft die Korrelationsmatrix **Q** der unabhängigen Variablen auf lineare Abhängigkeiten. Die einfachste Form einer linearen Abhängigkeit ist gegeben, wenn 2 unabhängige Variable mit 1.0 (oder beinahe 1.0) miteinander korrelieren. Es kann aber auch sein, dass eine Variable durch 2 oder mehrere andere Variable vollständig (oder beinahe vollständig) determiniert wird. Die lineare Abhängigkeit ist dann an den Korrelationskoeffizienten nicht erkennbar. In diesem Fall ist die Korrelationsmatrix nicht invertierbar und der Regressionskalkül nicht rechenbar. Almo bricht das Programm ab.

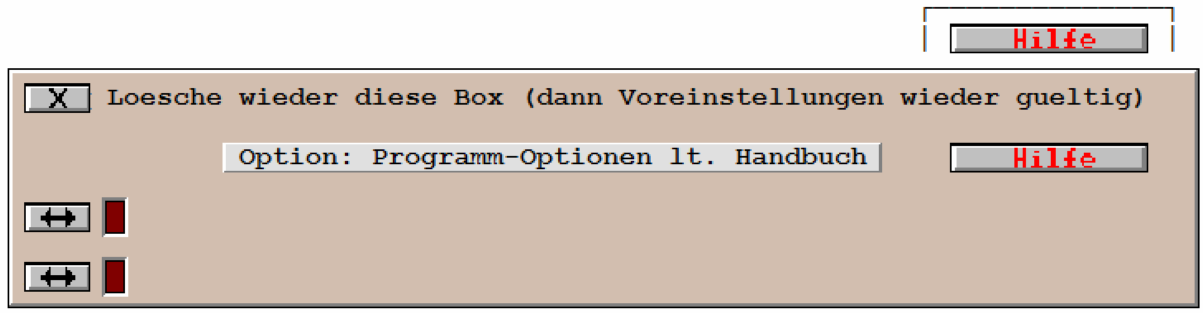
Almo erkennt eine Variable i als von anderen linear abhängig, wenn ihr Diagonalglied aus der Cholesky-Matrix kleiner ist als 0.0001. Der Wert 0.0001 ist voreingestellt. Dieser Schwellenwert ist sehr klein. Er wird eine vollständige lineare Abhängigkeiten in den meisten Fällen identifizieren können. Er ist jedoch zu klein um eine "Beinahe" Multikollinearität" zu entdecken.
Der Schwellenwert kann in der Box des Maskenprogramms verändert werden. Wenn Sie einen größeren Wert, z.B. 0.005 oder sogar 0.01 einsetzen, dann wird eine "Beinahe"-Multikollinearität früher entdeckt.

Almo gibt eine Warnung, wenn das Cholesky-Diagonalglied kleiner ist als 0.09. Der Wert 0.09 ist voreingestellt. Er kann in der Box verändert werden.

Eingabe-Box: Programm-Optionen lt. Handbuch



Box geöffnet:



Hier kann der Benutzer Optionen einsetzen, die nicht über eine der anderen Optionsboxen aktivierbar sind.

Sie können mehrere Angaben in ein Eingabefeld schreiben. Achten Sie aber darauf, dass Sie nicht Optionen einsetzen, die bereits über eine der Optionsboxen des Programms aktiviert wurden oder im Widerspruch zu diesen stehen.

Folgende Optionen sind möglich

OPTION 2 = x;
Wenn Sie z.B. OPTION 2=4; schreiben, dann wird die Korrelationsmatrix bis zu 4 Variablen "abgearbeitet", anders gesprochen: die Korrelationsmatrtix wird bis zu einer 4x4 Matrix abgearbeitet. Die Voreinstellung ist Option 2=2; dies braucht also nicht geschrieben zu werden.

OPTION 3 = 0;
Zusätzlich zu den standardisierten Regressionskoeffizienten werden auch die

nicht-standardisierten berechnet.
Option 3=1; Es werden nur die standardisierten Regressionskoeffizienten berechnet. Dies ist die Voreinstellung (braucht also nicht geschrieben zu werden.

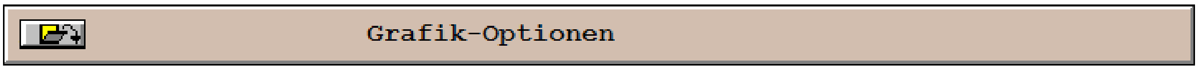
SA_Nenner = -1;
Bei der Berechnung der Standardabweichung soll im Nenner n-1 verwendet werden. Hinter "SA_Nenner=...;" kann vom Benutzer eine beliebige negative oder positive Zahl eingesetzt werden. Eine negative Zahl wird von n subtrahiert und eine positive zu n addiert. Die Voreinstellung ist "SA_Nenner=0;" d.h. im Nenner steht n.
Diese Option wirkt sich nicht auf die Korrelationsmatrix aus - nur auf die in der Ergebnisliste ausgegebenen Standardabweichungen.

MATRIX = QUASI_KORRELATION;
Die Korrelationsmatrix wird nach dem besonderen Kalkül der "Quasi-Korrelationsmatrix" berechnet. Das führt nur dann zu anderen Korrelations-Werten, wenn Kein-Wert-Fälle (fehlende Werte) auftreten.

Im Prinzip wären es auch möglich "Matrix=Kovarianz;" oder Matrix=Quadratsumme;" einzugeben. Dann würden allerdings nicht standar-disierte Padkoeffizienten, die nicht miteinander vergleichbar wären, berechnet werden.

SCHREIBE ERGEBNIS_MATRIX IN DATEI 5 ".\Progs\KorrPfad.mat";
Durch diese Anweisung im Programmparameter-Block wird die Matrix, auf die die Pfadanalyse angewendet wird in die angegebene Datei gerettet. Der Dateiname ist selbstverständlich beliebig. Die Datei-Nummer sollte zwischen 5 und 15 liegen.

Eingabe-Box "Aussehen" der auszugebenden Tabelle bzw. Matrix
Eingabe-Box Grafik-Optionen



Siehe zu diesen beiden Optionsboxen das Almo-Dokument 0 "Arbeiten mit Almo", Abschnitt P0.7.3 und P0.7.4.

Eingabe-Box: Basisstatistiken ausgeben

Optionsbox geöffnet:

Es können zusätzlich Basisstatitiken ausgegeben werden. Dies sind u.a.

Mittelwerte
Standardabweichungen
Zahl der diversen Werte je Variable
Zahl der fehlenden Werte je Variable

P25.1.7 Eingabe in Maskenprogramm P25m2 (Eingabe einer fertigen Korrelationsmatrix)

Dieses Programm entspricht dem bereits dargestellten Prog25m1 - mit dem Unterschied, dass nicht Untersuchungseinheiten eingelesen werden, sondern eine fertige, schon vorhandene Korrelationsmatrix.

```
                    Prog25m2.Msk
   Pfadanalyse mit Eingabe einer fertigen Korrelationsmatrix

Auf ein rekursives Kausalmodell wird die Regressionsanalyse
wiederholt angewendet

Beispiel:
Bei einer Befragung von Schülern wurden folgende Variable
erhoben:
         V1  Schichtzugehörigkeit
         V2  Durchschnittsnote aus Kernfächern
         V3  Beliebtheit in der Klasse
         V4  Einfluss in der Klasse
         V5  Schulzufriedenheit
         V6  Leistungswille

Die vermutete Kausal-Reihung der Variablen ist folgende
         V1 -> V2 -> V4 -> V3 -> V5 -> V6

BEACHTE: In der einzulesenden Korrelationmatrix stehem die
Variablen in der Reihenfolge V1,2,3,4,5,6 hintereinander.
Bei der vermuteten Kausal-Reihung sind jedoch V3 und V4
vertauscht. Das ist zulässig. Almo ordnet die Korrelations-
matrix entsprechend der vom Benutzer vorgegebenen Reihung um.

Almo rechnet ein volles rekursives System: Für jede Variable
werden alle in der Reihung vor ihr stehenden als unabhängige
Variable eingesetzt. Almo rechnet folgende Regressionsanalysen:

                      abhäng. Var     unabhäng. Var
1. Regress.analyse        V6            V1 bis V5
1. Regress.analyse        V5            V1, V2, V4, V3
1. Regress.analyse        V3            V1, V2, V4
1. Regress.analyse        V4            V1, V2
1. Regress.analyse        V2            V1

Wenn Regressionskoeffizienten nicht signifikant oder zu klein
sind dann können die entsprechenden Pfade aus dem Pfaddiagramm
herausgelöscht werden. Das Modell vereinfacht sich dadurch.
                                 Siehe Handbuch, Abschnitt P25
```

```
               Was ist ein Kurzprogramm ? --> Hilfe
                              Bedienung   --> Hilfe
```

1
```
              Speicher fuer x Variable       Hilfe

Vereinbare Variable= 20 ;
```

2
```
        Option: Weitere Vereinbarungen - nur wenn Almo dazu auffordert
```

3
```
              Datei der Variablennamen       Hilfe



        zeige                          zeige = Namensdatei in Output zeigen
                                       leer  = nicht
```

4


5


6


7

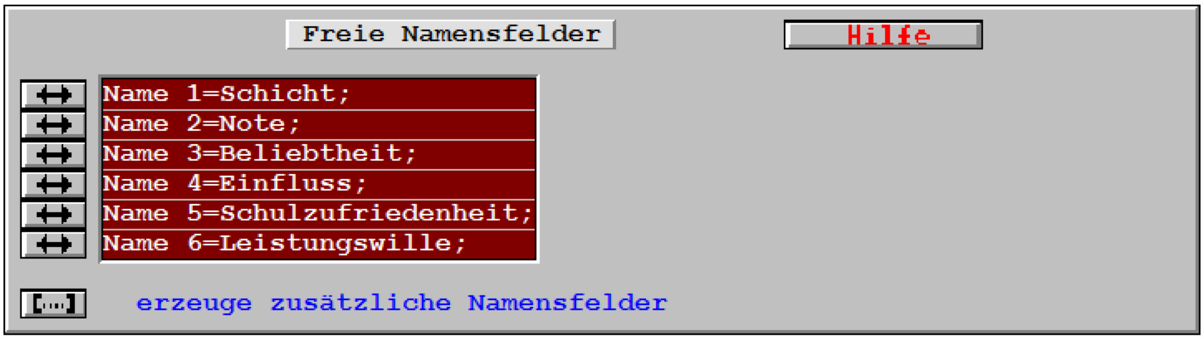

8


9


P25.1.7.1 Erläuterung zu den Boxen: Variablennamen, Form der einzugebenden Matrix

Die Boxen sind dieselben wie bei Prog25m1. Es sollen nur die folgende beiden Eingabeboxen erläutert werden:.

Eingabe-Box 4: Freie Namensfelder



Die Variablen-Nummern bzw. die Namens-Nummern müssen der Position der Variablen in der eingelesenen Matrix entsprechen. Also: Die Variable Schicht steht in der 1. Zeile/Spalte der Matrix, Note in der 2. Zeile/Spalte der Matrix, Beliebtheit in der 3. Zeile/Spalte usw.

Eingabe-Box 5: Datei aus der die Korrelationsmatrix gelesen wird

Geben Sie den vollen Pfad- und Dateinamen der Datei an, in der sich die fertige Korrelationsmatrix befindet. Die zu lesende Matrix muss in folgender Form in der Datei enthalten sein. Siehe dazu Handbuch, Teil 2, Abschnitt 43.

Beispiel für eine eingegebene Korrelationsmatrix

```
 *                           Stern

 487                         Zahl der Fälle, aus der die
                             Korrelationsmatrix gebildet wurde

 *                           Stern

 1.0
 0.4 1.0
 0.3 0.7 1.0                 untere Dreiecksmatrix
 0.1 0.2 0.7 1.0             mit Diagonale
 0.2 0.3 0.2 0.3 1.0
 0.3 0.2 0.2 0.1 0.8 1.0

 *                           Stern
 *                           Stern
```

Da die Korrelationsmatrix symmetrisch ist, wird nur das untere Dreieck einschließlich der Diagonale geschrieben.
Wichtig ist es die Zahl der Fälle anzugeben aus der die Korrelationsmatrix errechnet wurde. Ein Stern an dieser Stelle würde zu einem Programm-Abbruch führen.
Die Zahl der Kommastellen, mit der die Korrelationsmatrix geschrieben wird, ist beliebig.

P25.1.8 Ergebnisse der Pfadanalyse

Wir wollen die Ergebnisse aus Prog25m1 der auf das volle rekursive System angewandten multiplen Regression vortragen.

Die Daten sind einer Befragung von 751 12-und 13-jährigen Schülern entnommen. Siehe dazu Kurt Holm: Soziale Schicht und Schulverhalten in Hartfiel/Holm: Bildung und Erziehung in der Industriegesellschaft, UTB, 47. Die Variable "Schicht" wurde über den Beruf des Vaters bzw. der Mutter erfragt. Die Schulnote wurde aus den Zeugniszensuren der Kernfächer bei den Schülern selbst erfragt. Einfluss und Beliebtheit des Schülers im Klassenverband wurden durch einen soziometrischen Test erkundet. Schulzufriedenheit und Leistungswille wurden durch umfangreiche Fragebatterien ermittelt. Diese beiden *theoretischen*, *latenten* Variablen wurden faktorenanalytisch sorgfältig auf Eindimensionalität überprüft. Die Variablen dieser Untersuchung sind ihrem wissenschaftstheoretischen Status nach also empirisch als auch theoretisch. Auch die verwendeten Messverfahren sind sehr unterschiedlich. Es wäre interessant zu untersuchen, ob eine Pfadanalyse durch ein Strukturgleichungsmodell möglich ist und wenn ja, zu welchem Ergebnis sie führen würde.

Die Ausgabe aus Almo wird hier gekürzt wiedergegeben:

```
Diagonalglieder der  Choleskymatrix zur Ermittlung linearer Abhaengigkeiten

Variable
  1 Schicht                 1.0000
  2 Note                    0.8844
  4 Einfluss                0.8235
  3 Beliebtheit             0.6156
  5 Schulzufriedenheit      0.9382
  6 Leistungswille          0.8274
```

```
keine lineare Abhaengigkeiten
```

**************** **Erläuterung:**
Almo hat in der Korrelationsmatrix keine linearen Abhängigkeiten entdeckt

```
========================================================================

Koeffizienten fuer abhaengige Variable V6  Leistungswille

quadr.mult.Korr.koeff.          0.173
mult.Korr.koeff.                0.415
F-Wert f.mult.Korr.koeff.      31.042
Signif.niveau (1-p)*100        99.999
Schaetzfehler f.abh.Variable
bzw.Reg.koeff.f.Restvar.        0.910

Variable                      Regr.    Standard    t-Wert   Signifikanz
                              koeff    abweichg             (1-p)*100
  1 Schicht                 -0.0051     0.0355     0.1428     11.40
  2 Note                    -0.0560     0.0407     1.3766     83.11
  4 Einfluss                -0.0269     0.0426     0.6307     47.19
  3 Beliebtheit              0.0682     0.0425     1.6041     89.10
  5 Schulzufriedenhe         0.4183     0.0344    12.1502    100.00

Variable                  erklaerte   partieller    indirekte
                           Streuung   Korr.koeff      Effekte
  1 Schicht                  0.0000    -0.0052         0.0151
  2 Note                     0.0021    -0.0504         0.1160
  4 Einfluss                 0.0004    -0.0231         0.0369
  3 Beliebtheit              0.0029     0.0587         0.0018
  5 Schulzufriedenhe         0.1642     0.4069        -0.0083
```

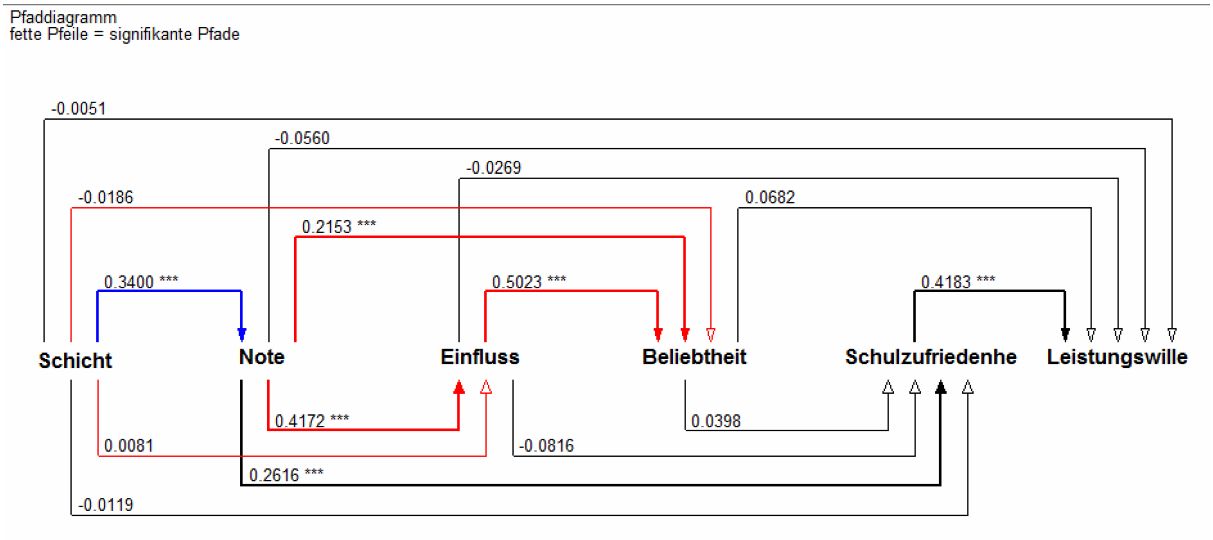
**************** **Erläuterung:**
Die abhängige Variable ist der Leistungswille der Schüler. Die einzige signifikante unabhängige Variable ist die Schulzufriedenheit. Alle anderen haben eine Signifikanz (1-p)*100 von unter 95 %. Wir werden diese Variablen nicht in das Pfaddiagramm aufnehmen, das wir zum Schluß zeichnen werden.

```
========================================================================

Koeffizienten fuer abhaengige Variable V5  Schulzufriedenheit

quadr.mult.Korr.koeff.          0.062
mult.Korr.koeff.                0.249
F-Wert f.mult.Korr.koeff.      12.279
Signif.niveau (1-p)*100        99.999
Schaetzfehler f.abh.Variable
bzw.Reg.koeff.f.Restvar.        0.969

Variable                      Regr.    Standard    t-Wert   Signifikanz
                              koeff    abweichg             (1-p)*100
  1 Schicht                 -0.0119     0.0377     0.3146     24.57
  2 Note                     0.2616     0.0422     6.1915    100.00
  4 Einfluss                -0.0816     0.0452     1.8036     92.86
  3 Beliebtheit              0.0398     0.0452     0.8801     62.10

Variable                  erklaerte   partieller    indirekte
                           Streuung   Korr.koeff      Effekte
  1 Schicht                  0.0001    -0.0115         0.0819
  2 Note                     0.0483     0.2212        -0.0216
  4 Einfluss                 0.0041    -0.0659         0.1316
  3 Beliebtheit              0.0010     0.0322         0.0602
========================================================================

Koeffizienten fuer abhaengige Variable V3  Beliebtheit

quadr.mult.Korr.koeff.          0.384
```

```
mult.Korr.koeff.                0.620
F-Wert f.mult.Korr.koeff.     155.281
Signif.niveau (1-p)*100        99.999
Schaetzfehler f.abh.Variable
bzw.Reg.koeff.f.Restvar.        0.785

Variable                      Regr.    Standard    t-Wert   Signifikanz
                              koeff    abweichg             (1-p)*100
  1 Schicht                 -0.0186     0.0305     0.6077    45.67
  2 Note                     0.2153     0.0333     6.4703   100.00
  4 Einfluss                 0.5023     0.0317    15.8698   100.00

Variable                  erklaerte   partieller    indirekte
                           Streuung   Korr.koeff      Effekte
  1 Schicht                  0.0003    -0.0222         0.1486
  2 Note                     0.0345     0.2305         0.2047
  4 Einfluss                 0.2078     0.5024         0.0877
========================================================================

Koeffizienten fuer abhaengige Variable V4  Einfluss

quadr.mult.Korr.koeff.          0.176
mult.Korr.koeff.                0.420
F-Wert f.mult.Korr.koeff.      80.029
Signif.niveau (1-p)*100        99.999
Schaetzfehler f.abh.Variable
bzw.Reg.koeff.f.Restvar.        0.907

Variable                      Regr.    Standard    t-Wert   Signifikanz
                              koeff    abweichg             (1-p)*100
  1 Schicht                  0.0081     0.0353     0.2306    18.09
  2 Note                     0.4172     0.0353    11.8173   100.00

Variable                  erklaerte   partieller    indirekte
                           Streuung   Korr.koeff      Effekte
  1 Schicht                  0.0001     0.0084         0.1419
  2 Note                     0.1540     0.3969         0.0028
========================================================================

Koeffizienten fuer abhaengige Variable V2  Note

quadr.mult.Korr.koeff.          0.116
mult.Korr.koeff.                0.340
F-Wert f.mult.Korr.koeff.      97.771
Signif.niveau (1-p)*100        99.999
Schaetzfehler f.abh.Variable
bzw.Reg.koeff.f.Restvar.        0.940

Variable                      Regr.    Standard    t-Wert   Signifikanz
                              koeff    abweichg             (1-p)*100
  1 Schicht                  0.3400     0.0344     9.8879   100.00

Variable                  erklaerte   partieller    indirekte
                           Streuung   Korr.koeff      Effekte
  1 Schicht                  0.1156     0.3400        -0.0000
========================================================================
```

P25.1.9 Das Pfaddiagramm

Almo zeichnet folgendes Pfaddiagramm:

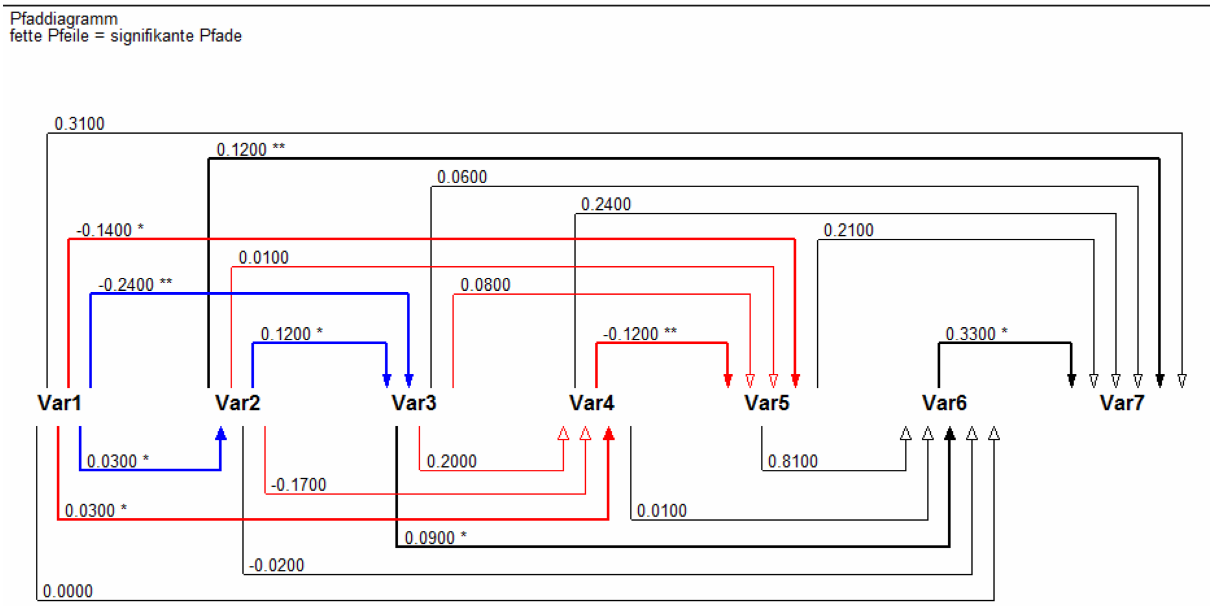


Interpretation:
Es existiert eine durchgehende "Achse" von der Schichtzugehörigkeit zum Leistungswillen.
Die soziale Schicht bestimmt die Durchschnittsnote. Diese bestimmt die Schulzufriedenheit. Und diese bestimmt den Leistungswillen der Schüler. Zwischen Note und Leistungswille besteht kein direkter Kausalzusammenhang, jedoch ein indirekter, der über die Zufriedenheit vermittelt wird. Der Einfluß, den ein Schüler in seiner Klasse hat, ist durch seine Note determiniert. Sein Einfluß bestimmt dann, wie beliebt er bei den anderen Schülern ist.

P25.1.9.1 Pfaddiagramm bearbeiten

Das Pfaddiagramm kann im Almo-Grafik-Editor in vielfältiger Weise bearbeitet werden.

Betrachten wir als Beispiel nachfolgendes Pfaddiagramm



Das Pfaddiagramm umfasst 7 Variable. Um die Pfeile optisch besser unterscheiden zu können werden Farben verwendet.

Betrachten wir die Pfeile im oberen Teil der Grafik.

```
Die Pfeile, die zu Var7 führen sind schwarz.
Die Pfeile, die zu Var5 führen sind rot
Die Pfeile, die zu Var3 führen sind blau
```

Im unteren Teil der Grafik gilt entsprechendes

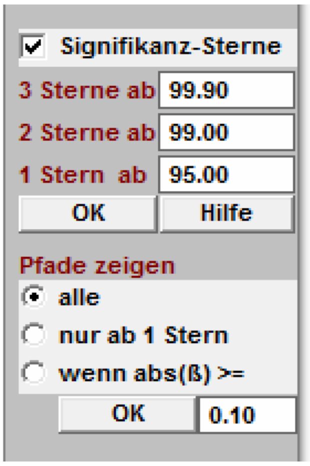
```
Die Pfeile, die zu Var6 führen sind schwarz.
Die Pfeile, die zu Var4 führen sind rot
```

Auf den Pfeilen stehen die (standardisierten) Regressionskoeffizienten. Signifikante Pfade sind durch 1 oder 2 oder 3 Sterne markiert. Die entsprechenden Pfeile sind fett gezeichnet. Standardmäßig werden an den Regressionskoeffizient angehängt:

```
   1 Stern wenn seine Sigifikanz (1-p)*100 >= 95 %    bzw. p <= 0.05
   2 Stern wenn seine Sigifikanz (1-p)*100 >= 99 %    bzw. p <= 0.01
   3 Stern wenn seine Sigifikanz (1-p)*100 >= 99,9 % bzw. p <= 0.001
```

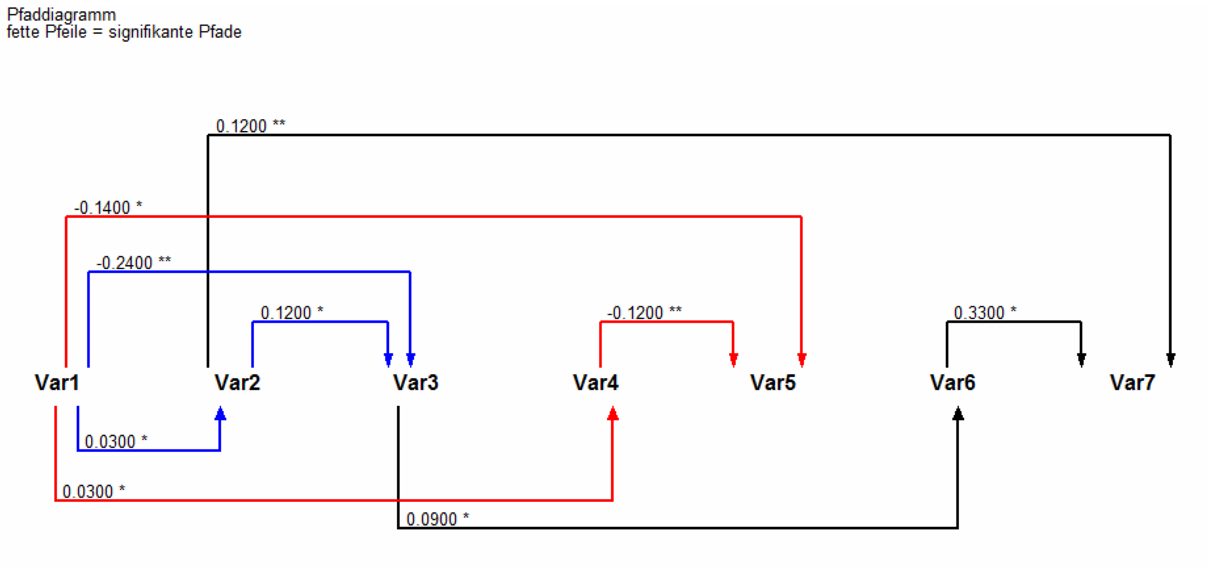
Pfade (Pfeile) eliminieren

Im Grafik-Editor in der rechten Werkzeug-Leiste werden folgende Möglichkeiten angeboten, das Pfaddiagramm zu verändern



Der Benutzer kann die Signifikanz je Stern anders definiert. Auch besteht die Möglichkeit, nicht-signifikante Pfade zu löschen und nur Pfade zu zeigen, die mindestens 1 Stern besitzen, also mit mindestens 95 % signifikant sind. Oder es können die Pfade aus dem Diagramm gelöscht werden, die einen bestimmten Wert des Regressionskoeffizienten - z.B. 0.10 - unterschreiten. Nach Klick auf den Knopf OK wird die Änderung in der Grafik ausgeführt.

Werden Pfade gelöscht, die eine Signifikanz von unter 95% besitzen, dann entsteht folgendes Pfaddiagramm.

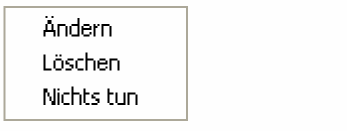


Für das Eliminieren von Pfaden scheint folgende Regel sinnvoll zu sein: Pfade, die nicht mit ca. 95 % signifikant sind, werden eliminiert.
Da bei großen Stichproben schon relativ kleine Pfadkoeffizienten und bei kleinen Stichproben umgekehrt erst relativ große Pfadkoeffizienten signifikant sind, ist es besser folgendes Entscheidungskriterium zu verwenden: Man sollte ß-Koeffizienten erst ab einer bestimmten Größe akzeptieren.

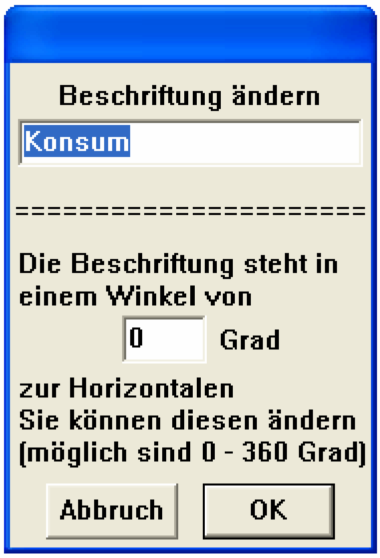
Zu viele Variable, lange Variablennamen
Pfaddiagramme mit 10 und mehr Variablen lassen sich im Prinzip auch noch darstellen. Der Platzbedarf für so große Diagramm überschreitet jedoch den Bildbereich im Grafik-Editor. Wenn man die Schieber "Größe" (der Grafik), "Höhe" und "Breite" in der linken Werkzeug-Leiste des Grafik-Editors entsprechend verschiebt, dann kann man das Diagramm eventuell noch in den angebotenen Bereich einpassen.

Wenn Variablennamen lang sind, etwa mehr als 12 Buchstaben umfassen, dann kann es geschehen, dass im Pfaddiagramm zwei hintereinander stehende Variablennamen sich berühren oder sogar überdecken. Hier gibt es 2 Abhilfen. Man kann das Diagramm breiter machen, indem man im Grafik-Editor in der linken Leiste den Schieber "Breite" verschiebt. Oder man macht die Variablennamen kürzer, indem man den Schieber "Var.namenbreite" nach rechts bewegt. Natürlich kann man beide Aktionen kombinieren.

Pfeile, Namen und Regressionskoeffizienten löschen oder verändern
Klicken Sie auf den Pfeil, Namen oder Regressionskoeffizienten, den Sie löschen oder verändern wollen. Dann Klick mit der rechten Maustaste. Es erscheint ein kleines Fenster.



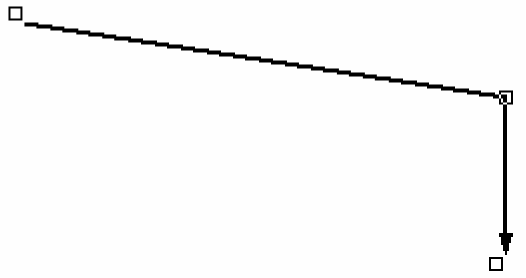
Klicken Sie nun auf "Löschen" (sofern Sie das betreffende Objekt löschen wollen). Wenn Sie den Namen ändern wollen, dann klicken Sie auf "Ändern". Es erscheint dann folgende Dialogbox



Sie können nun (in diesem Beispiel) den Namen "Konsum" in z.B. "Verbrauch" ändern.

Pfeile, Namen und Regressionskoeffizienten verschieben
Klicken Sie mit der linken Maustaste auf den Pfeil, Namen oder Regressionskoeffizienten. Mit gedrückter rechter Maustaste können Sie dann das betreffende Objekt verschieben. Wenn Sie auf einen Pfeil geklickt haben, dann sehen Sie folgendes

An den beiden Endpunken und am Eckpunkt des geknickten Pfeils befinden sich kleine Quadrate. Wenn Sie z.B. den Mauspfeil im Quadrat links plazieren, dann können Sie mit gedrückter linker Maustaste den Pfeilarm verdrehene, z.B. so



P25.1.10 Das Problem der Variablen-Reihung

Die Reihenfolge, in der die Variablen hintereinander gestellt werden, hängt von der Theorie ab, die man über den zu untersuchenden Gegenstand besitzt. Pfadanalyse ist "Regression plus Theorie". Ist eine Theorie nur rudimentär vorhanden, wie oft in den Sozialwissenschaftem, dann wird man mehrere konkurrierende Reihungen probieren und sich schließlich für das Pfaddiagramm entscheiden, das einem am plausibelsten erscheint.

Literatur

Bortz, J./**Schuster**, C.: Statistik für Human- und Sozialwissenschaftler, 2010, 7. Auflage, Kap. 24 Pfadanalyse, Springer Verlag

Holm, Kurt**:** Multiple lineare Regression und Pfadanalyse. in Holm: Die Befragung 5, Francke-Verlag, UTB 435

Reinecke, Jost: Strukturgleichungsmodelle in den Sozialwissenschaften, 2014, 2. Auflage, De Gruyter Oldenbourg