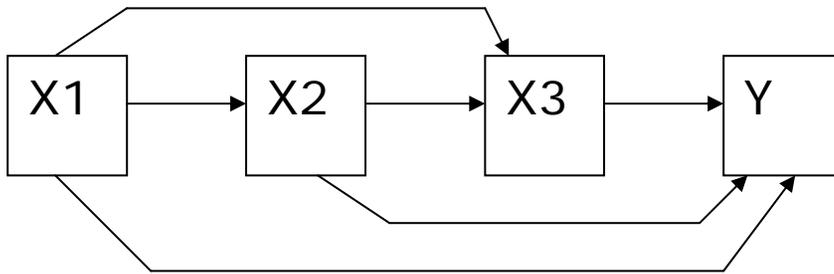


Pfadanalyse

Bacher, SoSe2007

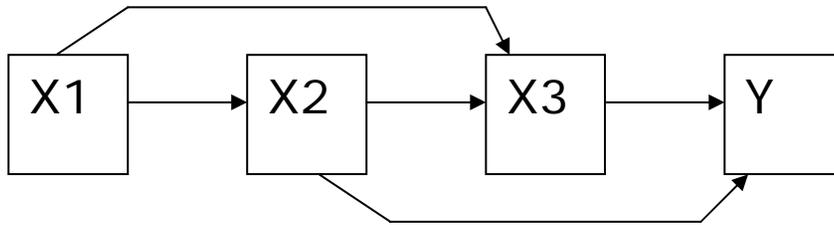
1. Grundlegende Verfahren

- **Explorative Pfadanalyse:** Kausale Beziehungen zwischen Variablen werden aufgedeckt, erforderlich ist eine kausale Anordnung der Variablen.
- **Konfirmatorische Pfadanalyse:** Kausalmodell *ist bekannt*, es wird geprüft, wie gut das theoretisch angenommene Modell den Daten angepasst ist.



explorativ:

Y hängt von X1, X2 und X3 ab, X3 hängt von X2 und X1 ab, X2 hängt von X1 ab



konfirmatorisch:

Y hängt von X2 und X3 ab, X1 hat keinen direkten Effekt auf Y, X3 hängt von X2 und X1 ab, X2 und X1 sollen gleich stark wirken, X2 hängt von X1 ab

- explorative Pfadanalyse kann mit Standardstatistikprogrammen gerechnet werden, für konfirmatorische Pfadanalyse ist eine Spezialsoftware erforderlich, z.B. LISREL¹, AMOS², MPLUS³

2. Aufgabe der explorativen Pfadanalyse

zentrale Aufgabe

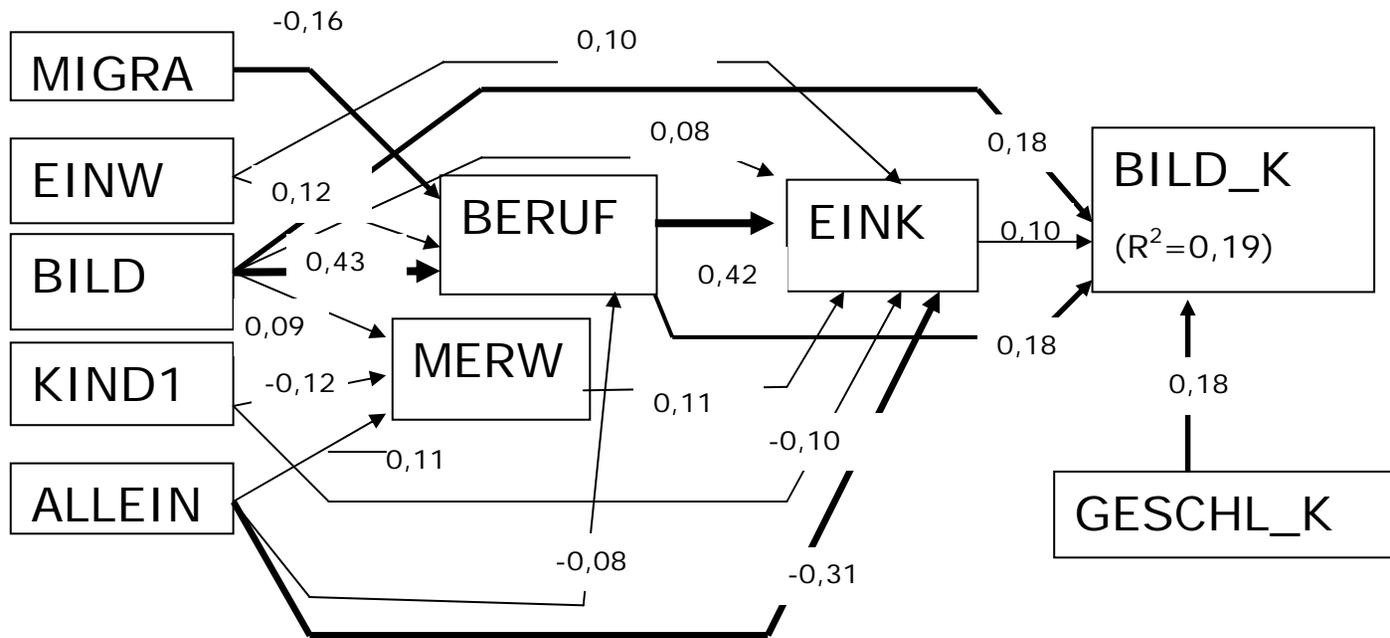
- Ermittlung der direkten (und indirekten) Wirkungen einer Variablen auf andere Variablen

Beispiel: siehe nächste Seite

¹ Informationen unter <http://www.ssicentral.com/lisrel/index.html>. Eine studentische Version ist frei erhältlich.

² Zusatzprodukt zu SPSS. http://www.spss.com/amos/loyalty_model_application.htm

³ Informationen unter <http://www.statmodel.com/>



MIGRA = Migrationshintergrund

EINW = Einwohnerzahl

BILD = höchste Bildung der Eltern

KIND19 =Zahl der Kinder bis 19 Jahre

ALLEIN = Alleinerzieherhaushalt

aus: Bacher (2003)

STATUS = berufliche Position der Eltern

EINK =Erwerbseinkommen der Eltern

MERW = mütterliche Erwerbstätigkeit

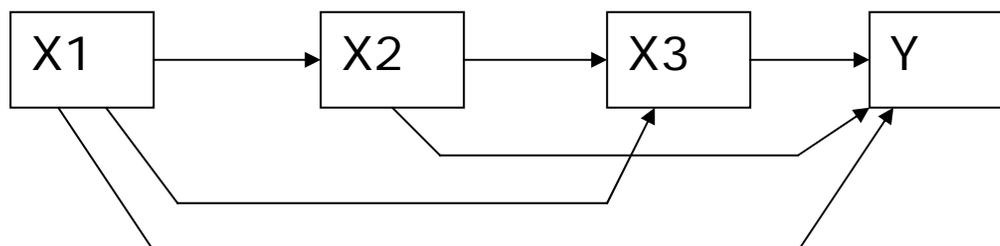
BILD_K = Bildungspartizipation des Kindes

GESCHL_K = Geschlecht des Kindes

3. Vorgehensweise und Umsetzung in ALMO⁴

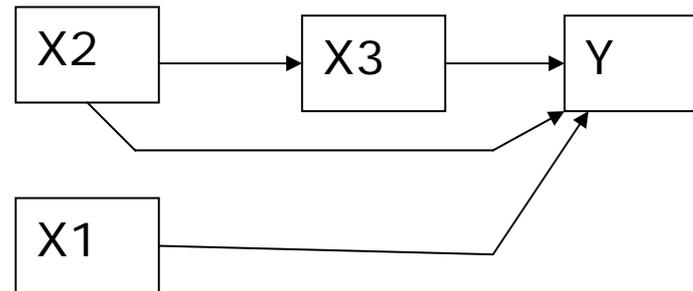
3.1. Allgemeine Vorgehensweise

(1.) Spezifikation eines (vollrekursiven) Variablen- bzw. Pfadmodells



Anmerkung: Das Modell muss nicht ganz abgearbeitet werden.

⁴ Umsetzung analog in anderen Statistikprogrammen, benötigt wird ein Programm zur multiplen Regression bzw. zum allgemeinen linearen Modell



(2.) Auswahl und Beschreibung der Variablen, falls erforderlich Bildung neuer Variablen, z.B. durch Faktorenanalyse

(3.) Schätzung der direkten (und indirekten) Effekte

(4.) Graphische Darstellung und Interpretation. Interpretation abhängig von den untersuchten Fragen, auch nicht-signifikante Effekte können von Bedeutung sein

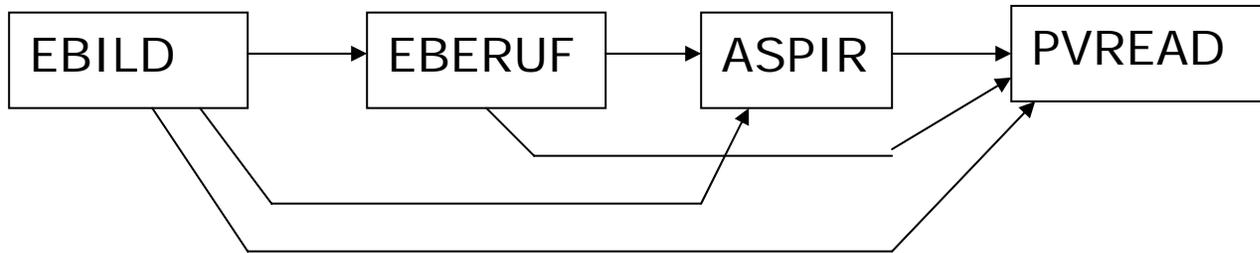
Schätzung der Effekte:

Abhängig vom Modell und dem Messniveau der Variablen:

- Programm P25 (nur quant. Variablen, nur eine Variable pro Stufe)
- Wiederholte Anwendung von Programm P20. Diese Technik kann auch in SPSS und anderen Statistikprogrammen angewendet werden.

3.2. Beispiel

(1.) Modellspezifikation



(2.) Auswahl, Beschreibung und gegebenenfalls Bildung neuer Variablen

Auswahl und Auszählung der relevanten Variablen

Q23 Which of the following do you expect to complete?

(Please <tick> as many as apply.)

- a) <ISCED level 2>
- b) <ISCED level 3B or C>
- c) <ISCED level 3A>
- d) <ISCED level 4>
- e) <ISCED level 5B>
- f) <ISCED level 5A or 6>

a) → V54, b) → V55 usw.

bei Datei-Problemen

Format der Daten

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

Wenn Dateiformat FIX oder Nicht-Standard-FREI

Option: Gruppierungsvariable

Option: Ein- und Ausschliessen von Untersuchungseinheiten

Option: Umkodierungen und Kein-Wert-Angaben

Loesche wieder diese Box

erzeuge zusätzliche Felder für Gewichtungs-Angaben

Variable 59 ST23Q06.Expect.I

Wert	Faelle	%	% ohne KW	% kumuliert
Kein Wert	11	1.10	-	-
1 Tick	239	23.90	24.17	24.17
2 No Tick	750	75.00	75.83	100.00
Summe	1000	100%		
Summe ohne KW	989		100%	

Variable 238 HISCED.Highest.e

Wert	Faelle	%	% ohne KW	% kumuliert
Kein Wert	21	2.10	-	-
0 None	2	0.20	0.20	0.20
1 ISCED 1	10	1.00	1.02	1.22
2 ISCED 2	54	5.39	5.51	6.73
3 ISCED 3B C	344	34.37	35.10	41.84
4 ISCED 3A ISCED 4	147	14.69	15.00	56.84
5 ISCED 5B	289	28.87	29.49	86.33
6 ISCED 5A 6	134	13.39	13.67	100.00
Summe	1001	100%		
Summe ohne KW	980		100%	

Bildung neuer Variablen

Berechnung einer neuen Variablen Bildungsaspiration (ASPIR): Dazu bestehen mehrere Möglichkeiten

ASPIR = V59(1=1;2=0); # V1003 #

oder ASPIR =

- 1 <ISCED level 2>
- 2 <ISCED level 3B or C>
- 3 <ISCED level 3A>
- 3 <ISCED level 4> (z.B. HTL).
- 4 <ISCED level 5B>
- 5 <ISCED level 5A or 6>

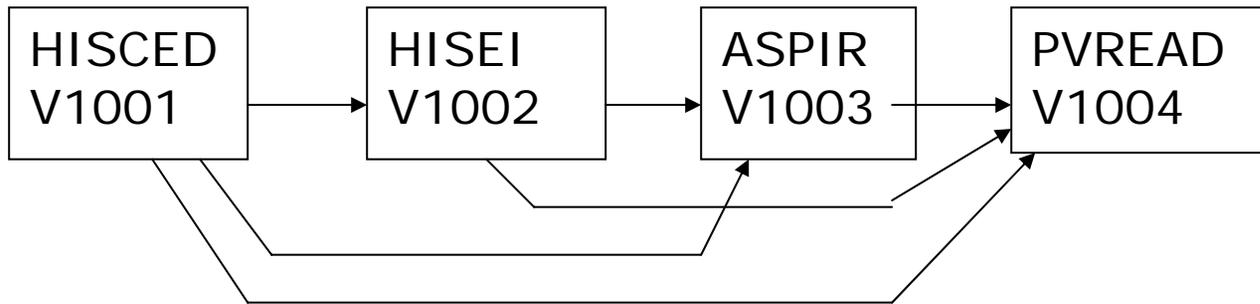
Weitere Variablen:

PVREAD = V302; # V1004 #

EBILD= V238; # V1001 #

EBERUF = V231; # V1002 #

(3.) Wiederholte Regressionsanalysen



Drei Analysen:

V1004 = abh. Variable, alle Variablen links unabh. Variablen

V1003 = abh. Variable, alle Variablen links unabh. Variablen

V1002 = abh. Variable, alle Variablen links unabh. Variablen

Datei der Variablennamen

Hilfe

"D:\pisa2003\PisaAustria\Variablennamen.nam"

zeige = Namensdatei in Output zeigen
leer = nicht zeigen

Freie Namensfelder

Hilfe

Leere alle Eingabefelder dieser Sub-Box

n1001=ebild;
 n1002=eberuf;
 n1003=aspir; ;
 n1004=pvread;

[...] erzeuge zusätzliche Namensfelder

quantitative abhängige Variable

↔ punread

ordinale abhängige Zielvariable

Hilfe

↔

nominale abhängige Zielvariable

Hilfe

↔

Analyse-Variable: Unabhängige Variable

Hilfe

nominale unabhängige Variable

Hilfe

↔

↑ ↓

Interaktionen x. Ordnung zwischen den
nominalen unabhängigen Variablen bilden
oder einige ausgewählte Interaktionen bilden
∅ =keine Interaktionen bilden

Hilfe

↔

paarweise Vergleiche (Kontraste) für die
nominalen unabhängige Variablen rechnen

quantitative unabhängige Variable

Hilfe

↔ ebild, eberuf, aspir

ebild=v238;
eberuf=U231;
aspir=U59(1=1;2=0;sonst=kw);
puread=U302;

erzeuge zusätzliche Felder für Umkodierungen / Kein_Wert-Angaben

obige Umkodierungen und Kein-Wert-Angaben in eine Datei speichern

0 = Datei mit diesem Namen **neu anlegen**
obige Umkodierungen und Kein-Wert-Angaben darin speichern
1 = Datei mit diesem Namen **besteht schon**
Umkodierungen und Kein-Wert-Angaben hinten anfügen

Kontrollieren, ob Umkodierung so erfolgt wie gewünscht

Option: Spezielle Kein-Wert-Behandlung

Loesche wieder diese Box

Untersuchungseinheiten gewichten
(nicht möglich wenn ordinale Variable in Analyse)

Gewicht1=v317*U403;

erzeuge zusätzliche Felder für Gewichtungs-Angaben

(4.) Interpretation der Ergebnisse

analog zur multiplen Regression, verwendet werden:

- standardisierte Regressionskoeffizienten (wenn nur quant. Var.)

oder

- partielle Korrelationskoeffizienten (bei gemischten Var., aber auch bei nur quant. Var.)

und

- erklärte Varianz R^2

1. Analyse

Koeffizienten fuer quantitative/ordinale Variable aus univariater Analyse

hinsichtlich der abhaeng. Var. V1004 pvread

Variable	standard.			Signifikanz	
	Regr. koeff.	Regr. koeff.	part. Korrel.	p	(1-p)100
V1001 ebild	0.0411	3.4251	0.0411	0.2011	79.89
V1002 eberuf	0.2607	1.6658	0.2518	0.0000	100.00
V1003 apsir	0.2891	69.3882	0.2910	0.0000	100.00

Zusammenfassung

hinsichtlich der abhaengigen Variablen V1004 pvread

Streuungsquelle	Streuung	Korrel Koeff.	F-Wert	df	Signifikanz p	(1-p)100
Gesamtstreuung	10234816.1341					
Fehlerstreuung	8018128.2612			965		
alle unabh. Var. zusammen	2216687.8729	0.4654	88.9278	3	0.0000	99.9995
V1001 ebild	13589.3307	0.0411	1.6355	1	0.2011	79.8882
V1002 eberuf	542593.2707	0.2518	65.3023	1	0.0000	99.9995
V1003 apsir	741909.0602	0.2910	89.2904	1	0.0000	99.9995

$$R^2 = .4654^2 = 0,217$$

2. Analyse

hinsichtlich der abhaeng. Var. V1003 apsir

Variable	standard.			Signifikanz	
	Regr. koeff.	Regr. koeff.	part. Korrel.	p	(1-p)100
V1001 ebild	0.2083	0.0722	0.1981	0.0000	100.00
V1002 eberuf	0.2222	0.0059	0.2107	0.0000	100.00

Zusammenfassung
hinsichtlich der abhaengigen Variablen V1003 apsir

Streuungsquelle	Streuung	Korrel Koeff.	F-Wert	df	Signifikanz	
					p	(1-p)100
Gesamtstreuung	176.1555					
Fehlerstreuung	152.8198			958		
alle unabh. Var. zusammen	23.3357	0.3640	73.1438	2	0.0000	99.9995
V1001 ebild	6.2408	0.1981	39.1224	1	0.0000	99.9995
V1002 eberuf	7.1011	0.2107	44.5157	1	0.0000	99.9995

3. Analyse

hinsichtlich der abhaeng. Var. V1002 eberuf

Variable	standard.		part. Korrel.	Signifikanz	
	Regr. koeff.	Regr. koeff.		p	(1-p)100
V1001 ebild	0.4286	5.5840	0.4286	0.0000	100.00

Zusammenfassung

hinsichtlich der abhaengigen Variablen V1002 eberuf

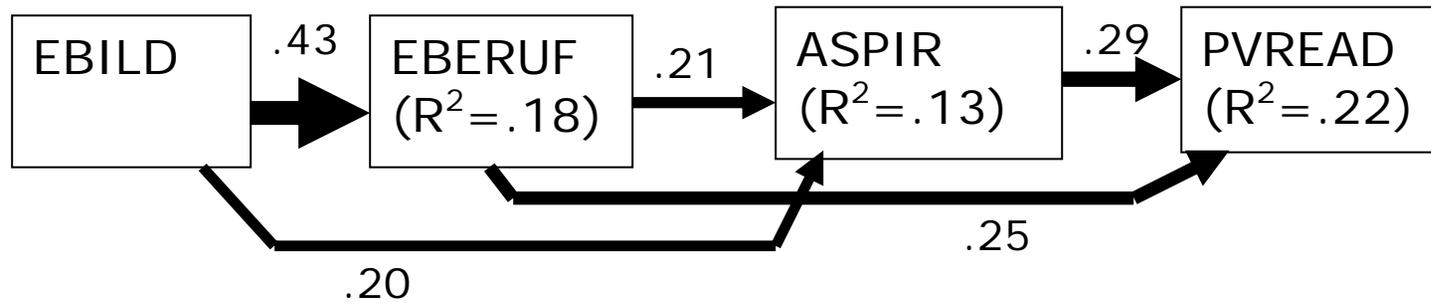
Streuungsquelle	Streuung	Korrel Koeff.	F-Wert	df	Signifikanz	
					p	(1-p)100
Gesamtstreuung	245256.6635					
Fehlerstreuung	200199.3239			946		
alle unabh. Var. zusammen	45057.3396	0.4286	212.9090	1	0.0000	99.9995
V1001 ebild	45057.3396	0.4286	212.9090	1	0.0000	99.9995

Graphische Darstellung:

Basis: signifikante Wirkung und häufig zusätzlich bestimmte Anforderungen bzgl. Mindestgröße des Absolutbetrages des Koeffizienten, z.B. nur Koeffizienten mit einem Absolutbetrag größer 0,10.

Falls Signifikanzprüfung nicht zulässig (z.B. abhängige dichotome Variable, komplexes Stichprobendesign), dann kann nur ein Schwellenwert verwendet werden.

Abbildung 1: Ergebnisse der explorativen Pfadanalyse (a)



(a) eingetragen wurden nur Pfadkoeffizienten mit einem Absolutbetrag größer 0,10. Als Pfadkoeffizienten wurden partielle Korrelationskoeffizienten verwendet

Quelle: PISA2003, eigene Berechnungen, gewichtete Daten, n=948 – 969

4. Übungsaufgaben

Erweitern Sie das Modell um folgende Variable: Geschlecht des Kindes, Migrationshintergrund, beengte Wohnverhältnisse und kulturelles
Zeichnen sie das erweiterte Pfaddiagramm. Schätzen Sie das Modell und interpretieren Sie das Ergebnis.

Literatur:

- Bacher, J., 2003: Soziale Ungleichheit und Bildungspartizipation im weiterführenden Schulsystem Österreichs. Österreichische Zeitschrift für Soziologie, 28. Jg., Nr. 3, 3-33.
- Holm, K., 1977: Lineare multiple Regression und Pfadanalyse. In: Holm, K. (Hg.): Die Befragung 5. München, S. 7-102
- Holm, K., 1979: Das allgemeine lineare Modell. In: Holm, K. (Hg.): Die Befragung 6. München, S. 11-213