

Institut für Soziologie
Benjamin Gedon

Methoden 2

Kausalanalyse





1. Kausalanalyse
2. Anwendungsbeispiel
3. Wiederholung
4. Übungsaufgabe



Kausalität:

- Identifizieren von Ursache-Wirkungs-Beziehungen
 - Ursache geht Wirkung zeitlich voraus
 - Es besteht ein innerer Zusammenhang zwischen Ursache und Wirkung
- Im Experiment wird hierzu die Einflussvariable vom Forscher beeinflusst (kontrolliert) und der Effekt auf die abhängige Variable gemessen.
- Auf Grund der Komplexität unserer Fragestellungen und aus ethischen Gründen ist dies in der Sozialwissenschaft selten möglich.
- Alternative: Theoretische Modellierung und darauf aufbauend multivariate Analyse



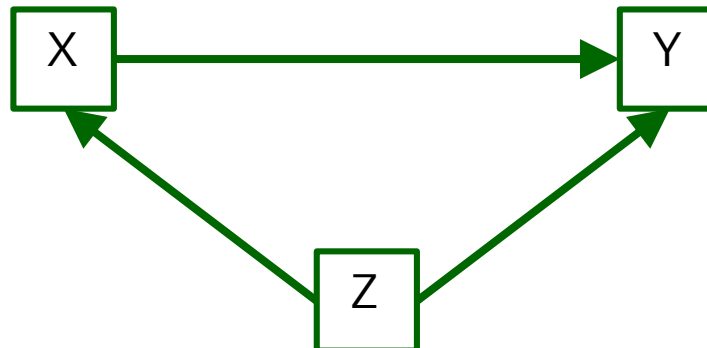
Theoretische Modellierung muss folgende Fragen beantworten (1):

- Warum soll X auf Y wirken (und nicht umgekehrt)?



Warum beeinflusst die Abiturnote die Bachelornote?

- Welche Variablen (Z) beeinflussen sowohl X als auch Y? (Konfundierung)

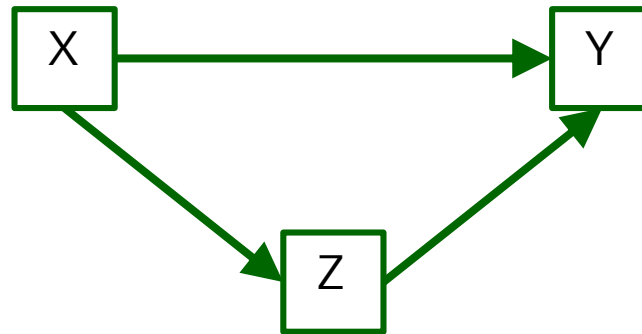


Was beeinflusst Abiturnote und Bachelornote? Evtl. die Bildung der Eltern.



Theoretische Modellierung muss folgende Fragen beantworten (2):

- Welche Variablen beeinflussen Y werden aber selbst von X beeinflusst?
(Intervention)



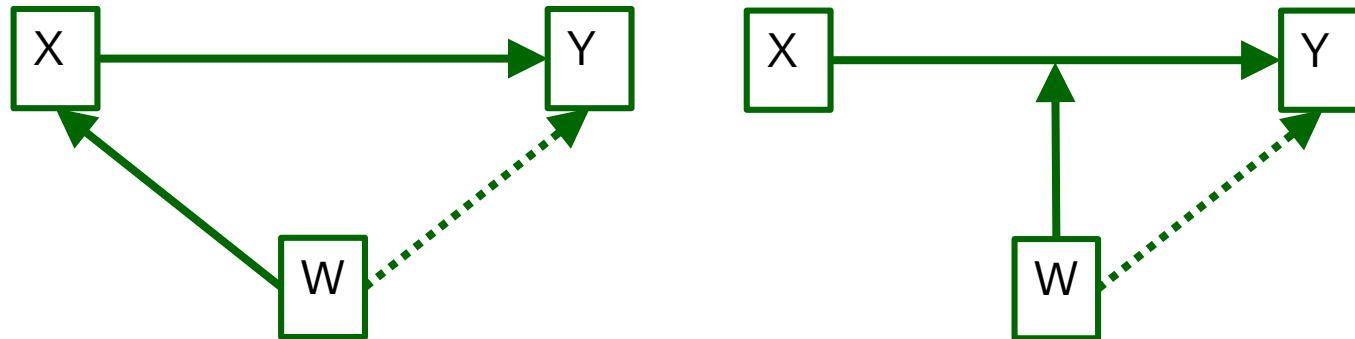
Das Vorwissen wird von der Abiturnote beeinflusst, hat aber gleichzeitig Einfluss auf die Bachelornote.



Theoretische Modellierung muss folgende Fragen beantworten (3):

- Ist der Zusammenhang zwischen X und Y für bestimmte Ausprägungen einer Variable (W) anders als für andere Ausprägungen? (Interaktion)

Hierfür gibt es unterschiedliche Notationen:

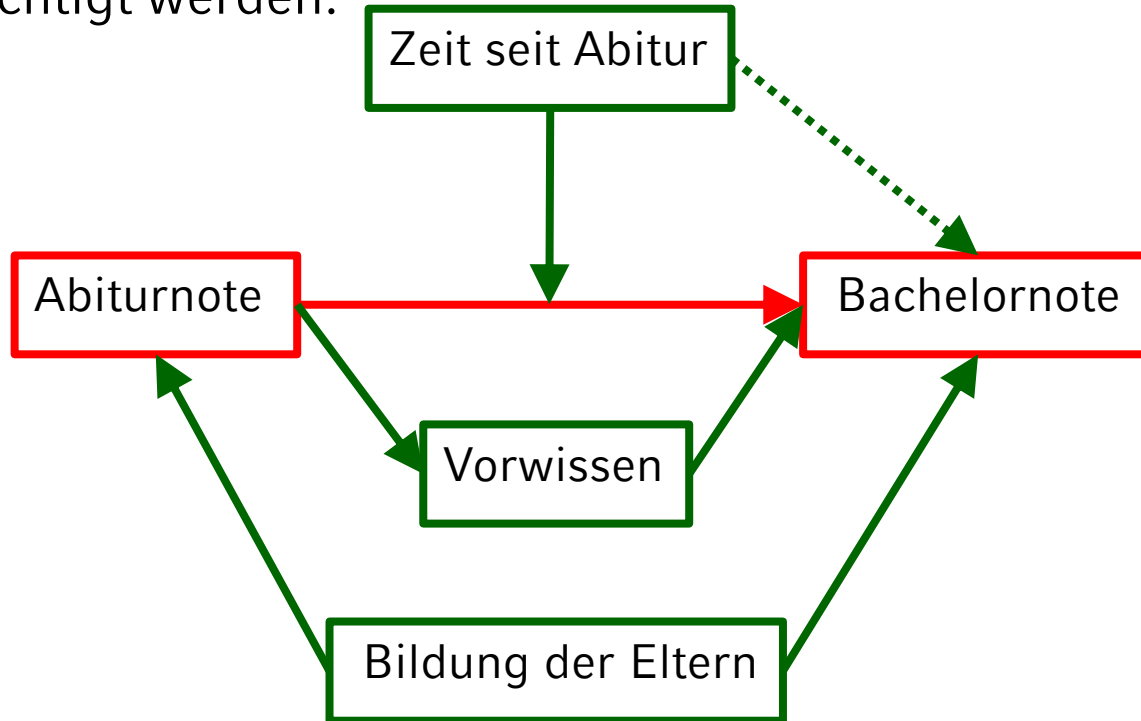


Die Zeit die seit dem Abitur vergangen ist könnte den Einfluss der Abiturnote auf die Bachelornote abschwächen.



Wir interessieren uns für den **Zusammenhang** zwischen der Abitur- und der Bachelornote.

Hierfür müssen bei der Analyse die **grünen** Beziehungen berücksichtigt werden.



Hinweis: Echte Kausalschlüsse sind mit Querschnittsdaten – also auch mit den Absolventendaten – nicht möglich.



Zunächst: Überprüfen, ob überhaupt ein Zusammenhang zwischen AV und UV in den Daten zu finden ist.

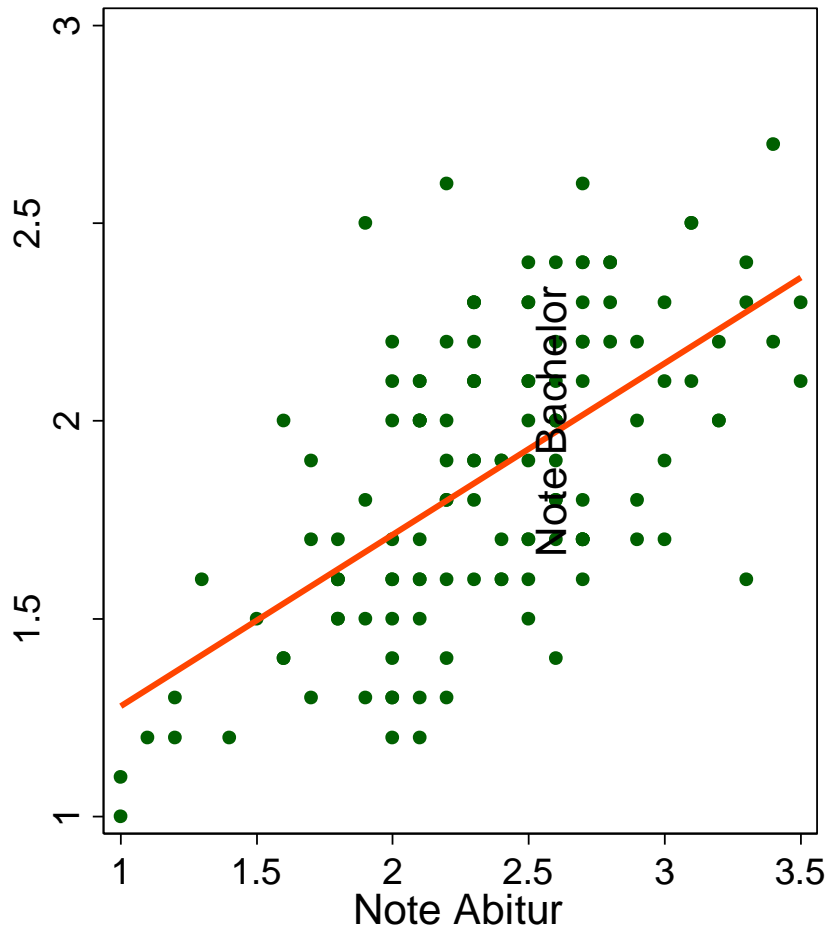
```
. pwcorr banote abinote, sig obs
```

	banote	abinote
banote	1.0000	
	132	
abinote	0.6100	1.0000
	0.0000	
	132	132

Starker Zusammenhang zwischen Abitur- und Bachelornote.
Dieser ist höchst signifikant.

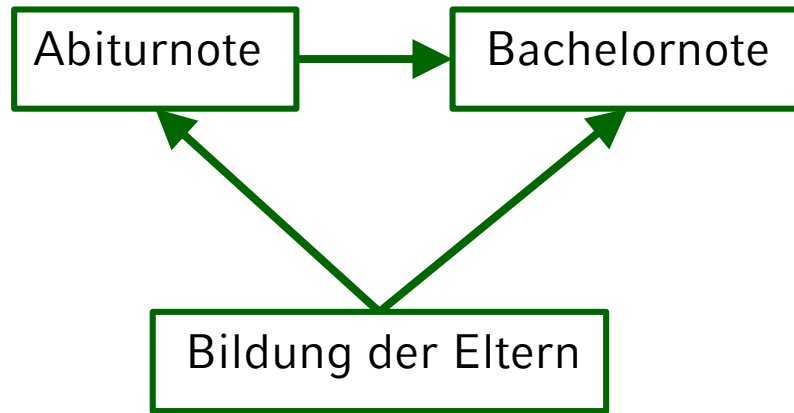


Grafische Deskription des Zusammenhangs:





Einfachregression zwischen Abitur- und Bachelornote.



	AV: Gesamtnote BA-Studium
	Modell 1
Abiturnote zentriert	0,433 (0,000)
Konstante	1,876 (0,000)
Fallzahl	132
R ²	0,372

p-Werte in Klammern

Dieser Zusammenhang wird aber – laut obigem Modell – durch die konfundierende Variable „Bildung der Eltern“ beeinflusst (verzerrt).

Um den Effekt der Abitur- auf die Bachelornote zu schätzen muss also die Bildung der Eltern kontrolliert werden.



Bildung der Eltern:

- Im Folgenden operationalisiert über Abitur der Mutter.
 - Andere Messungen:
 - Abitur Vater
 - Hochschulabschluss Vater oder Mutter
- Diese haben – auch in Kombination – keine anderen Effekte als die Bildung der Mutter.

Hinweis: Auf der vorangegangenen Folie wurde (und auf folgenden Folien wird) die zentrierte Abiturnote verwendet. Hierfür wird die durchschnittliche Abiturnote von der individuellen Abiturnote abgezogen. Im vorliegenden Fall konnten so Multikollinearitätsprobleme ausgeräumt werden. Dies ist aber Stoff von fortgeschritteneren Veranstaltungen. Auf die Interpretation der Koeffizienten hat die Zentrierung keine Auswirkung; lediglich die Konstante muss anders interpretiert werden.



Bildung der Eltern:

Zunächst Korrelation zwischen Abi- und BA-Note für Personen mit bzw. ohne Mutter mit Abitur.

```
. bysort abimut: pwcorr banote abinote, sig obs
```

```
-> abimut = nein
```

	banote	abinote
banote	1.0000	
	63	
abinote	0.6092	1.0000
	0.0000	
	63	63

In beiden Gruppen gleiche Zusammenhangsstärke. Der postulierte Effekt liegt folglich nicht vor.



Regression
zwischen Abitur-
und Bachelornote
unter Kontrolle der
Bildung der Eltern.

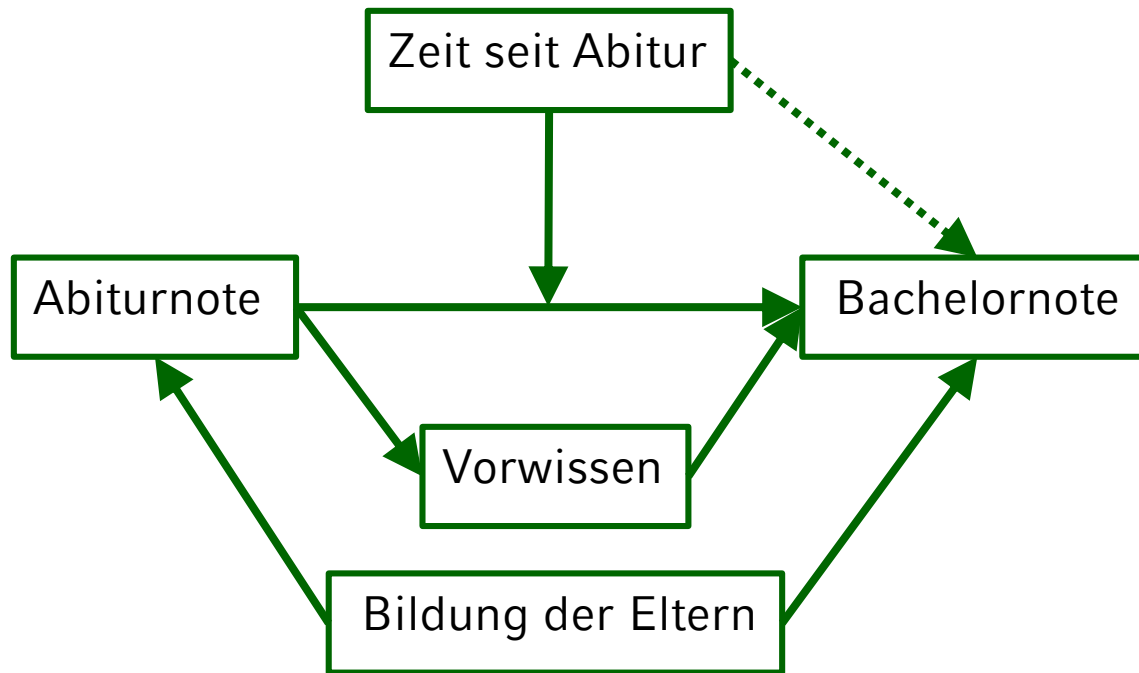
```
outreg2 [modell1  
modell2] using tab2,  
replace label word  
pvalue noaster
```

	AV: Gesamtnote BA-Studium	
	Modell 1	Modell 2
Abiturnote zentriert	0,433 (0,000)	0,442 (0,000)
Mutter mind. Abitur?		0,0443 (0,403)
Konstante	1,876 (0,000)	1,854 (0,000)
Fallzahl	132	132
R ²	0,372	0,376

p-Werte in Klammern

Der Effekt der Abiturnote auf die Bachelornote bleibt konstant und höchstsignifikant. Die Bildung der Eltern hat keinen Einfluss auf den Zusammenhang zwischen Abitur- und Bachelornote.

In diesem Modell gibt es keinen Zusammenhang zwischen elterlicher Bildung und Bachelornote.



Das Vorwissen wurde im Fragebogen leider nicht abgefragt. Daher können wir in den Daten nicht differenzieren welcher Teil des Effekts direkt auf die Bachelornote wirkt und welcher Teil durch das größere Vorwissen vermittelt wird.



Zeit zwischen Abiturnote und Bachelorabschluss – Ungenaue Messung

- bekannte Werte:
 - Dauer Erwerbstätigkeit nach Abitur aber vor Studienbeginn
 - Studienbeginn
 - Studienende

- unbekannte Werte:
 - Arbeitslosigkeit
 - FSJ u.ä.
 - Weltreise, Work + Travel



	AV: Gesamtnote BA-Studium		
	Modell 1	Modell 2	Modell 3
Abiturnote zentriert	0,433 (0,000)	0,442 (0,000)	0,436 (0,00193)
Mutter mind. Abitur?		0,0443 (0,403)	0,0448 (0,403)
in Jahren: BA-Studium + Erwerbstätigkeit nach Abitur			-0,0105 (0,710)
Abinote * BA-Studium + Erwerbstätigkeit nach Abitur			0,00115 (0,974)
Konstante	1,876 (0,000)	1,854 (0,000)	1,889 (0,000)
Fallzahl	132	132	132
R ²	0,372	0,376	0,377

p-Werte in Klammern



Interpretation:

- Effekt der Abiturnote bleibt über alle Modelle hinweg nahezu konstant.
- Im Modell mit Interaktionsterm ist der Abitureffekt immer noch hoch in den anderen Modellen ist er höchst signifikant.
- Die elterliche Bildung hat keinen Einfluss auf den Zusammenhang zwischen Abinote und BA-Note. Vermutung: Kinder von Nichtakademikern haben aber schlechtere Chancen einen Uniabschluss zu erreichen.

	AV: Gesamtnote BA-Studium		
	Modell 1	Modell 2	Modell 3
Abiturnote zentriert	0,433 (0,000)	0,442 (0,000)	0,436 (0,00193)
Mutter mind. Abitur?		0,0443 (0,403)	0,0448 (0,403)



Interpretation:

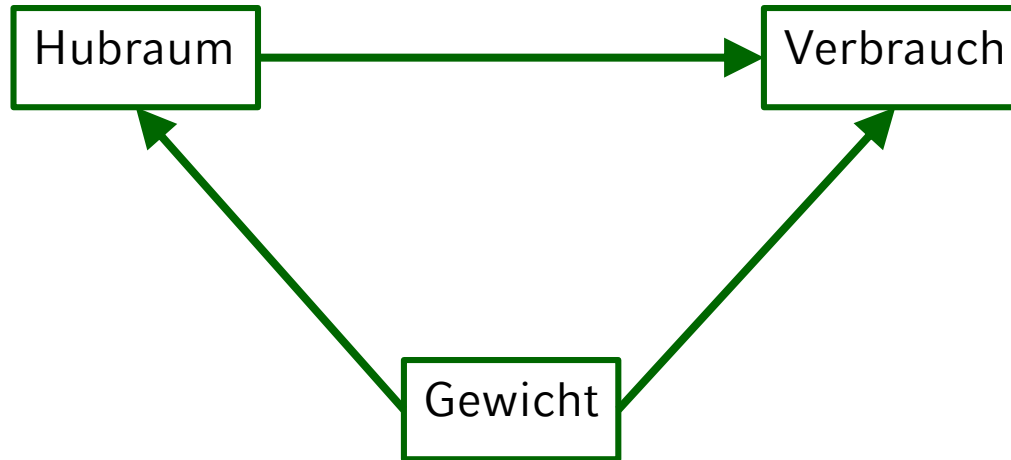
- Die Zeit die zwischen Abitur und Bachelorabschluss verstreicht hat keinen Einfluss auf die Richtung oder Stärke des Effekts der Abiturnote auf die Bachelornote.
- Alle Modelle erklären durch die aufgenommenen Variablen zwischen 37 und 38 % der Varianz der Bachelornote. Die Konstanz dieses Wertes ist ein weiterer Hinweis für die Abwesenheit der theoretisch postulierten Drittvariablen-Effekte.

Abinote * BA-Studium + Erwerbstätigkeit nach Abitur			0,00115 (0,974)
Konstante	1,876 (0,000)	1,854 (0,000)	1,889 (0,000)
Fallzahl	132	132	132
R ²	0,372	0,376	0,377

p-Werte in Klammern

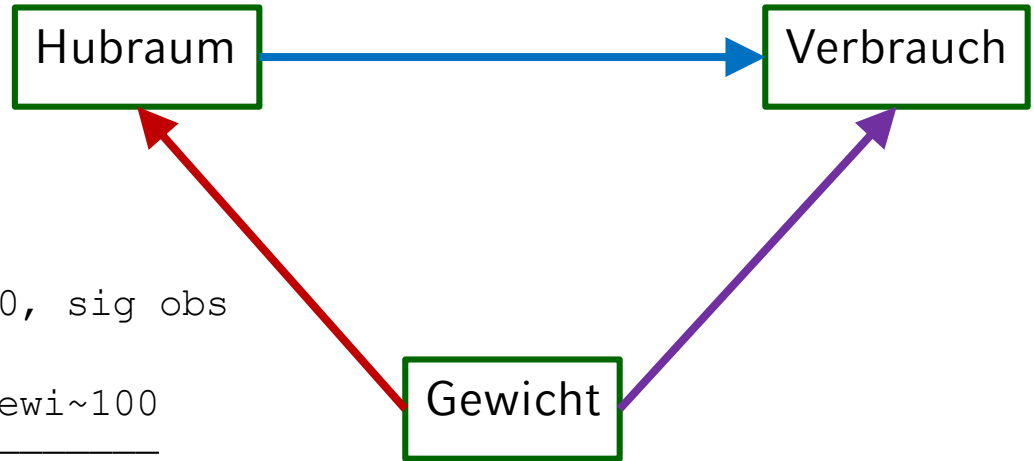


Ein weiteres Beispiel aus dem Autodatensatz:



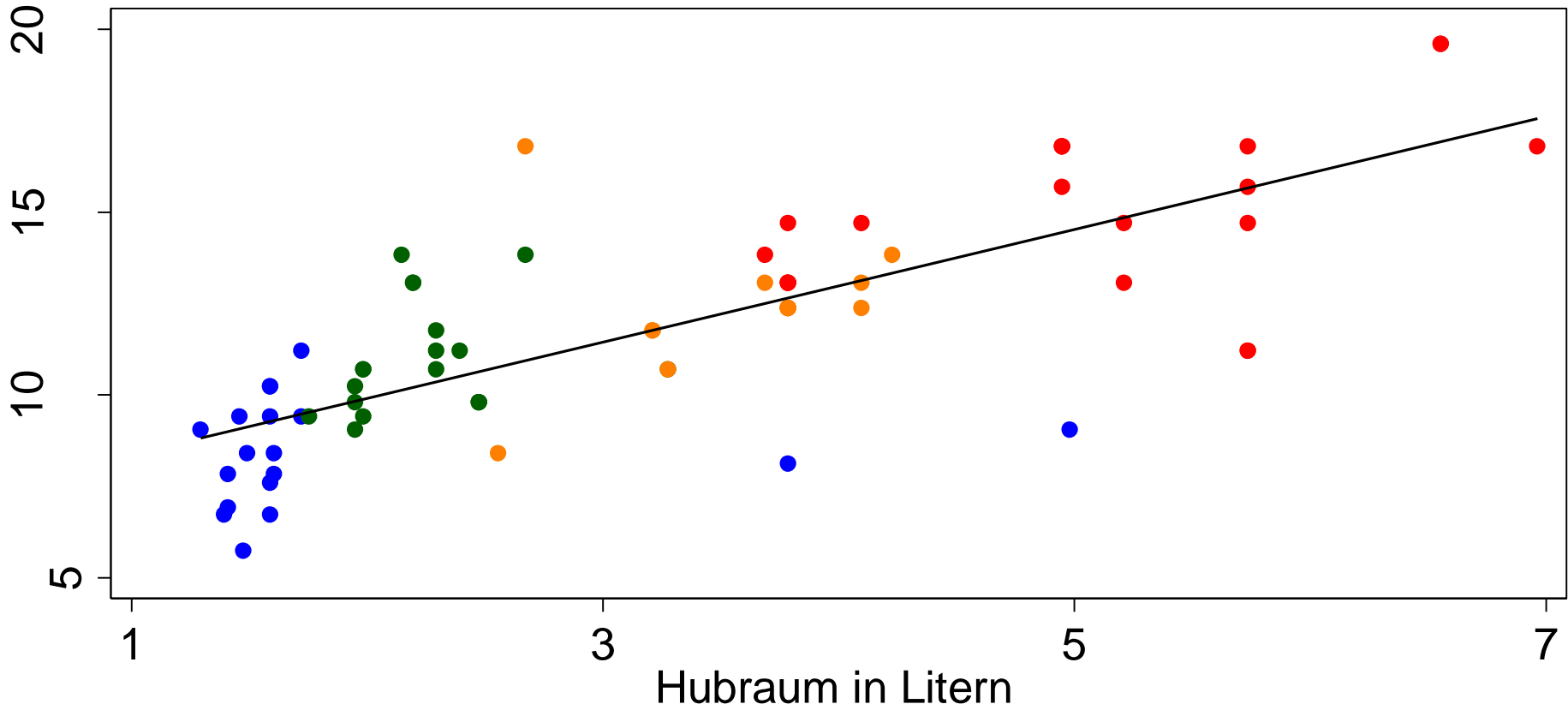
Wir interessieren uns dafür, warum Autos mit größerem Hubraum mehr Benzin verbrauchen.

Wir vermuten, dass Autos mit größerem Gewicht zum einen einen größeren Hubraum, zum anderen aber auch einen größeren Verbrauch haben. Wird dadurch aber der ganze Zusammenhang zwischen Hubraum und Verbrauch erklärt?



```
. pwcorr lp100km hubraum gewicht100, sig obs
```

	lp100km	hubraum	gewi~100
lp100km	1.0000		
	74		
hubraum	0.7712	1.0000	
	0.0000		
	74	74	
gewicht100	0.8544	0.8949	1.0000
	0.0000	0.0000	
	74	74	74



Quelle: STATA-Autodatensatz



Interpretation Modell 1:

- Pro Liter Hubraum nimmt der Verbrauch pro 100 km um 1,5 Liter zu. Dieser Effekt ist höchst signifikant.
- Durch den Hubraum können 59,5 % der Varianz des Verbrauchs erklärt werden.

	Verbrauch in Litern pro 100 km	
	Automodell 1	Automodell 2
Hubraum in Litern	1,543 (0,000)	0,0663 (0,811)
Gewicht in 100 kg		0,704 (0,000)
Konstante	6,818 (0,000)	1,945 (0,0393)
Fallzahl	74	74
R ²	0,595	0,730

p-Werte in Klammern

Interpretation Modell 2:

- Der Effekt des Hubraums wird durch Kontrolle des Gewichts komplett aufgelöst. Es handelte sich also um eine Scheinkorrelation.
- Pro 100 kg nimmt der Verbrauch um 0,7 Liter höchst signifikant zu.
- Jetzt werden durch die beiden Variablen 73 % der Varianz des Verbrauchs erklärt.



Regressionsmodell mit metrischen und dichotomen Variablen

Source	SS	df	MS	Number of obs	=	91
Model	3952.30849	3	1317.43616	F(3, 87)	=	1.85
Residual	61844.1943	87	710.852808	Prob > F	=	0.1435
Total	65796.5027	90	731.072253	R-squared	=	0.0601
				Adj R-squared	=	0.0277
				Root MSE	=	26.662

ek_std	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mon_ausl	-.2400585	1.06057	-0.23	0.821	-2.348056	1.867939
frau	-9.079167	6.023129	-1.51	0.135	-21.05079	2.892453
abinote	-9.900503	5.206308	-1.90	0.061	-20.2486	.4475971
_cons	104.9309	13.60798	7.71	0.000	77.88356	131.9782



Interaktion Dummy * metrisch

Source	SS	df	MS	Number of obs	=	168
Model	8,01195168	3	2,67065056	F(3, 164)	=	29,26
Residual	14,9675126	164	,091265321	Prob > F	=	0,0000
Total	22,9794643	167	,137601583	R-squared	=	0,3487
				Adj R-squared	=	0,3367
				Root MSE	=	,3021

v29_1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Frau	-,3946252	,2238154	-1,76	0,080	-,8365563	,047306
NoteAbi	,2932869	,0765436	3,83	0,000	,1421489	,4444249
IntNoteFrau	,1660298	,0936539	1,77	0,078	-,0188932	,3509528
_cons	1,212252	,1848855	6,56	0,000	,8471895	1,577315



Dummyvariablen:

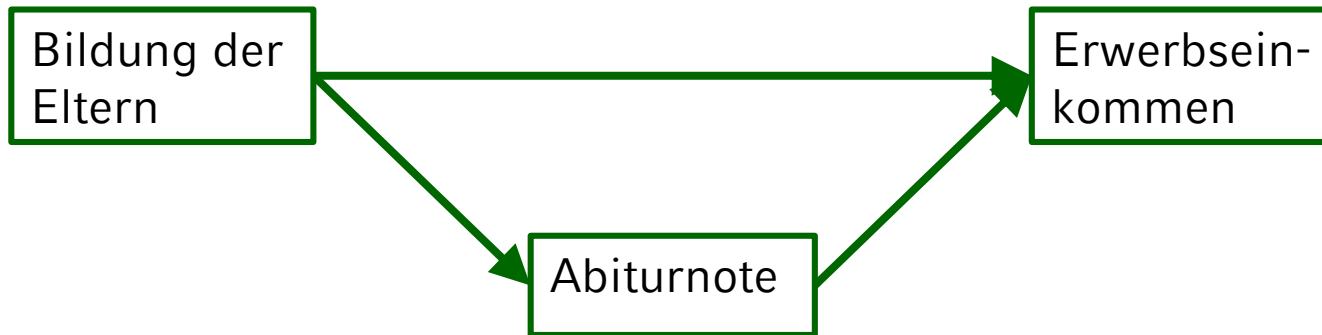
	Niedrig	Mittel	Hoch
Noch Schüler	.	.	.
Kein Abschluss	1	0	0
Hauptschule beendet ohne Abschluss	1	0	0
(Qualifizierender) Hauptschulabschluss	1	0	0
Mittlere Reife	0	1	0
Fachabitur	0	0	1
Abitur	0	0	1



1. Sie interessieren sich dafür, wie sich die Bildung der Eltern auf das Erwerbseinkommen auswirkt.
 - Bilden Sie das Erwerbseinkommen als Stundenlohn wie in der letzten Übungsaufgabe.
 - Erstellen Sie aus den Variablen v132_1 bis v132_14 insgesamt 4 Dummyvariablen, die den Wert 1 für folgende Ausprägung annehmen:
 - Vater Abitur
 - Vater Hochschulabschluss oder promoviert
 - Mutter Abitur
 - Mutter Hochschulabschluss oder promoviert
 - Schätzen Sie ein entsprechendes Regressionsmodell und interpretieren Sie die Koeffizienten, die Konstante und die Modellgütemaße.



2. Sie vermuten, dass sich die Bildung der Eltern auch auf die Abiturnote auswirkt. Die Abiturnote wiederum beeinflusst das Einkommen.



- Berücksichtigen Sie in obigem Regressionsmodell also zusätzlich die Abiturnote und vergleichen Sie die Modelle anschließend.
 - Welchen Einfluss hat die Abiturnote auf die Beziehung zwischen elterlicher Bildung und Einkommen?
 - Wie ändert sich die Erklärungskraft des Modells durch die Berücksichtigung der Abiturnote?