LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Querschnittsdatenanalyse:

Angewandte Regressionsanalyse mit STATA

Prof. Dr. Josef Brüderl
LMU München
Wintersemester 2018/19

Inhalt

Lernziele

- Kenntnis verschiedener Querschnitts-Regressionsmodelle
  - Keine Herleitungen, keine Berechnungen per Hand
  - Sondern ein anwendungsorientierter Überblick
- Darstellung und Interpretation der Regressionsergebnisse
  - Insbesondere die graphische Darstellung der Regressionsergebnisse wird betont („Das Zeitalter der Regressionstabelle ist vorbei“)
- Interpretation von Interaktionen
  - Selbst im linearen Modell wird hier viel falsch gemacht
  - Bei nicht-linearen Modellen ist es noch komplizierter
- Praktische Umsetzung der Regressionsmodelle mit STATA
  - Die grundlegenden STATA-Befehle sind in den Folien enthalten
  - Zusätzlich kann man anhand der begleitenden STATA Do-Files die Berechnungen nachvollziehen

ALLBUS 2002

- Bevölkerungsumfrage alle 2 Jahre seit 1980 (N ~ 3.000)
  - Von GESIS als Service für die Sozialforschung
  - Trenddaten
- ALLBUS 2002 (N=2.820)
  - Einwohnermelderegisterstichprobe
  - Ostdeutsche überrepräsentiert
  - GG: alle deutschsprachigen Personen über 18, wohnhaft in D in Privathaushalten
  - Ausschöpfung: 47%
  - Mündliches Interview (CAPI)
  - Infos: http://www.gesis.org/allbus

ALLBUS 2002: Datenaufbereitung

- Für den Kurs wurden einige Variablen aufbereitet
  - Abgespeichert im Datensatz: **`AllbReg.dta`**

Daten: ALLBUS 2002
Do-File: 00 Datenaufbereitung.do

- Abhängige Variablen
  - `eink`: monatliches Nettoeinkommen in Euro
  - `rechts`: Links-Rechts Selbsteinstufung (Skala 1-10)
  - `arblos`: Arbeitslosigkeit in den letzten 10 Jahren (0=nein, 1=ja)
  - `partei`: Wahlabsicht (CDU, SPD, FDP, Grüne, PDS)
  - `oecdeink`: Nettoäquivalenzeinkommen in Euro
- Unabhängige Variablen
  - `bild`: schulische und berufliche Bildung in Jahren
  - `alter`: Alter in Jahren
  - `exp`: Berufserfahrung (Berechnung: alter - bild - 6)
  - `prestv`: Berufsprestige des Vaters (Magnitude-Skala)
  - `frau`: Dummy für Frau
  - `ost`: Dummy für Ostdeutscher
  - `beruf`: berufliche Stellung (Arbeiter, Angest., Beamter, Selbst.)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

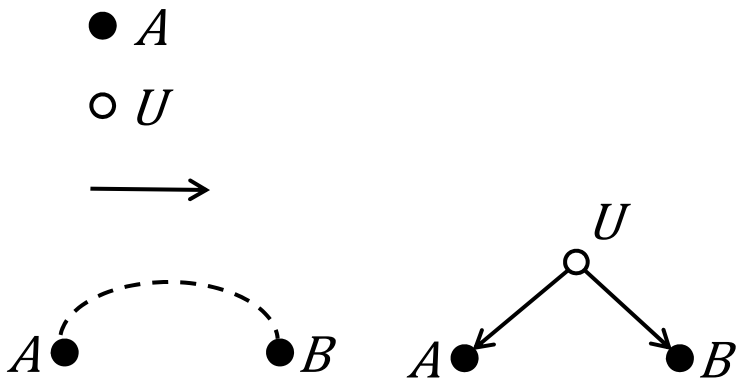
Kapitel 1:
Kausalität in den Sozialwissenschaften

Ziele empirischer Sozialforschung

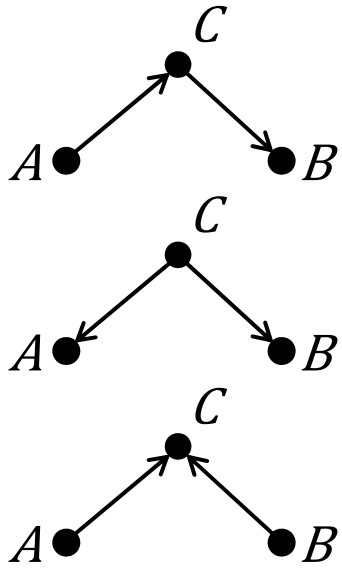
- Beschreibung (Deskription)
  - Die genaue Beschreibung der sozialen Welt kommt zuerst
    - „It is a capital mistake to theorize in advance of the facts“ (Warnung von Sherlock Holmes an Dr. Watson)
- Erklärung
  - Hat man die Fakten geklärt, kann man nach den Ursachen fragen
    - Dies ist die Suche nach Kausaleffekten (Kausalanalyse)
- Politikberatung
  - Hat man Fakten und kausale Zusammenhänge, so kann man politische Maßnahmen empfehlen
    - Dazu benötigt man aber Vorstellungen über den Soll-Zustand (politisches Ziel). Da dies normativ ist, nicht Teil der Wissenschaft.
  - Bsp.: Anstieg der Einkommensungleichheit
    - In welchen Einkommensgruppen verändert sich was?
    - Dann kann man nach den Ursachen fragen
    - Dann kann man politische Maßnahmen empfehlen

Kausaldiagramme (DAGs)

- Knoten: beobachtete Variable
- Offener Knoten: latente Variable
- Gerichtete Kante: **Kausalbeziehung**
- Bidirektionale Kante: **Assoziation** durch gemeinsame latente Ursache („confounder“)



- Die drei grundlegenden Muster der Kausalbeziehung dreier Variablen
  - Intervenierende Variable
  - Konfundierende Variable
  - Collider Variable



Was ist Wissenschaft?

- Ziel: Wissenschaft will Wissen schaffen
  - Nicht nur moralisieren und/oder politisieren
- Methode I:

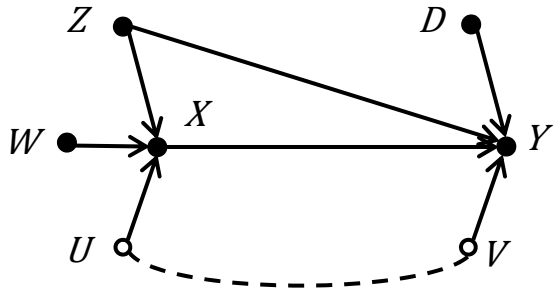
Lee Smolin, Professor für theoretische Physik

"[W]ir müssen darauf bestehen, sie [die Ideen] testen – oder falsifizieren – zu können. Darauf beruht der Fortschritt der Wissenschaft in den vergangenen 400 Jahren. Wenn man eine theoretische Struktur hat, die nichts erklärt und nichts vorhersagt, hört man auf, Wissenschaft zu betreiben. Dann haben wir es mit der Gefahr von nicht überprüfbaren Theorien [...] zu tun.

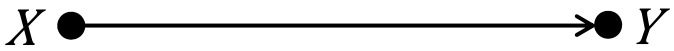
- Methode II: Konsequentes Anzweifeln aller Ergebnisse (auch der eigenen!)
  (aus: Richtlinien der LMU München zur Selbstkontrolle in der Wissenschaft)

Theorie

- Kausalanalysen basieren auf Theorien
  - Eine “Theorie” ist eine Menge von miteinander verknüpften und logisch konsistenten Aussagen, von denen eine nichtleere Teilmenge empirisch prüfbare Aussagen (Hypothesen) sind



- Das Geschäft der (kausalanalytischen) Sozialforschung ist die empirische Überprüfung der Gültigkeit von (aus Theorien abgeleiteten) Hypothesen
  - Z.B.:



Exkurs: Mängel soziologischer Theorien

- Viele soziologische Theorien sind allerdings wissenschaftlich unfruchtbar
  - Oft ist unklar, welche theoretischen Konzepte eigentlich eine wesentliche Rolle spielen
  - Viele Theorien sind zu ungenau, um falsch zu sein
  - Viele Theorien lassen keine Herleitung von empirisch überprüfbaren Hypothesen zu, weil sie u.a. Konstrukte verwenden,
    - welche keine Entsprechung in der Realität haben
    - und/oder kaum operationalisier- und messbar sind

Empirische Überprüfung I

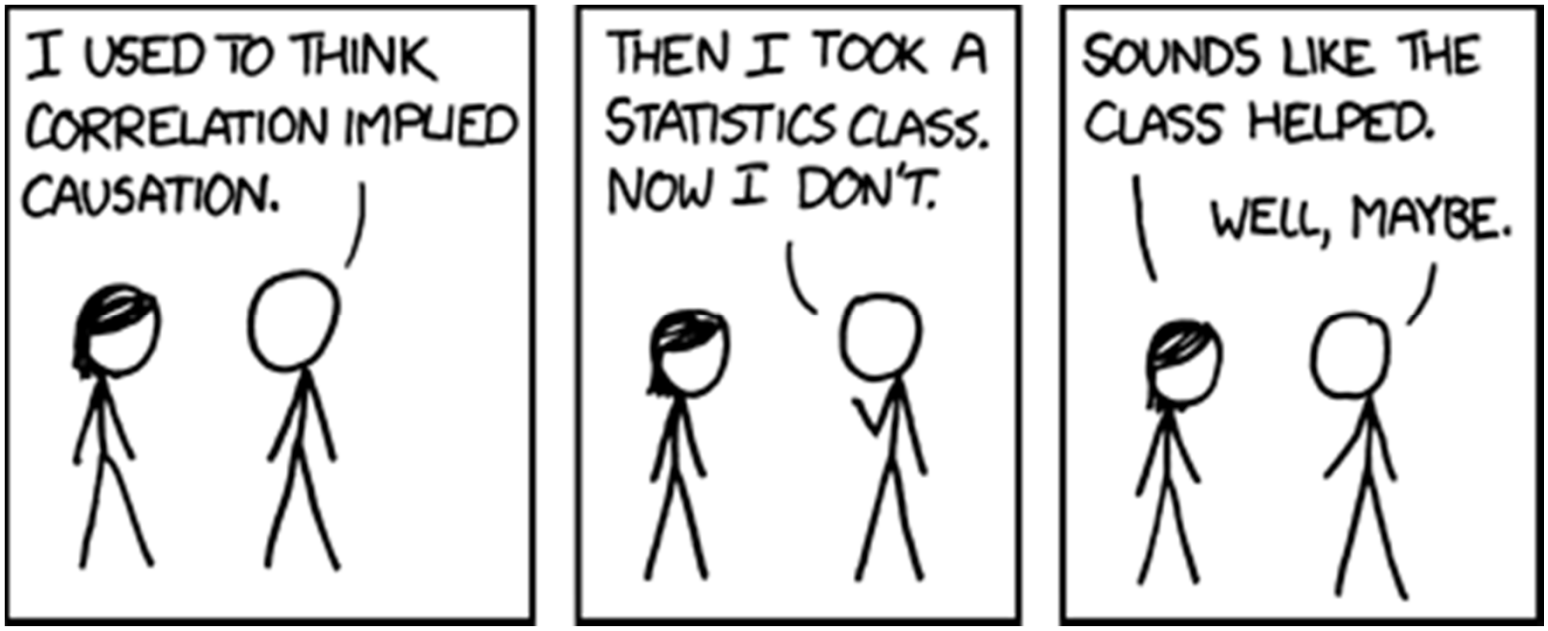
- DIE wissenschaftliche Methode: das Experiment
  - Versuchs- und Kontrollgruppe
  - Randomisierung
    - Dadurch unterscheiden sich Versuchs- und Kontrollgruppe nicht
    - Keine unbeobachtete Heterogenität
  - Kontrollierte Stimulussetzung durch Forscher
    - Damit ist sichergestellt, dass die uV der aV zeitlich vorgeht
    - Keine Endogenität
- Ein sauber durchgeführtes Experiment erlaubt einen sicheren Kausalschluss
- Experimente sind aber in den Sozialwissenschaften oft nicht praktikabel

Empirische Überprüfung II

- Deshalb erhebt man oft Daten über X und Y ex-post-facto und berechnet deren Korrelation
- Korrelation ist aber nicht gleich Kausalität
  - Es könnte auch eine „Scheinkorrelation“ vorliegen (s.u.)

- Um von einer Korrelation auf Kausalität schließen zu können, müssen folgende Bedingungen gelten:
  - X und Y sind korreliert
  - X geht Y zeitlich voran (keine Endogenität)
  - Die Korrelation von X und Y bleibt erhalten, auch wenn man für dritte Variablen kontrolliert (keine unbeobachtete Heterogenität)

Korrelation ist nicht gleich Kausalität



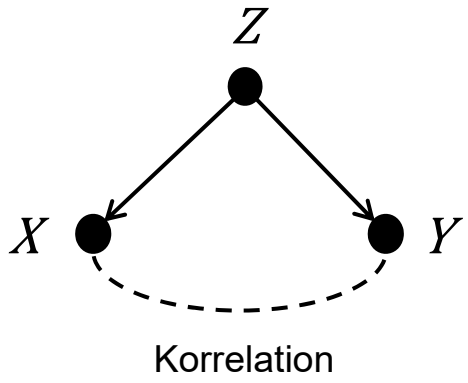
https://xkcd.com/552/

Das Problem: Selbstselektion

- Bei einem Experiment werden die Vpn vom Forscher den beiden Gruppen per Randomisierung zugewiesen
- Bei ex-post facto Designs überlässt man es den Personen selbst, in welche Gruppe sie gehen (Selbstselektion)
- Das führt leicht zu unbeobachteter Heterogenität
- Selbstselektion ist **das** allgegenwärtige methodische Problem in der Sozialforschung!
  - Beispiele, bei denen dieses Problem nicht bedacht wurde und die so (oder ähnlich) in Studien berichtet wurden und werden(!)
    - Ehemänner leben länger
    - Ehemänner verdienen mehr
    - Ärmere Menschen leben kürzer
    - Bewohner von Betonblöcken sind häufiger krank
    - Häufiges Fernsehen auf Privatsendern macht dumm
    - Zähneputzen senkt das Herzinfarktrisiko

Folge: Scheinkorrelation/Konfundierung

- X und Y korrelieren zwar, aber Grund hierfür ist eine dritte Variable Z, die sowohl X als auch Y kausal verursacht
  - Die Korrelation ist „echt“, aber die Kausalität ist „scheinbar“ (Scheinkausalität)
- Schematisch anhand eines DAG



- Z ist eine „antezedierende“ Variable (Drittvariable, Confounder)
- Durch die beiden Kausaleffekte entsteht eine Korrelation von X und Y
- Es wäre ein Fehler diese Korrelation als kausal zu interpretieren

- Durch Kontrolle von Z (Drittvariablenkontrolle) kann man das Problem beheben

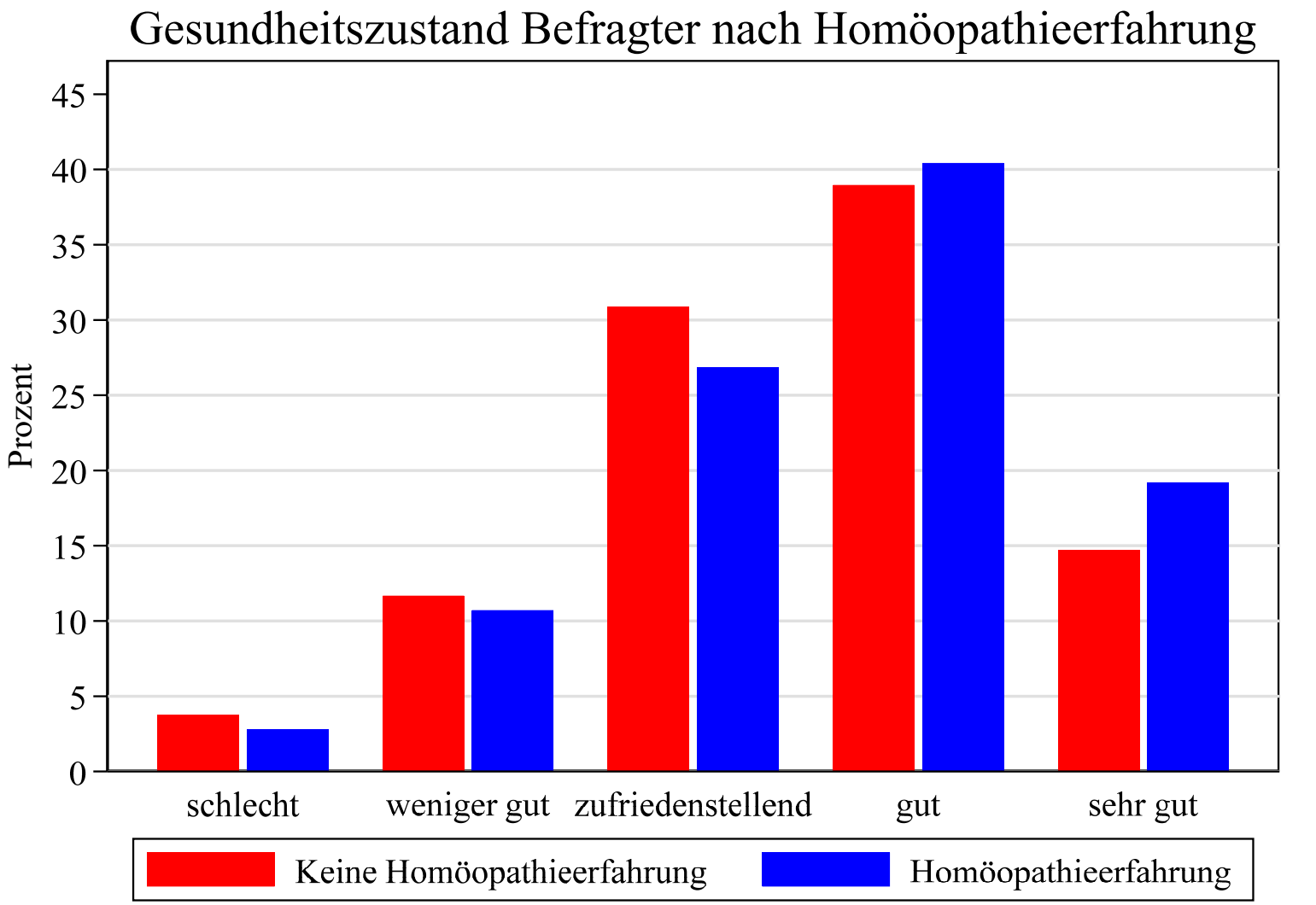
Drittvariablenkontrolle

- Z.B. mittels konditionaler Kreuztabellen (Partialtabellen)
  - Z wird konstant gehalten: Für jede Ausprägung von Z wird eine eigene Kreuztabelle (X × Y) erstellt (dreidimensionale Kreuztabelle)
  - Damit erhält man für jede Ausprägung von Z einen eigenen, konditionalen Korrelationskoeffizienten:

$$r_{XY \bullet Z_1}, r_{XY \bullet Z_2}, \text{usw.}$$

  - Die messen die Korrelation von X und Y unter Kontrolle von Z
    - Damit ist Z jeweils konstant und kann nicht mehr die Ursache für eine eventuelle Korrelation von X und Y sein
    - Sind die konditionalen Korrelationskoeffizienten ungleich Null, so können wir von Kausalität ausgehen
- Problem: es kann mehrere Drittvariablen geben
  - Lösung: multivariate Analyseverfahren (Regression)
- Regression hat in der Sozialforschung eine zentrale Rolle: Sie ist **der** Ersatz für das Experiment

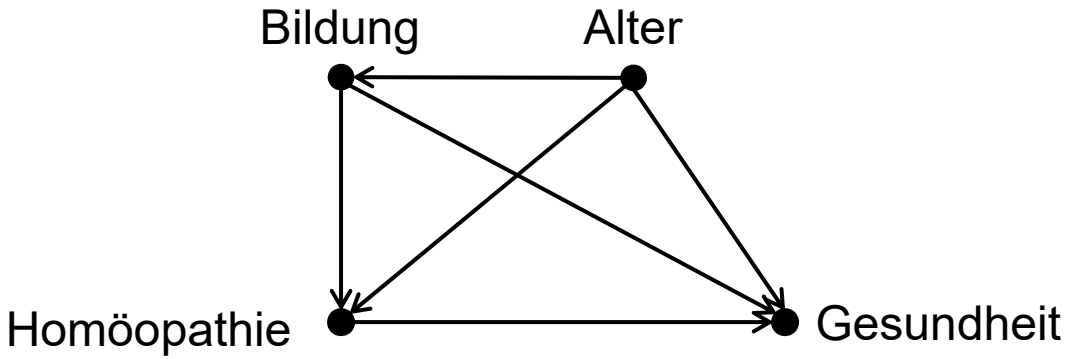
Bsp.: Macht Homöopathie gesund?



Quelle: Idee von Stefanie Heyne
Daten: Allbus 2012
Do-File: 00a Homoeopathie

Bsp.: Macht Homöopathie gesund?

- Dieser Zusammenhang ist evtl. eine Scheinkorrelation
  - Zwei potentielle Confounder: Bildung und Alter



Bsp.: Macht Homöopathie gesund?

	(1)	(2)	(3)
Erfahrung mit homoe	0.129***	-0.002	-0.029
	(0.035)	(0.035)	(0.033)
Niedr. Bildung (Ref.)			
Mittlere Bildung		0.390***	0.196***
		(0.040)	(0.040)
Hohe Bildung		0.673***	0.438***
		(0.043)	(0.043)
Alter			-0.017***
			(0.001)
_cons	3.491***	3.194***	4.185***
N	3419	3419	3419
R-sq	0.004	0.073	0.148

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Daten: Allbus 2012
Do-File: 00a Homoeopathie

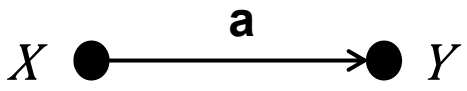
Abhängige Variable: Gesundheitseinschätzung 1 - 5

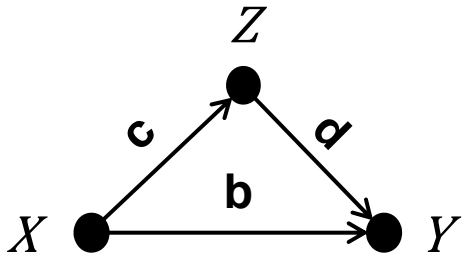
Methoden der Kausalanalyse

- „Selection on observables“: Confounder gemessen
  - Man kann für sie kontrollieren
    - Regression **(1. Sem.: „Querschnittsdatenanalyse“)**
    - Matching **(3. Sem.: „Kausalanalyse“)**
- „Selection on unobservables“: Confounder nicht gemessen
  - Regressions- und Matching-Schätzer sind verzerrt
    - Unbeobachtete Heterogenität, „omitted variable bias“
  - Unverzerrte Schätzer kann man erhalten mit
    - Instrumentalvariablen Ansatz **(3. Sem.: „Kausalanalyse“)**
      - Exogene Variation identifiziert den Kausaleffekt
    - Regression Discontinuity Ansatz **(3. Sem.: „Kausalanalyse“)**
      - Homogenität von Personen an einer Schwelle
    - Within-Panelanalyse **(2. Sem.: „Längsschnittdatenanalyse“)**
      - Homogenität einer Person über die Zeit

Mediation

- Bei Konfundierung ist Z „antezedierend“ (zeitlich vor X und Y)
- Ist Z intervenierend (zeitlich zwischen X und Y), so liegt eine Mediation vor (auch „Intervention“ genannt)
- Eine Mediation ist im Unterschied zur Konfundierung kein Problem der Kausalanalyse, sondern der 2. Schritt
  - Z ist ein „kausaler Mechanismus“
  - Man weiß nun, wie der Kausaleffekt von X auf Y zustande kommt





- (Totaler) Kausaleffekt: a
- Direkter Kausaleffekt: b
- Indirekter Kausaleffekt: c · d
- Es gilt: a = b + c · d
- Kausalanalyse
  - 1. Schritt: schätze den totalen KE
  - 2. Schritt: kontrolliere für Z
    - b < a : Z ist ein Mechanismus

Beispiel: Kirchgangshäufigkeit

- Mit dem ALLBUS 1994 untersuchen wir, wie sich der Wohnort (West/Ost) auf den Kirchgang auswirkt

Kirchgang nach Wohnort

	West	Ost
Selten/nie	55%	84%
Öfter	45%	16%
N	2339	1104

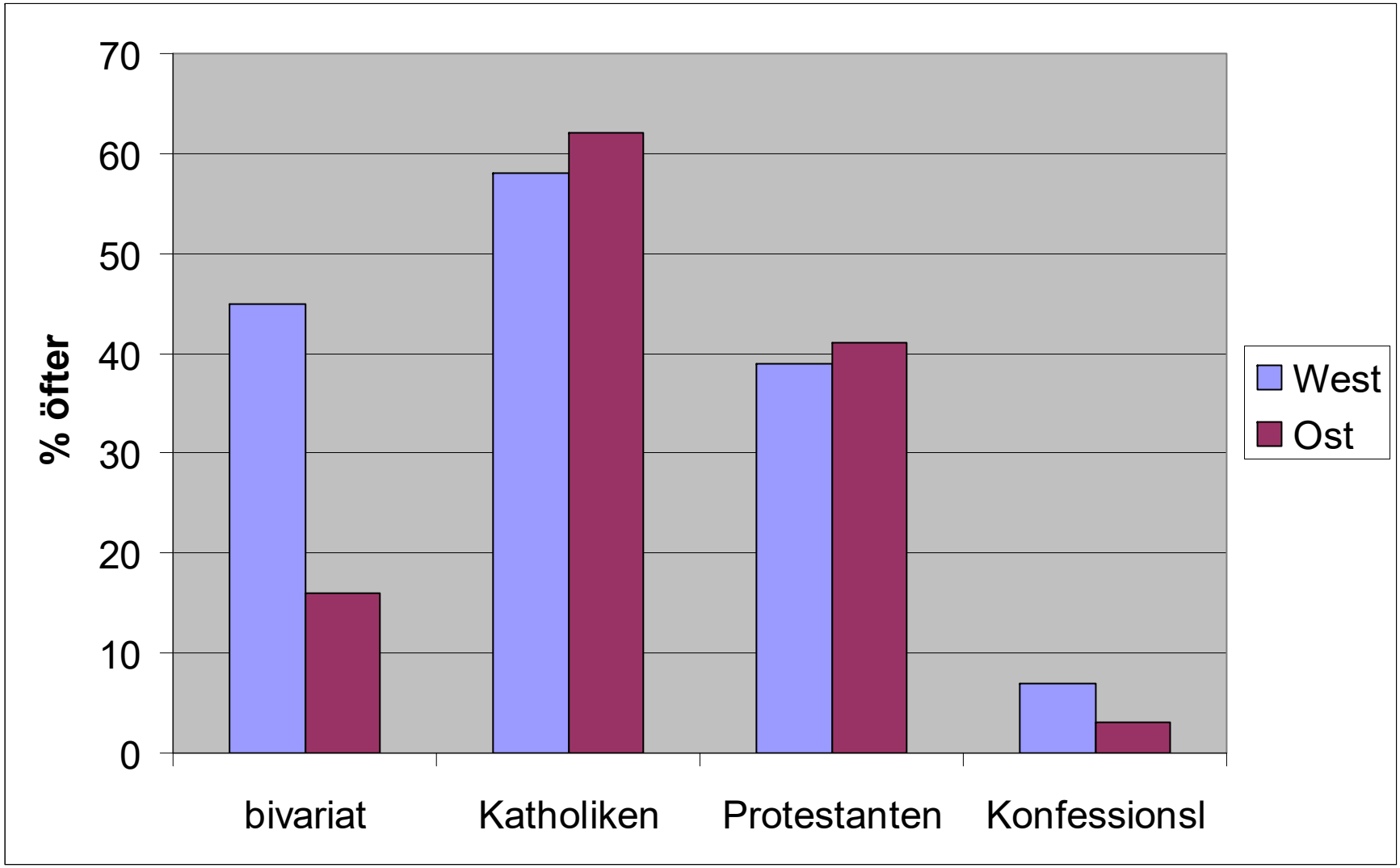
V = 0,28

- Könnte „Konfession“ ein intervenierender Mechanismus sein? Deshalb kontrollieren wir für Konfession (2 × 2 × 3-Tabelle)

Kirchgang nach Konfession und Wohnort

	Katholiken		Protestanten		Konfessionslose	
	West	Ost	West	Ost	West	Ost
Selten/nie	42%	38%	61%	59%	93%	97%
Öfter	58%	62%	39%	41%	7%	3%
	V = 0,01		V = 0,01		V = 0,07	

Beispiel: Kirchgangshäufigkeit



Beispiel: Kirchgangshäufigkeit

- Um den Kausalmechanismus ganz zu verstehen, erstellen wir noch die Kreuztabellen X × Z und Z × Y

Konfession nach Wohnort

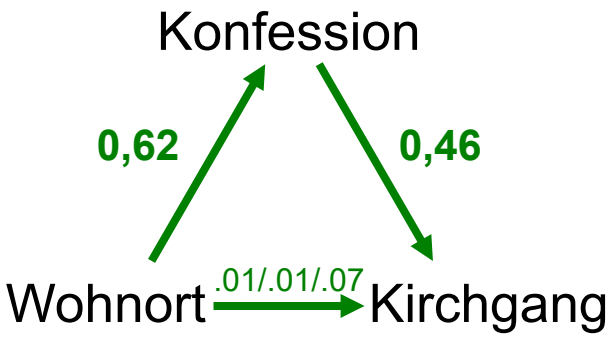
	West	Ost
Katholik	47%	3%
Protestant	41%	26%
Konfessionsl.	12%	71%

V = 0,62

Kirchgang nach Konfession

	selten	öfter
Katholik	41%	59%
Protestant	60%	40%
Konfessionsl.	96%	4%

V = 0,46



Das gesamte Kausalmodell präsentieren wir übersichtlich in einem „Pfaddiagramm“. An die Pfeile schreiben wir die bivariaten Korrelationskoeffizienten. Die konditionalen Korrelationskoeffizienten von „Wohnort“ auf „Kirchgang“ sind fast null und machen deutlich, dass hier praktisch kein direkter Kausaleffekt vorliegt. „Konfession“ ist der kausale Mechanismus, der den KE von „Wohnort“ auf „Kirchgang“ vollständig erklärt.

Extrem: Suppression

- Falls der indirekte Effekt ein anderes Vorzeichen hat, als der direkte Effekt, fällt der totale Effekt kleiner aus

$$a = b + c \cdot d$$

  - In extremen Fällen ist der totale Effekt gar nahe Null (verdeckte Korrelation, Suppression)
  - Oder hat gar das entgegengesetzte Vorzeichen

- Bsp.: Diskriminierung von Frauen bei der Studienzulassung?
  - Aus: Krämer, Walter (1995) Denkste! Campus-Verlag
  - Fiktives Beispiel

	M	F	Σ
nicht zugel.	400	450	850
zugelassen	100 (20%)	50 (10%)	150
Σ	500	500	1000

Φ = (-) 0.14

Beispiel: Studienzulassung

Kontrolliert man für die intervenierende Variable „Studienfach“ verschwindet der Effekt: es gibt nur einen schwachen direkten Effekt.

Mathe

	M	F	Σ
nicht zug.	100	10	110
zug.	80 (44%)	10 (50%)	90
Σ	180	20	200

Φ = (+) 0.03

SoWi

	M	F	Σ
nicht zug.	300	440	740
zug.	20 (6%)	40 (8%)	60
Σ	320	480	800

Φ = (+) 0.04

Beispiel: Studienzulassung

Frauen bewerben sich häufiger für Sowi

	M	F	Σ
Mathe	180	20	200
Sowi	320	480	800
Σ	500	500	1000

Φ = (+) 0.40

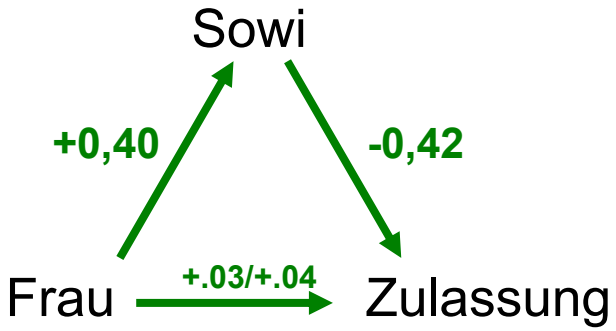
Sowi hat niedrigere Zulassungsquoten als Mathe

	Mathe	Sowi	Σ
nicht zug.	110	740	850
zug.	90	60	150
Σ	200	800	1000

Φ = (-) 0.42

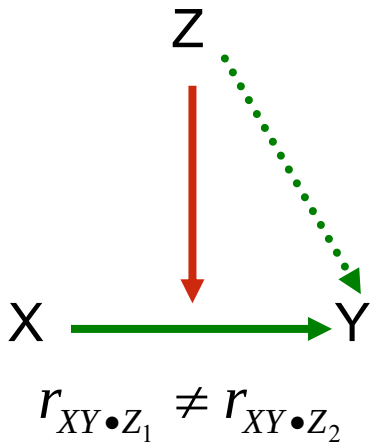
- Der direkte Effekt ist Null
- Der indirekte Effekt ist negativ
- Deshalb ist der totale Effekt negativ

- Der totale Effekt wird vollständig durch die Studienfachwahl erklärt
  - Es gibt keine Diskriminierung



Moderation (Interaktion)

Die Beziehung von X und Y fällt unterschiedlich aus, je nachdem welchen Wert Z annimmt (Z heißt auch „Moderator“)



Beispiele:

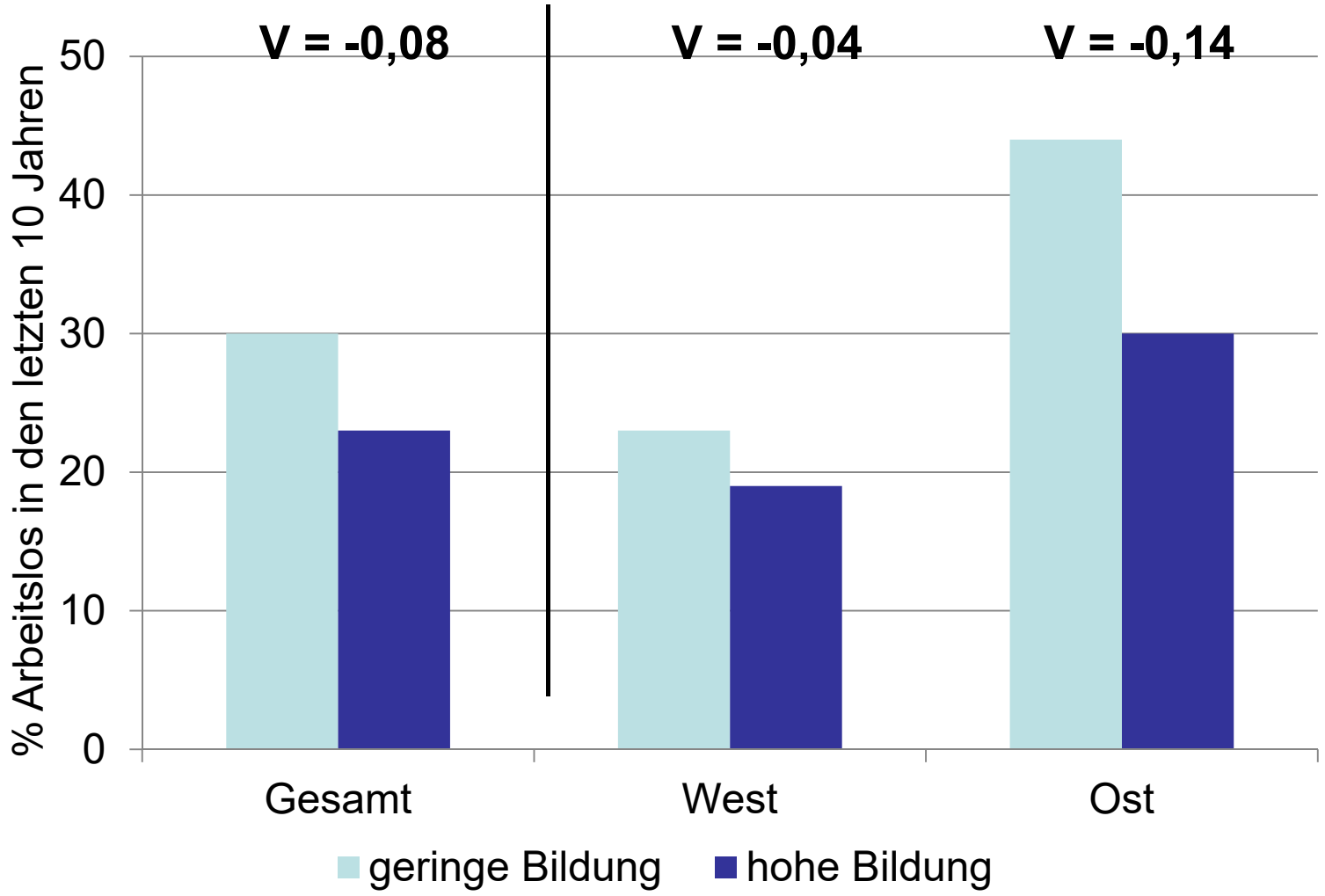
- Sport (X), Gesundheit (Y), Erkältung (Z)
- Einsatz (X), Erfolg (Y), Motivation (Z)

Beispiel: M. Halbwachs (1930) Les Causes du Suicide

Halbwachs stellte fest, dass es einen Zusammenhang zwischen Konfession und Selbstmordrate gibt: Katholiken 19,9 Selbstmorde (pro 100.000), Protestanten 39,6 Selbstmorde (pro 100.000). Kontrolliert man den Wohnort, so verschwindet der Zusammenhang für Städte, auf dem Land wird er stärker.

Wohnort	Katholik	Protestant
Stadt	39,9	37,8
Land	8,8	41,4
Alle	19,9	39,6

Beispiel: Arbeitslosigkeit in Abhängigkeit von Bildung



Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 2:
Explorative Datenanalyse

Anscombes Quartett

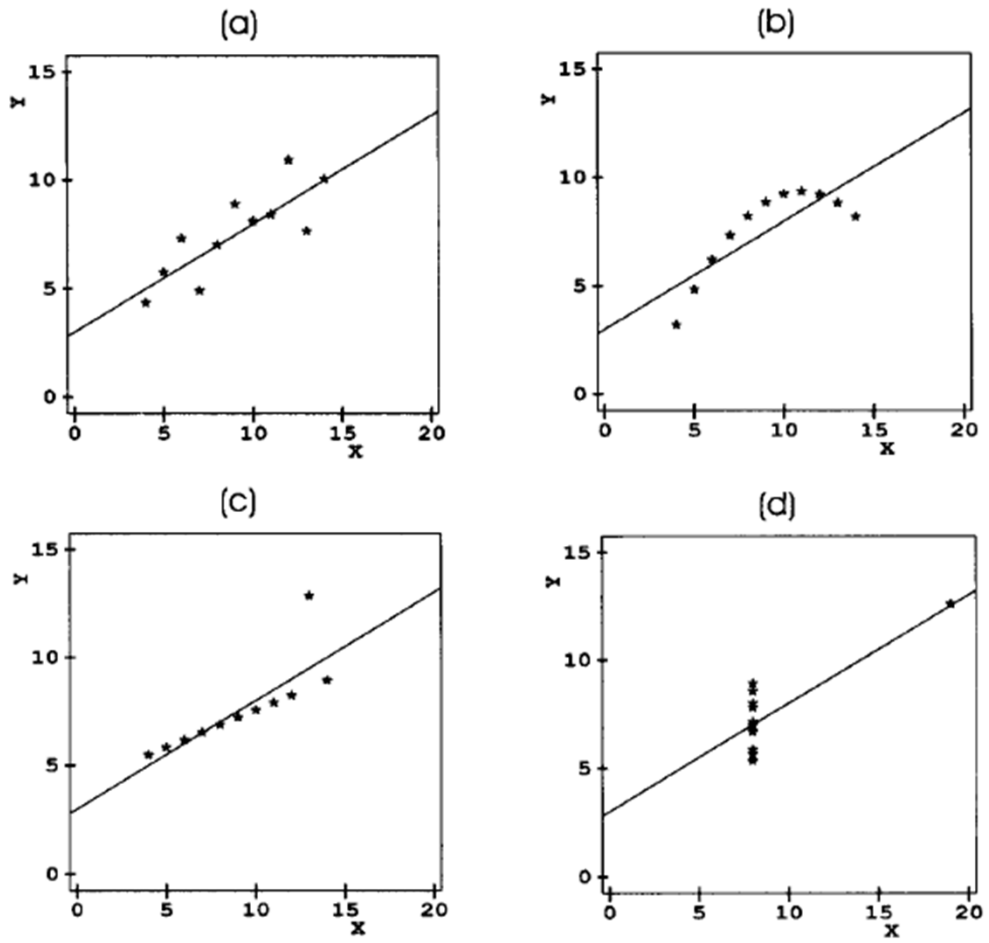
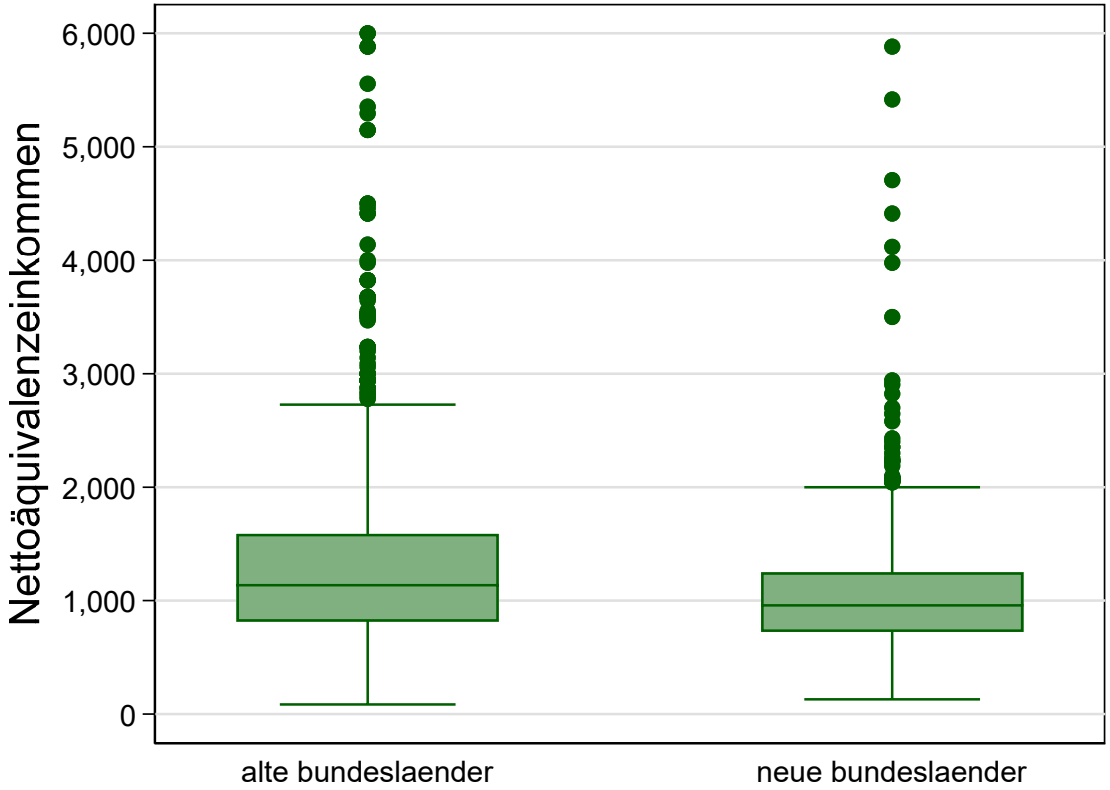


Figure 3.1. Four datasets, due to Anscombe (1973), with identical least-squares regressions. In (*a*), the linear regression is an accurate summary; in (*b*), the linear regression distorts the curvilinear relationship between Y and X; in (*c*), the linear regression is drawn toward an outlier; in (*d*), the linear regression "chases" the influential observation at the right.

Univariate Verteilungen
Bsp.: Armut in Deutschland

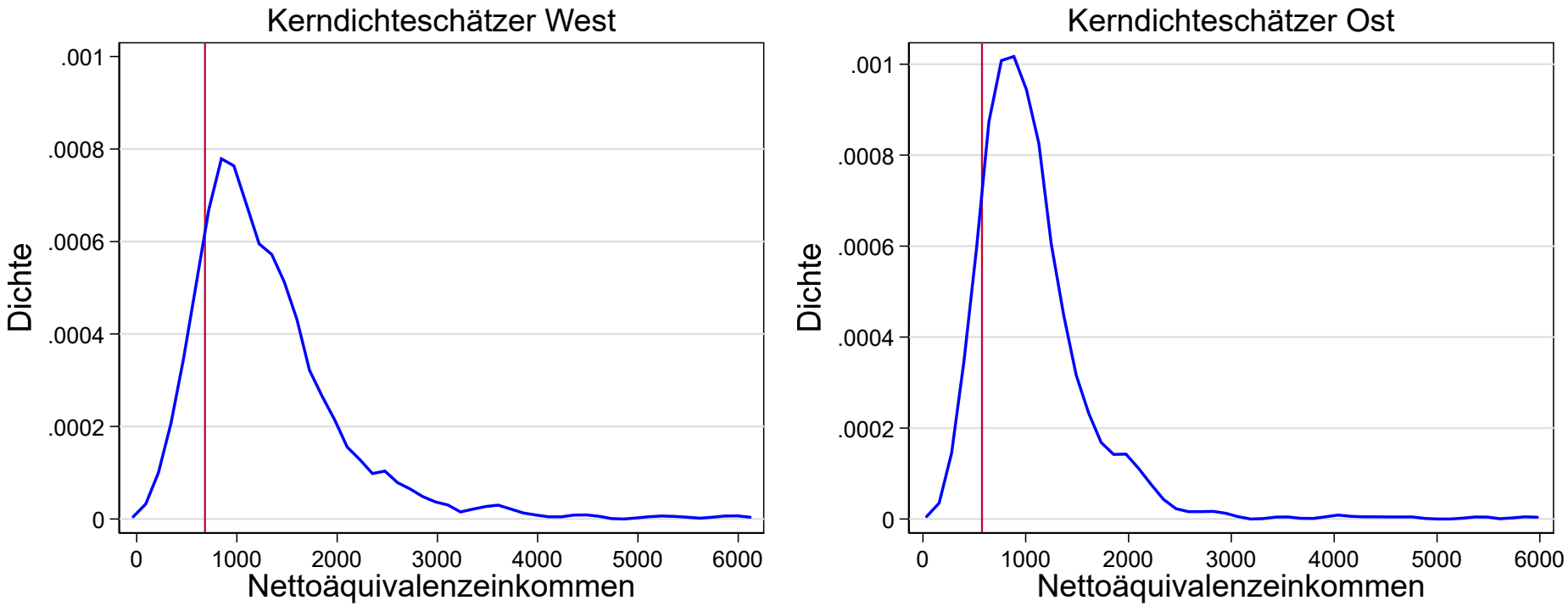
```
* Gruppiertes Boxplot West/Ost
graph box oecdeink, over(v3)
```



	West	Ost
Armutsgrenze (60% Median)	682	575
Armutsquote	15,9%	12,1%

Daten: ALLBUS 2002
Do-File: 00b Armut.do

Beispiel: Armut in Deutschland



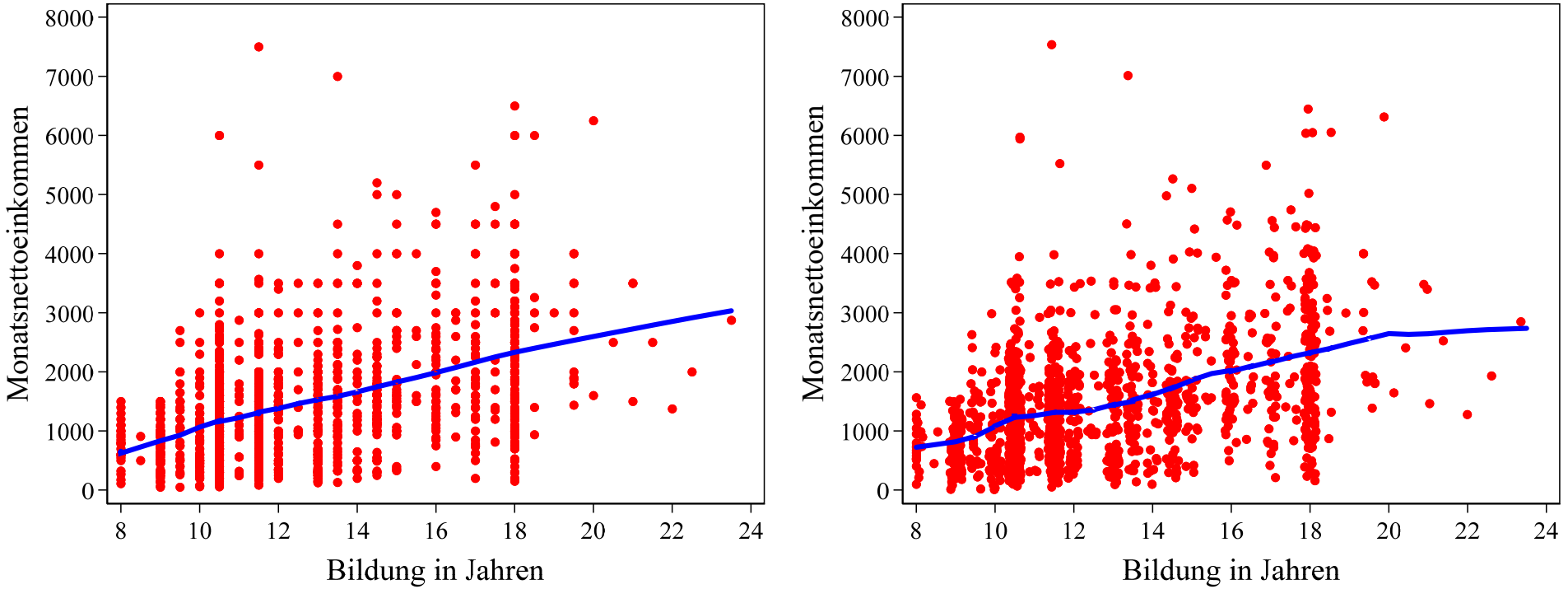
Eingezeichnet sind die Armutsgrenzen

Daten: ALLBUS 2002
Do-File: 00b Armut.do

Bivariate Verteilungen: Scatterplot

- Bivariate Zusammenhänge veranschaulicht man am besten mit einem Streudiagramm
  - Überdecken sich die Daten stark, so, „jittered“ man am besten
  - Einen Eindruck von der Art des Zusammenhangs bekommt man mittels einer nicht-parametrischen Regression. Bewährt hat sich hierfür der Lowess-Smoother (locally weighted scatterplot smoother).
    - An der Stelle x_i wird eine lineare Regression berechnet, in die die Daten in der Umgebung gewichtet eingehen. Die Breite der Umgebung ist steuerbar durch ,,bandwidth“ (z.B. bwidth=0.8). Es wird trikubisch gewichtet. Anhand der Regressionsparameter wird dann $\widehat{y_i}$ berechnet. Dies wird für alle X-Werte gemacht. Die Verbindung der $(x_i, \widehat{y_i})$ ergibt die Lowess-Kurve. Je kleiner die Umgebung, desto näher an den Daten ist die Kurve.

Bivariate Verteilungen: Scatterplot

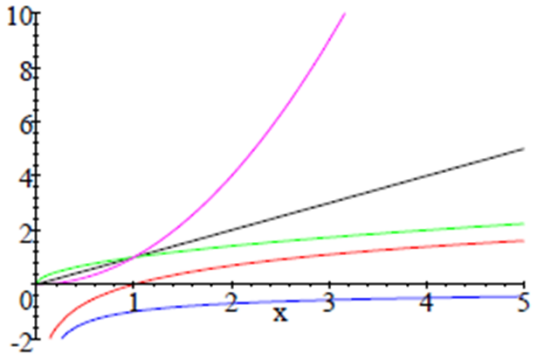


Daten: ALLBUS 2002
Do-File: 1 Regression.do

- Beispiel: Einkommen in Abhängigkeit von Bildung
  - Nur Westdeutschland, max. 8 Tsd. Euro (N=1.468)
  - Links ist nicht gejittered, es kommt zu starker Überdeckung. Rechts ist gejittered („jitter(2)“: 2% der Zeichenfläche)
  - Die blaue Kurve ist der Lowess-Smoother
    - Links werden zur Berechnung jeweils 80% der Fälle in der Umgebung verwendet, rechts nur 30%. Die rechte Kurve folgt deshalb wesentlich genauer den Daten, ist dafür aber unregelmäßiger.
  - In beiden Fällen erkennt man kaum Nicht-Linearitäten

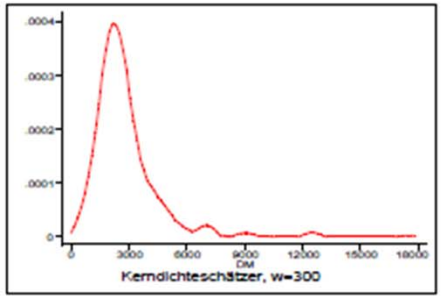
Exkurs: Datentransformationen

- Schiefe und Ausreißer sind für Regressionen ein Problem
- Durch Potenz-Transformationen kann man Schiefe reduzieren
- Tukeys “ladder of powers”

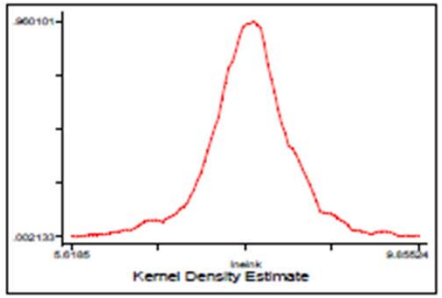


x^3	$q = 3$		produziert
$x^{1.5}$	$q = 1.5$	cyan	Rechtsschiefe
x	$q = 1$	schwarz	
$x^{.5}$	$q = .5$	grün	produziert
$\ln x$	$q = 0$	rot	Linksschiefe
$-x^{-.5}$	$q = -.5$	blau	

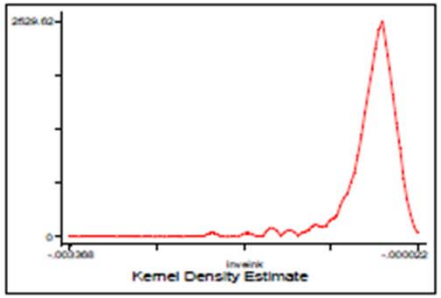
Beispiel: Einkommensverteilung



q=1



q=0



q=-1

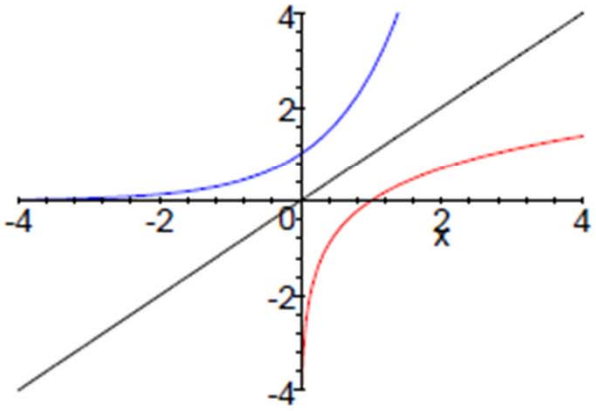
Exkurs: Potenzfunktionen, ln und e

$$x^{0.5} = x^{\frac{1}{2}} = \sqrt[2]{x}, \qquad x^{-0.5} = \frac{1}{x^{0.5}} = \frac{1}{\sqrt[2]{x}}, \qquad x^0 = 1$$

Mit $\ln$ notieren wir den (natürlichen) Logarithmus zur Basis $e = 2,71828\ldots$:

$$y = \ln x \Leftrightarrow e^y = x$$

Daraus folgt $\ln(e^y) = e^{\ln y} = y$.



Rechenregeln

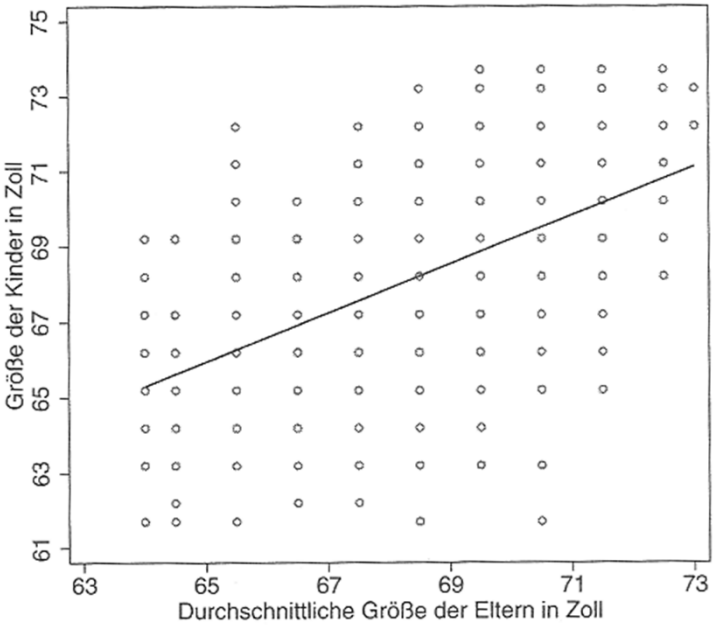
$e^x e^y = e^{x+y}$	$\ln(xy) = \ln x + \ln y$
$e^x / e^y = e^{x-y}$	$\ln(x/y) = \ln x - \ln y$
$(e^x)^y = e^{xy}$	$\ln x^y = y \ln x$

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 3:
Einführung in die Regression

Zum Begriff „Regression“

- Regression toward the mean:
  „The stature of the adult offspring must on the whole, be more *mediocre* than the stature of their parents“ (Sir Francis Galton, 1889)
- Galton fittete (visuell!) eine Gerade
  - Sein Ergebnis: Steigung von 0,67
- Später wurde dies mit OLS gemacht und auf das Fitten von Geraden mit OLS der Begriff „Regression“ übertragen
  - OLS Ergebnis: Steigung von 0,64
- Die erste inhaltliche Anwendung prägte also den Begriff für ein statistisches Verfahren!



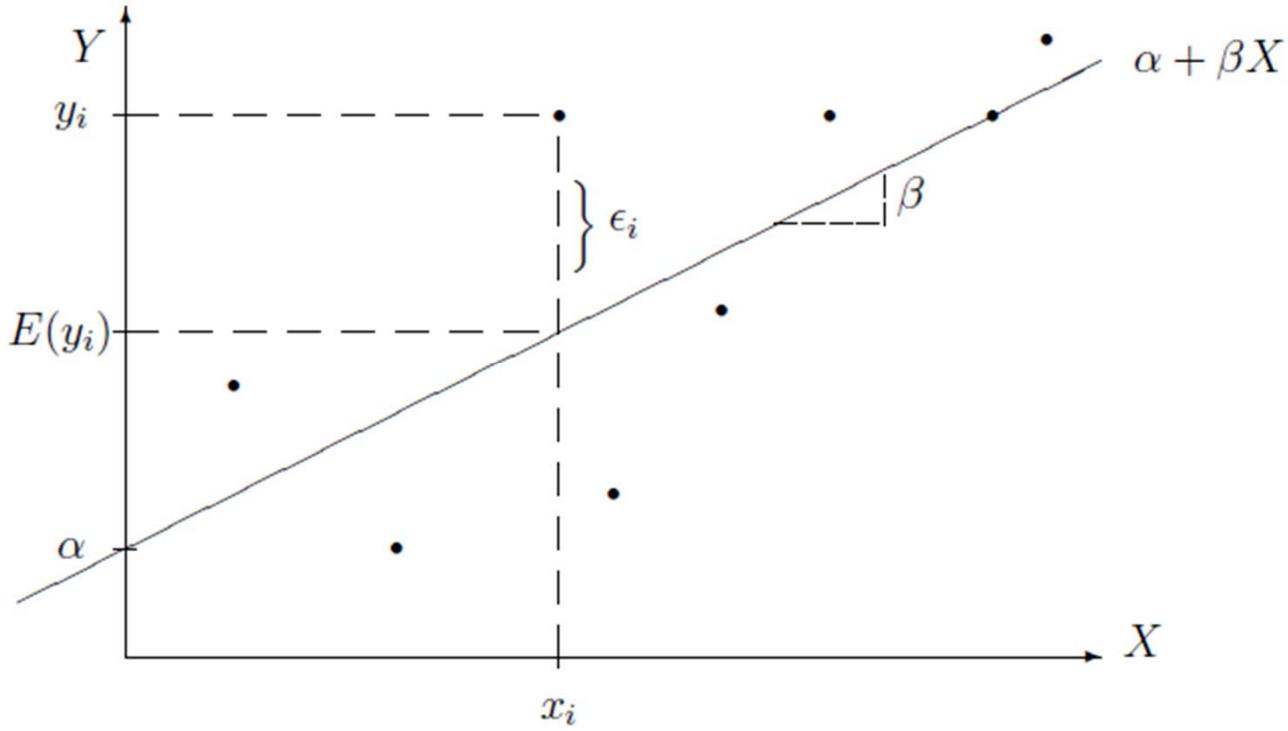
Galtons Daten mit Regressionsgerade
Quelle: Fahrmeir et al. (2007)

Das einfache Regressionsmodell

- Man formuliert folgendes lineare Modell des Zusammenhangs:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

- α und β sind die Regressionskoeffizienten
  - α: Achsenabschnitt, β: Steigung
  - β: um wie viel Einheiten ändert sich Y, wenn X um eine Einheit steigt
- ε_i ist der Fehlerterm (Abweichung der Daten von der Modellgerade)



OLS-Schätzer

- Man schätzt die Regressionskoeffizienten, indem man die Fehlerquadratsumme minimiert (ordinary least squares, OLS)

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \varepsilon_i^{\,2} = \min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

- Ableiten dieses Ausdrucks, Nullsetzen und Auflösen der beiden daraus resultierenden Gleichungen, liefert die OLS-Schätzer:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

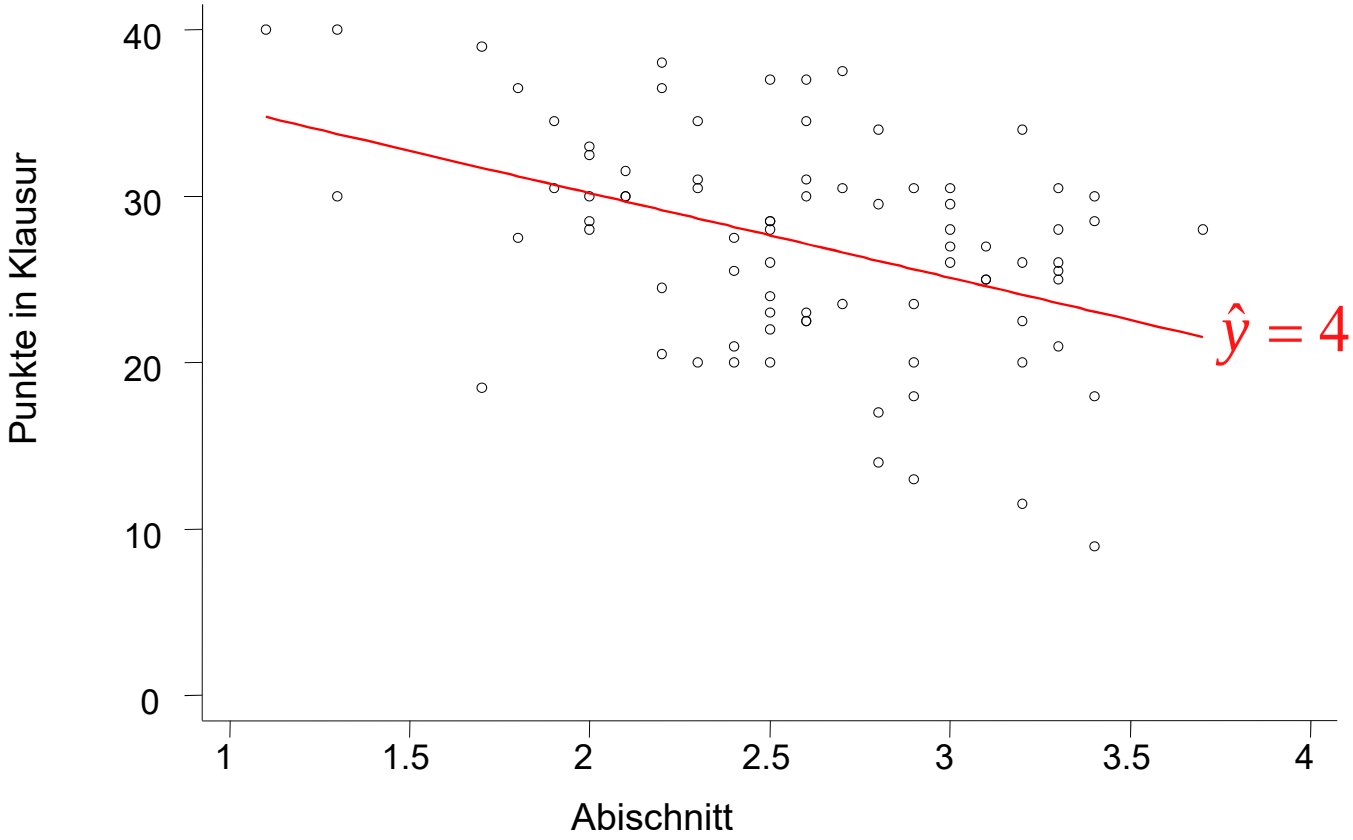
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{XY}}{s_X^{\,2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^2}$$

- Die vom Regressionsmodell vorhergesagten Werte sind $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

- Die geschätzten Fehler (Residuen) sind damit $\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$

Beispiel

Abinote und Klausurerfolg



$$\hat{y} = 40{,}4 - 5{,}1 \cdot x$$

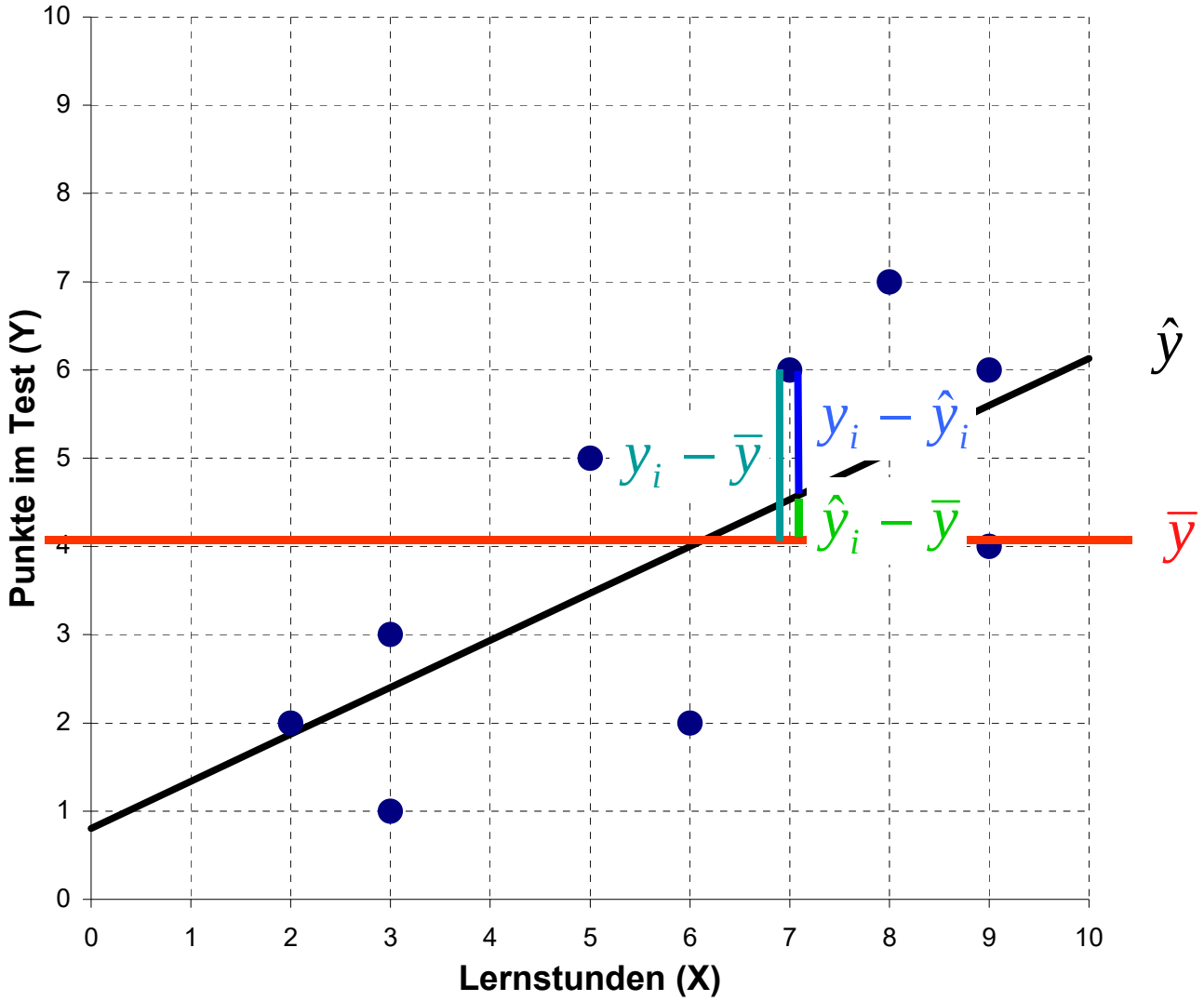
Das Bestimmtheitsmaß R^2

- Wie gut passt das Regressionsmodell auf die Daten?
- Die Grundidee ist: Welcher Anteil der Streuung von Y wird durch das Regressionsmodell „erklärt“?
- Streuungszerlegung
  - Total sum of squares (TSS): $$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$
  - Model sum of squares (MSS): $$MSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$
  - Residual sum of squares (RSS): $$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
  - Die gesamte Streuung kann damit in zwei Teile zerlegt werden

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$TSS = MSS + RSS$$

Graphische Interpretation der Streuungszerlegung



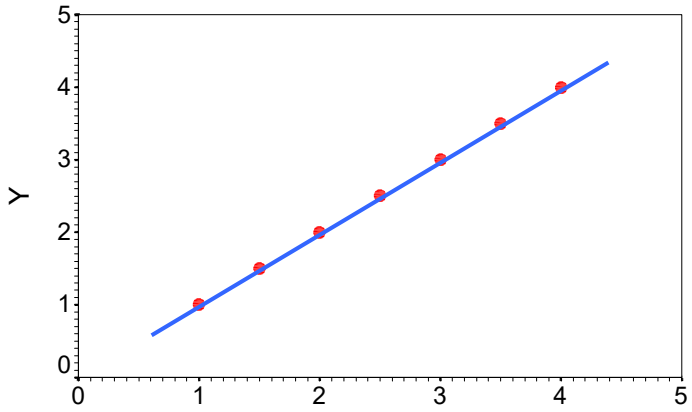
Das Bestimmtheitsmaß R^2

- Das Bestimmtheitsmaß ist nun definiert als

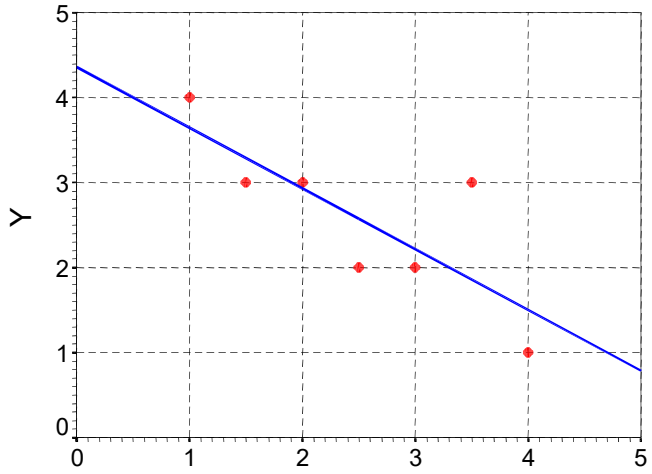
$$R^2 = \frac{\text{erklärte Streuung}}{\text{gesamte Streuung}} = \frac{MSS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- Es gilt: $0 \leq R^2 \leq 1$
- R^2 lässt sich interpretieren als der Anteil der Varianz, der durch die Regressionsgerade (und damit durch X) erklärt wird
- Es gilt: $R^2 = r^2$

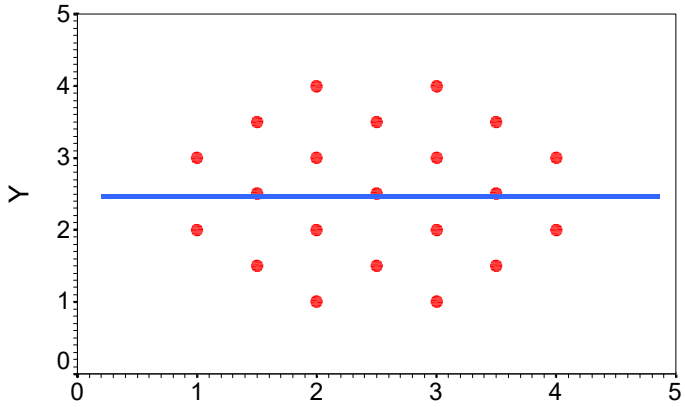
r und R^2



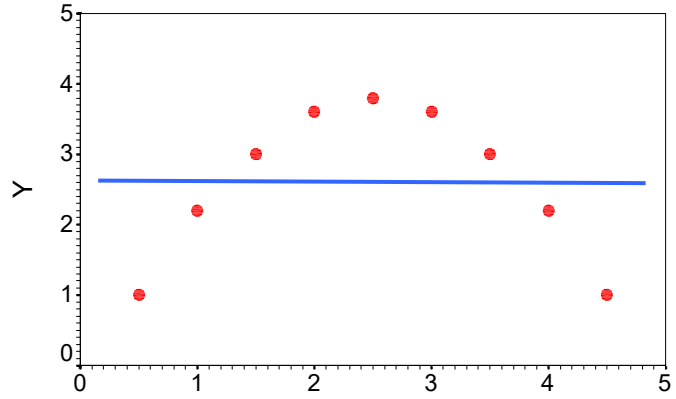
r = 1, R^2 = 1



r = -0.79 , R^2 = 0,62



r = 0 , R^2 = 0



r = 0 , R^2 = 0

Signifikanztest für $\hat{\beta}$

- $\hat{\beta}$ ist ein Schätzer
  - Mit einer Stichprobenverteilung
  - Und einem Standardfehler $\hat{\sigma}_{\hat{\beta}}$
- Damit kann man auch ein Konfidenzintervall schätzen
- Ebenso kann man einen Signifikanztest durchführen
  - Nullhypothese: X hat keinen Einfluss auf Y (kein Zusammenhang)
    H_0: β = 0
  - Die Teststatistik (t-Wert) ist

$$T = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \sim t(n-2)$$

  - Die H_0 wird abgelehnt, falls $|T| > t_{1-\alpha/2}(n-2)$
    - Ab n>30 das $z_{1-\alpha/2}$ Quantil (Faustregel für α=5%: |T| > 2)
  - Können wir die H_0 verwerfen, so spricht man davon, dass X einen signifikanten Einfluss auf Y hat

Annahmen der Regression

- A1: Linearitätsannahme (**Linearität**)

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

- A2: Im Mittel ist der „Fehler“ null

$$E(\varepsilon_i) = 0, \quad \text{für alle i}$$

- A3: Die Fehlervarianz ist konstant (**Homoskedastizität**)

$$V(\varepsilon_i) = \sigma^2, \quad \text{für alle i}$$

- A4: Die Fehlerkovarianzen sind null (**keine Autokorrelation**)

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \quad \text{für alle i} \neq \text{j}$$

- A5: Regressor und Fehler sind unkorreliert (**Exogenität**)

$$Cov(x_i, \varepsilon_j) = 0, \quad \text{für alle i und j}$$

- A6: Fehler normalverteilt (für Sig.tests) (**Normalverteilung**)

$$\varepsilon_i \sim N(0, \sigma^2)$$

Eigenschaften der OLS-Schätzer

- Bei Gültigkeit von A1 bis A5 haben die OLS-Schätzer gewisse wünschenswerte Eigenschaften: Sie sind
- unverzerrt (erwartungstreu): $E(\hat{\beta}) = \beta$
- in der Klasse der linearen, unverzerrten Schätzer die mit der kleinsten Stichprobenvarianz
  - best linear unbiased estimate (BLUE)
  - Gauß-Markov Theorem
- Dabei bedeutet:
  - „linear“: die Schätzer lassen sich als lineare Funktionen der Daten berechnen
  - „unbiased“: die Schätzer sind erwartungstreu
  - „best“: die Schätzer sind effizienter als alle anderen linearen Schätzer

Standardisierte Regressionskoeffizienten

- β hängt von der Maßeinheit von X und Y ab
- Um Vergleichbarkeit herzustellen, wählt man manchmal die Standardabweichung als Maßeinheit
  - Standardisierung von Y und X (Z-Transformation)

$$y_i^* = \frac{y_i - \bar{y}}{s_Y}, \quad x_i^* = \frac{x_i - \bar{x}}{s_X}$$

  - Die Regressionsgleichung lautet nun

$$y_i^* = \alpha^* + \beta^* x_i^* + \varepsilon_i^*$$

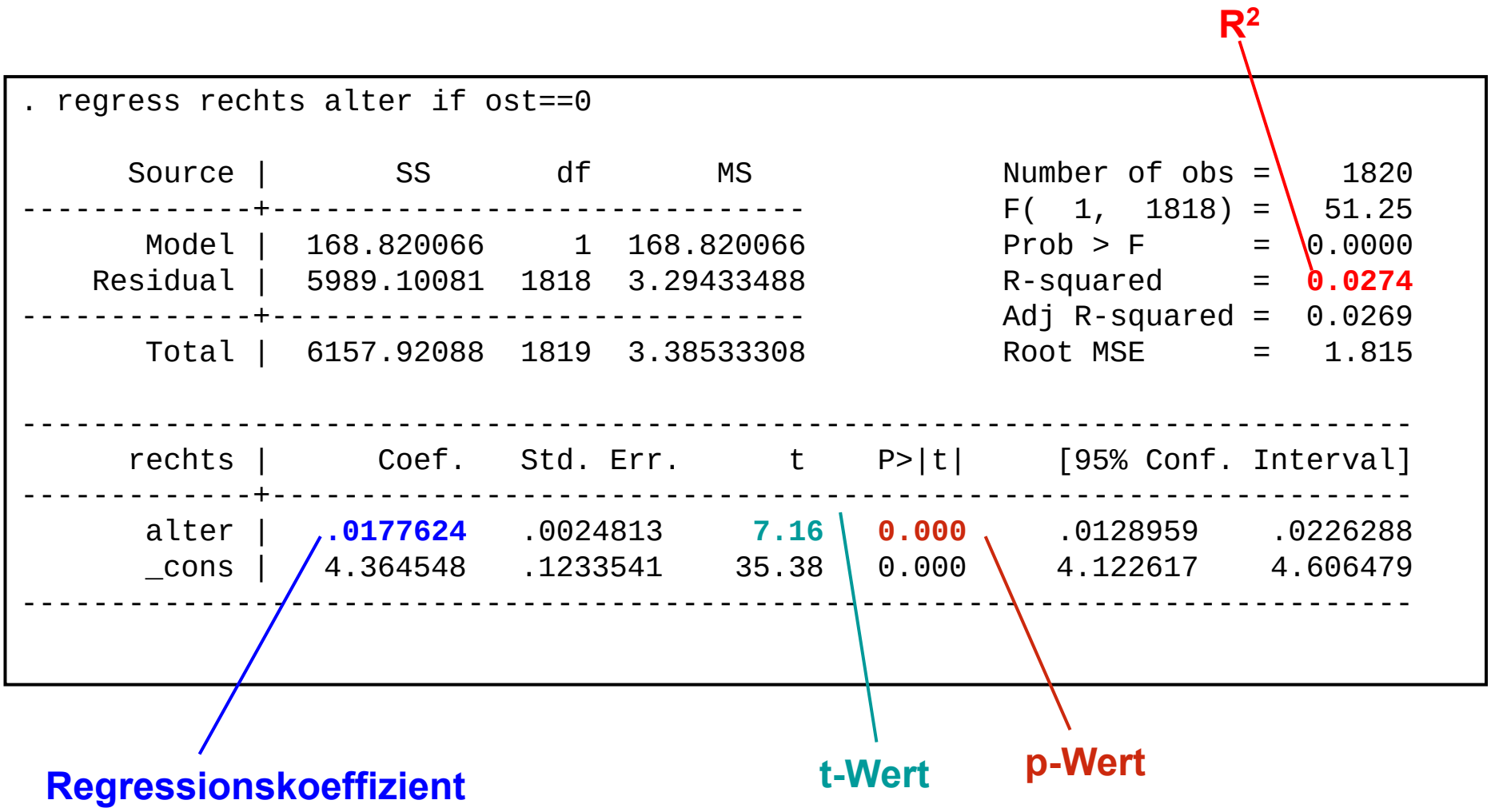
  - Für die standardisierten Regressionskoeffizienten ergibt sich

$$\hat{\alpha}^* = \bar{y}^* - \hat{\beta}^* \bar{x}^* = 0$$

$$\hat{\beta}^* = \frac{s_{X^*Y^*}}{s_{X^*}^2} = r$$

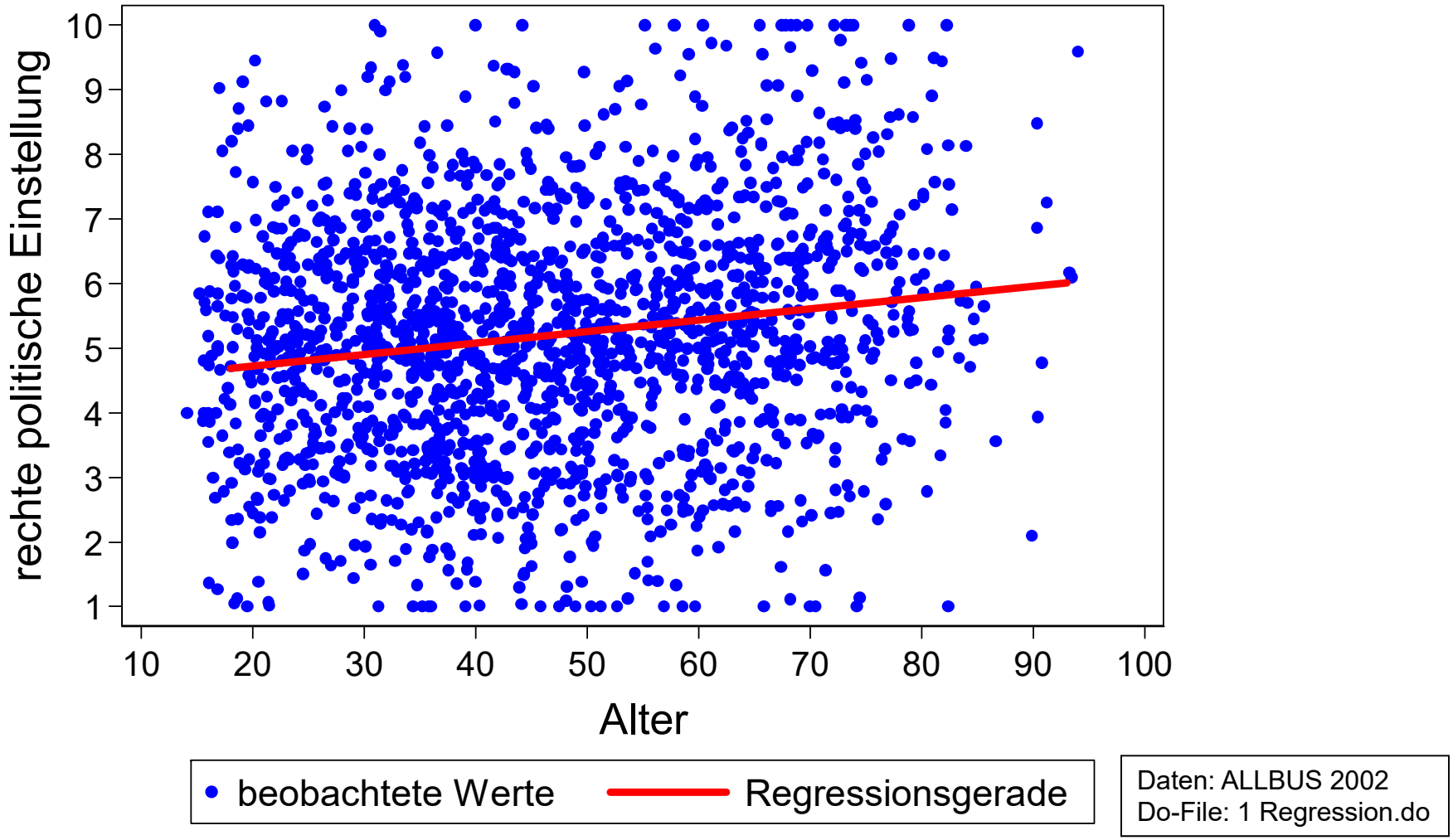
  - Beispiel „Abinote und Klausurerfolg“: r = -0,43
    - Steigt die Note um eine Standardabweichung, so verringert sich die Punktzahl um 0,43 Standardabweichungen

Stata-Bsp.: Politische Einstellung auf Alter



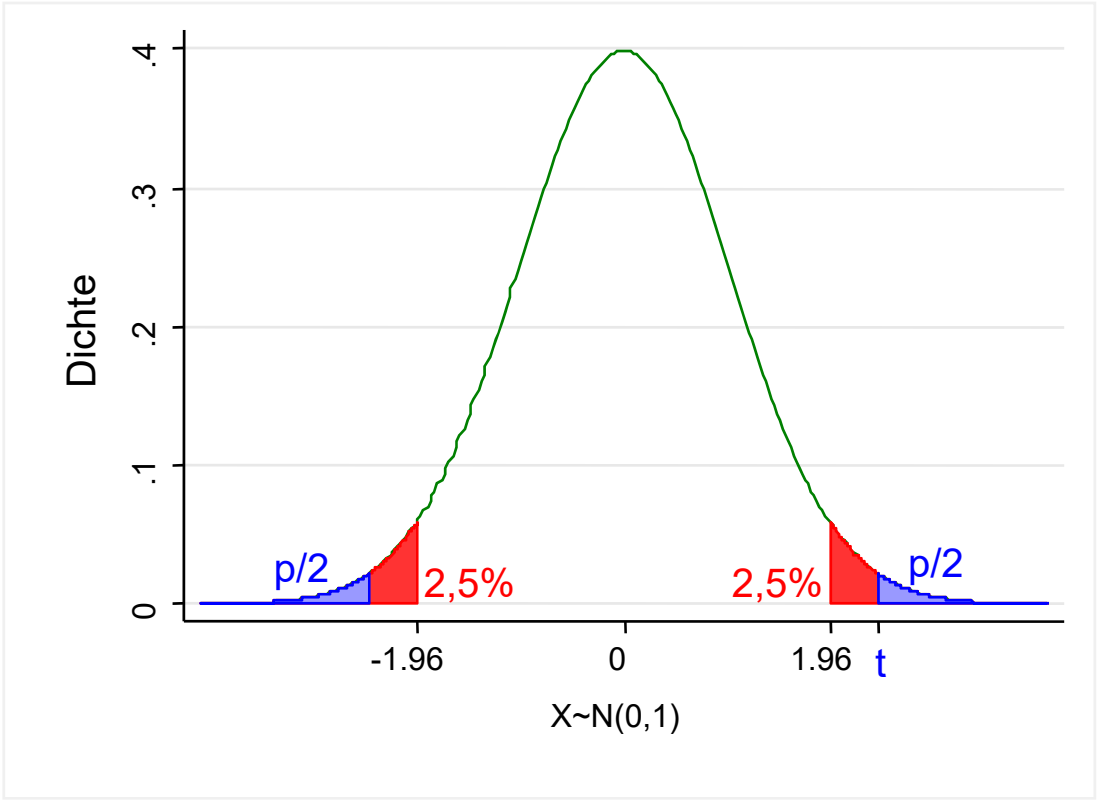
Daten: ALLBUS 2002
Do-File: 1 Regression.do

Regression: politische Einstellung auf Alter



Exkurs: p-Wert

- Der p-Wert gibt bei Gültigkeit der H_0 die Wahrscheinlichkeit an, dass die Teststatistik den berechneten Wert oder einen, der noch weiter in Richtung der Alternativhypothese liegt, annimmt
- Die Nullhypothese wird dann verworfen, wenn **$p \leq \alpha$**



H_0 : ß = 0

Prüfverteilung: Bei n>30 ist die t-Verteilung eine Standardnormal-verteilung

Die kritischen Werte sind auf dem 5%-Niveau -1,96 und 1,96

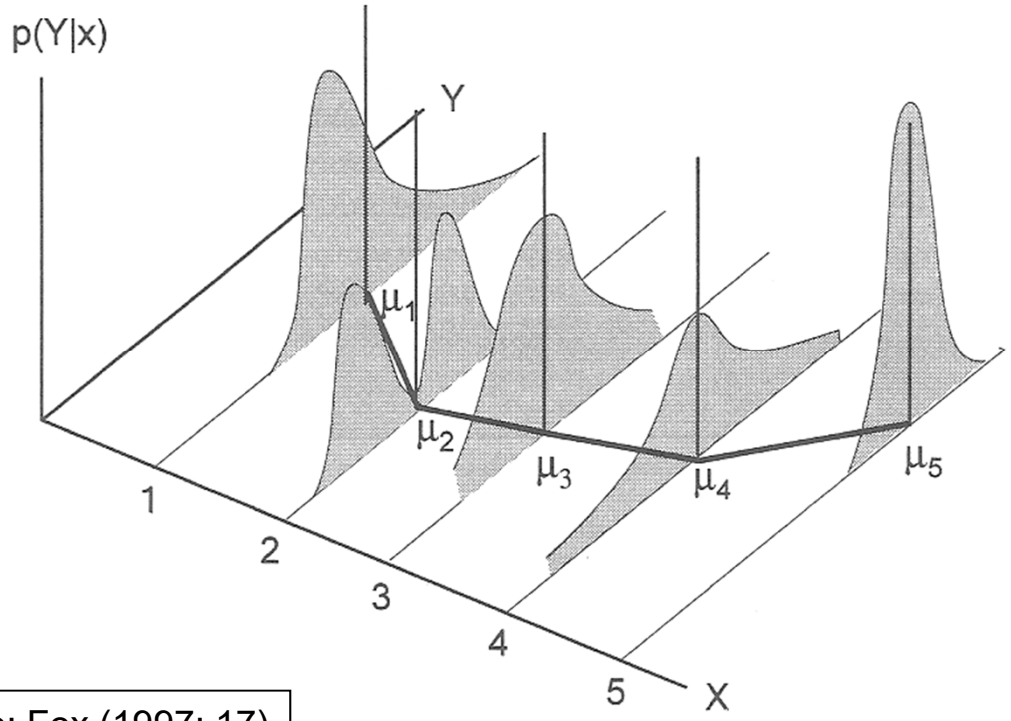
Liegt die Teststatistik **t** z.B. bei 2,4, so kann die H_0 abgelehnt werden

Rechts von **t** liegt p/2 (die Whs., dass noch was Extremeres rauskommt) (hier ist p=0,016)

Da offensichtlich p < 0,05, kann die H_0 abgelehnt werden

Exkurs: Regression als bedingte Verteilung

- Zwei Variablen Y und X
  - mit Realisierungen (y_i, x_i), für i=1,…,n
- Die Regression von Y auf X
  - ist die bedingte Verteilung: $f(Y \mid X=x)$



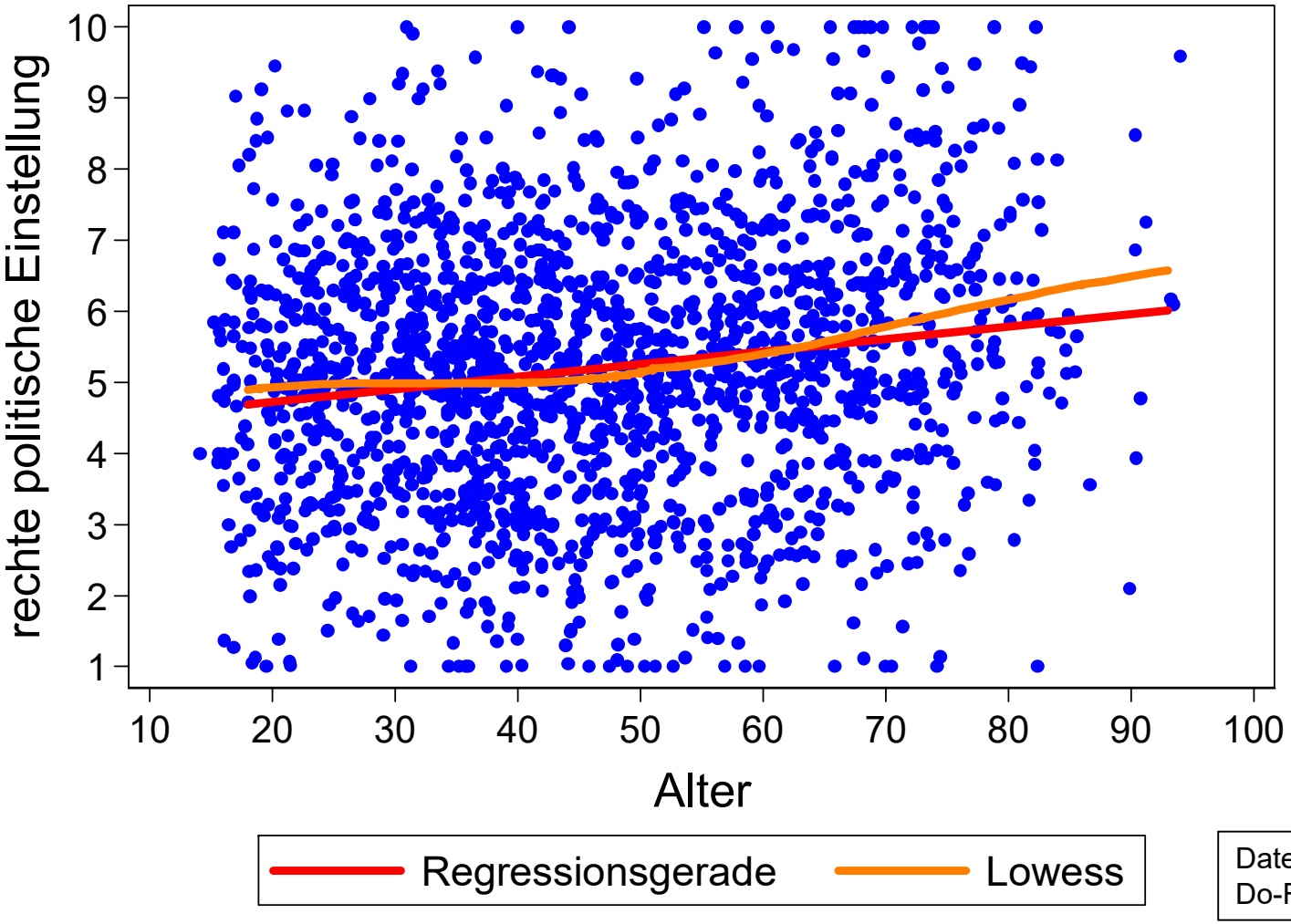
Quelle: Fox (1997: 17)

- X ist hier kategorial.
- Die bedingten Verteilungen haben sehr unterschiedliche Form.
- Eingetragen sind auch die bedingten Mittelwerte. Der Zusammenhang ist U-förmig.

Exkurs: Regression als bedingte Verteilung

- Eine solche allgemeine Regression enthält zu viel Information
- Informationsreduktion:
  Charakterisierung der Verteilung durch geeignete Kennzahlen
  - Y metrisch: bedingtes arithmetisches Mittel
  - Y metrisch, ordinal: bedingtes Quantil
  - Y nominal: bedingte Häufigkeiten (Kreuztabelle)
- Nicht-parametrische Regression
  - Benutze die Y-Werte in einer Umgebung von x zur Berechnung der Kennzahl (local averaging)
    - Lokale mean (median) Regression
    - Lowess Smoother
- Parametrische Regression
  - Weitere Informationsreduktion: man nimmt an, dass die bedingten Kennzahlen einer Funktion folgen
    - Lineare Mittelwertsregression (OLS Regression)
    - Lineare Medianregression
    - Quantilsregression

Nicht-parametrische und parametrische Regression



Daten: ALLBUS 2002
Do-File: 1 Regression.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 4:
Das multiple lineare Regressionsmodell

Das multiple lineare Regressionsmodell

Das Modell:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \qquad i = 1, \ldots, n$$

- ε_i ist ein Fehlerterm
- β_0 heißt Regressionskonstante
- Die anderen Regressionskoeffizienten definieren eine p-dimensionale Regressionsebene
- Interpretation: β_j gibt an,
  um wie viel Einheiten sich Y ändert,
  wenn sich X_j um eine Einheit erhöht,
  unter Kontrolle der anderen im Modell enthaltenen X-Variablen
  - Synonym: β_j sagt uns, welcher Effekt verbleibt, wenn wir die anderen X-Variablen konstant halten

OLS Schätzung

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

mit :

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Annahmen :

$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$ vier Annahmen über die Fehlerverteilung (normalverteilt, im Mittel 0, Homoskedastizität, keine Autokorrelation)

$Cov(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathbf{0}$ Exogenität von X

$rg(\mathbf{X}) = p + 1$ keine linearen Abhängigkeiten (bzw. Multikollinearität)

$$\text{OLS Schätzer : } \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Multiples R^2

- Die vorhergesagten Werte ergeben sich aus

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_p x_{ip}$$

- Multiples Bestimmtheitsmaß

$$R^2 = \frac{\text{erklärte Streuung}}{\text{gesamte Streuung}} = \frac{MSS}{TSS} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

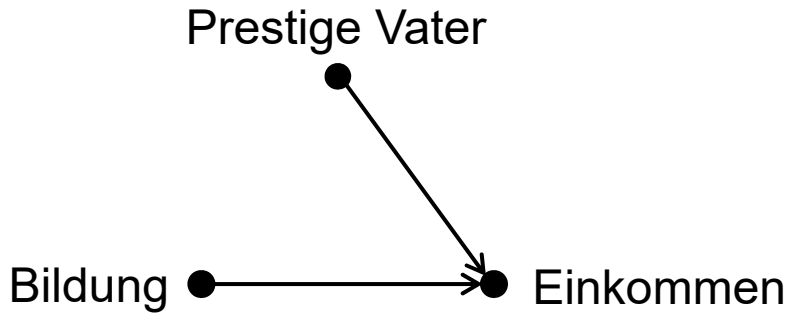
- Es besagt, welcher Anteil der Varianz von Y durch alle Regressoren zusammen erklärt wird
  - Fügt man einen weiteren Regressor hinzu, so ist das Bestimmtheitsmaß des erweiterten Modells mindestens genauso groß wie zuvor
  - Ist allerdings die Erklärungskraft der hinzugefügten Variable, gegeben die bereits im Modell enthaltenen Variablen, gering, so wird sich R^2 nur minimal erhöhen
  - Das Hinzufügen weiterer Variablen verbessert das Modell somit nur, wenn diese Variablen einen eigenständigen Erklärungsbeitrag leisten

Signifikanztests

- Test eines einzelnen Regressionskoeffizienten
  - Nullhypothese: X_j hat keinen Einfluss auf Y (kein Zusammenhang)
    $$H_0: \beta_j = 0$$
  - Die Teststatistik (t-Wert) ist
    $$T = \frac{\hat{\beta}_j}{\hat{\sigma}_j} \sim t(n-p-1)$$
  - Die H_0 wird abgelehnt, falls $|T| > t_{1-\alpha/2}(n-p-1)$
    - Ab n>30 das $z_{1-\alpha/2}$ Quantil (Faustregel für α=5%: |T| > 2)
- Test des gesamten Modells: overall F-Test
  - Nullhypothese: keine X-Variable hat einen Einfluss auf Y
    $$H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0$$
  - Die Teststatistik (F-Wert) ist
    $$F = \frac{\text{MSS}/p}{\text{RSS}/(n-p-1)} \sim F(p, n-p-1)$$
  - Die H_0 wird verworfen, falls: $F > F_{1-\alpha}(p,n-p-1)$

Beispiel: Statuszuweisungsmodell

- Blau/Duncan (1967) "The American Occupational Structure"
  - Wie erlangt man seine soziale Position?
    Durch „achievement“ oder Statusvererbung?
  - ALLBUS 2002:
    - Abhängige Variable: monatliches Netto-Einkommen in Euro
      (nur Westdeutsche, Vollzeit)
    - Status des Vaters: Magnitude-Prestigeskala (Werte von 20-187)
    - „Achievement“: eigene Schul- und Berufsbildung (Werte von 8-23,5)

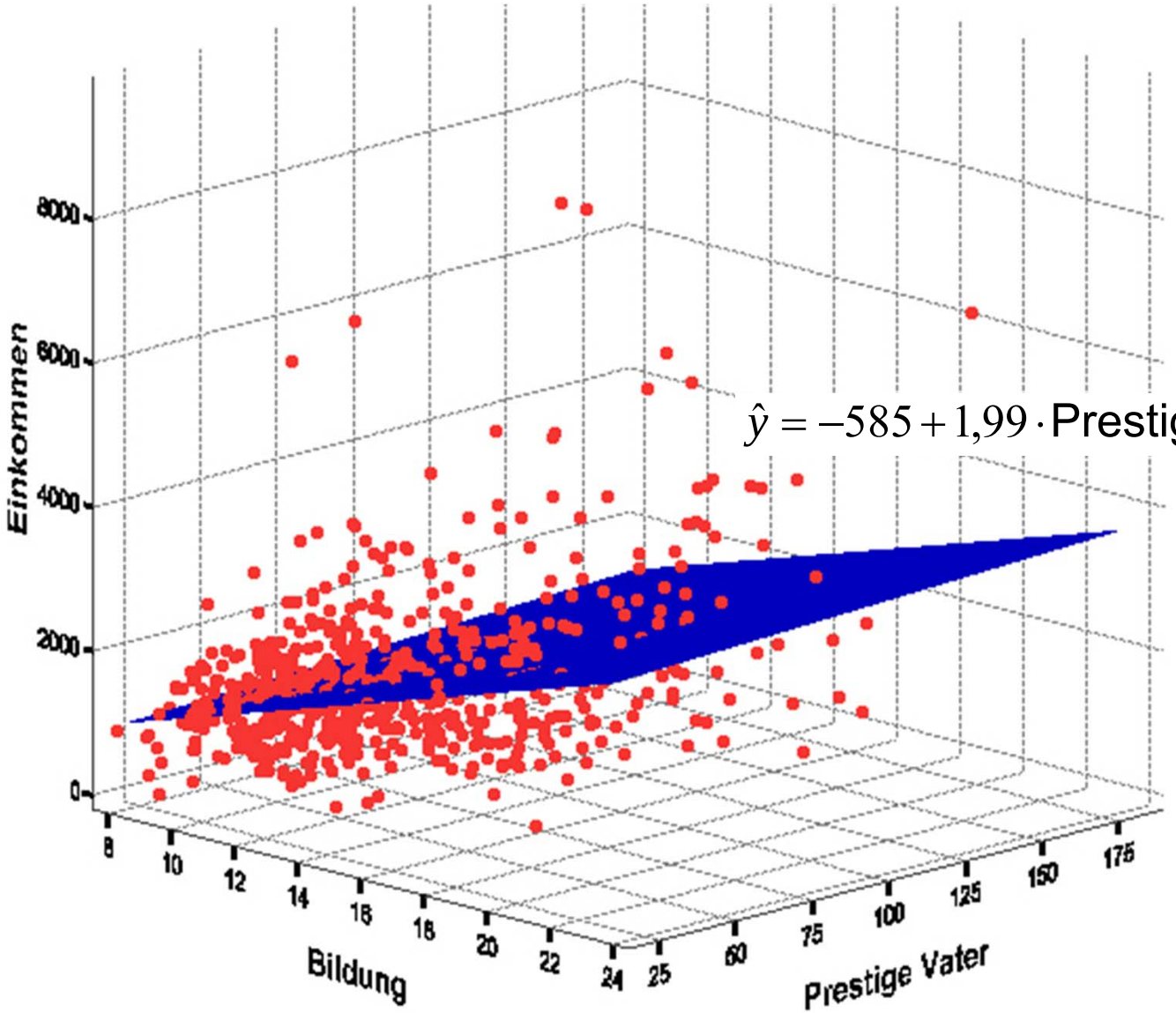


Beispiel: Stata-Output

```
. regress    eink prestv bild

      Source |       SS       df       MS              Number of obs =     670
-------------+------------------------------           F(  2,   667) =   64.99
       Model |   215244302     2   107622151           Prob > F      =  0.0000
    Residual |  1.1045e+09   667  1655904.67           R-squared     =  0.1631
-------------+------------------------------           Adj R-squared =  0.1606
       Total |  1.3197e+09   669  1972694.65           Root MSE      =  1286.8

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      prestv |    1.99162   2.004457     0.99   0.321    -1.944187    5.927426
        bild |   180.2345   17.73502    10.16   0.000     145.4113    215.0577
       _cons |  -585.4664   224.3659    -2.61   0.009    -1026.015   -144.9179
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 2 LinReg Modell.do

Beispiel: Regressionsebene



$$\hat{y} = -585 + 1{,}99 \cdot \text{Prestige} + 180 \cdot \text{Bildung}$$

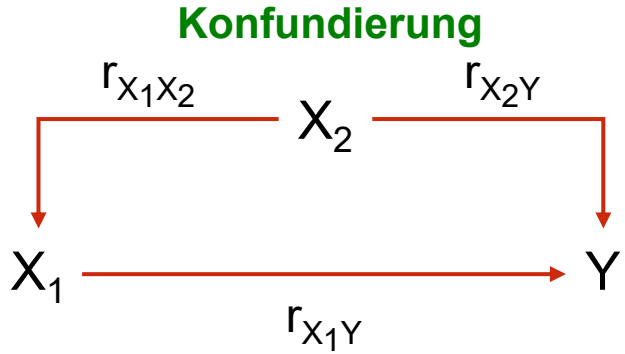
Was bedeutet „unter Kontrolle“?

- β_j ist der Effekt von X_j unter Kontrolle der anderen im Modell enthaltenen X-Variablen
  - Man sagt auch: „unter Konstanthaltung“ der anderen im Modell enthaltenen X-Variablen
- Der bivariate Effekt wird von Konfundierungen „bereinigt“

bivariater Effekt **Konfundierung**

$$\hat{\beta}_1^* = \frac{r_{X_1Y} - r_{X_1X_2} r_{X_2Y}}{1 - r_{X_1X_2}^2}$$

Standardisierung



- Spezialfall: X_1 und X_2 sind nicht korreliert (Multikausalität)

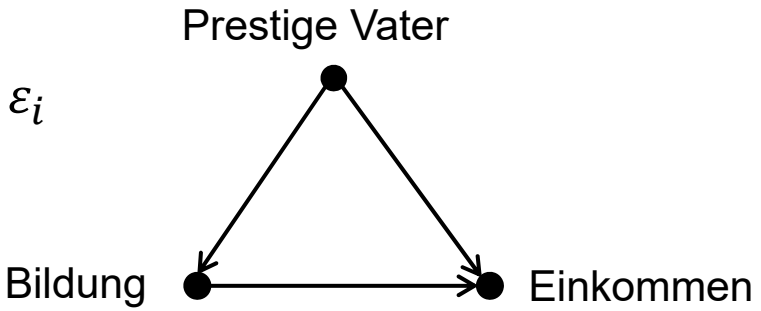
$$\hat{\beta}_1^* = r_{X_1Y}$$

  - Der multiple Regressionskoeffizient ist gleich dem bivariaten

Was bedeutet „unter Kontrolle“?

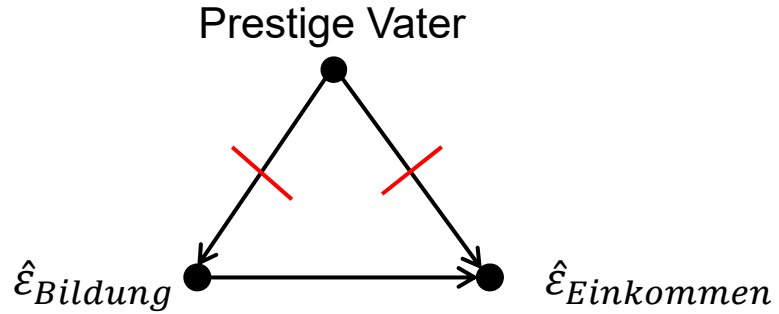
- Das multiple Modell mit einem Confounder (C)

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 c_i + \varepsilon_i$$



- Der Effekt des Confounders (C) wird „herauspartialisiert“
  - $y_i = \alpha + \beta c_i + \varepsilon_{iY}$ → $\widehat{\varepsilon_{iY}} = y_i - \hat{\alpha} - \hat{\beta} c_i$
    - Die auf C beruhende Variation ist herausgerechnet: $Cov(\widehat{\varepsilon_{iY}}, c_i) = 0$
  - $x_i = \gamma + \delta c_i + \varepsilon_{iX}$ → $\widehat{\varepsilon_{iX}} = x_i - \hat{\gamma} - \hat{\delta} c_i$
    - Die auf C beruhende Variation ist herausgerechnet: $Cov(\widehat{\varepsilon_{iX}}, c_i) = 0$
- Die Regression mit den Residuen liefert ebenfalls $\widehat{\beta_1}$

$$\widehat{\varepsilon_{iY}} = \beta_1 \widehat{\varepsilon_{iX}} + \varepsilon_i$$



Beispiel: Kontrolle der Herkunft

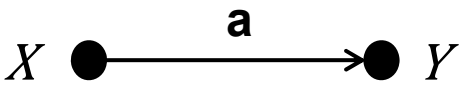
	(1) Bivariate Regression	(2) Multiple Regression	(3) Residuen Regression
bild	186.82***	180.23***	180.23***
	(16.45)	(17.74)	(17.72)
prestv		1.99	
		(2.00)	
_cons	-555.64*	-585.47**	0.00
	(222.35)	(224.37)	(49.68)
N	670	670	670
R-sq	0.162	0.163	0.134

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

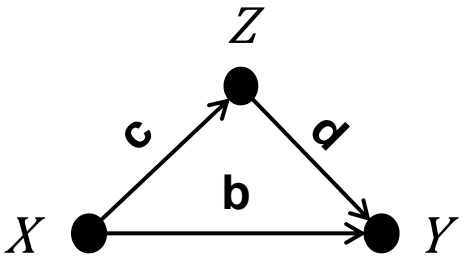
Daten: ALLBUS 2002
Do-File: 2 LinReg Modell.do

Regel I: Welche Variablen kontrollieren?

- Oft wird ein Standard-Set an „Kontrollvariablen“ verwendet
  - Geschlecht, West/Ost, Alter, Herkunft, Beruf, Bildung, …
  - Die „kontrolliere Alles auf Verdacht“ Strategie
  - Das ist keine sinnvolle Vorgehensweise!
- Regel: um den (totalen) Kausaleffekt zu identifizieren, kontrolliere nur (!) Confounder
- Wenn man Mediatoren mitkontrolliert: overcontrol-bias
  - Es wird nicht der totale, sondern nur der direkte Effekt geschätzt



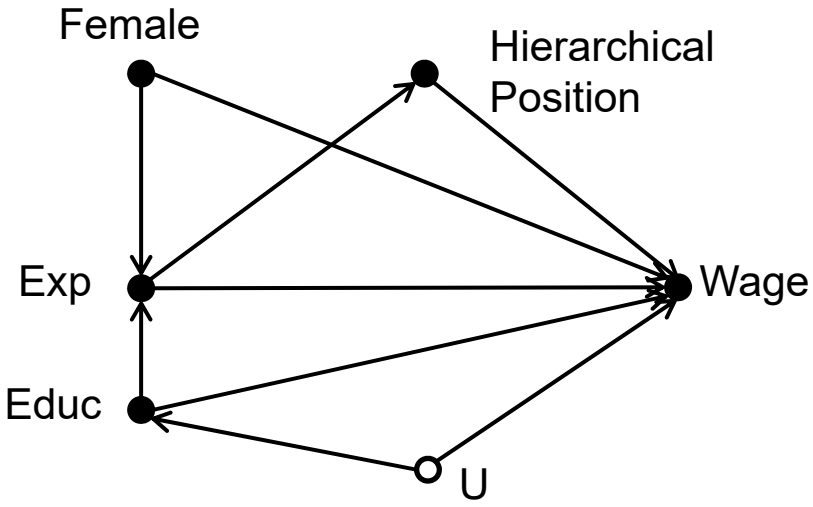
$$y_i = \beta_0 + \boldsymbol{a} x_i + \varepsilon_i$$



$$y_i = \beta_0 + \boldsymbol{b} x_i + d z_i + \varepsilon_i$$

Regel I: Welche Variablen kontrollieren?

- Ziel: Einen (totalen) Kausaleffekt zu identifizieren
  1. Ein theoretisches Modell (um den Kausaleffekt) entwickeln
  2. Daraus ein maßgeschneidertes Regressionsmodell ableiten
- Beispiel: Kausaleffekt der Berufserfahrung auf Lohn



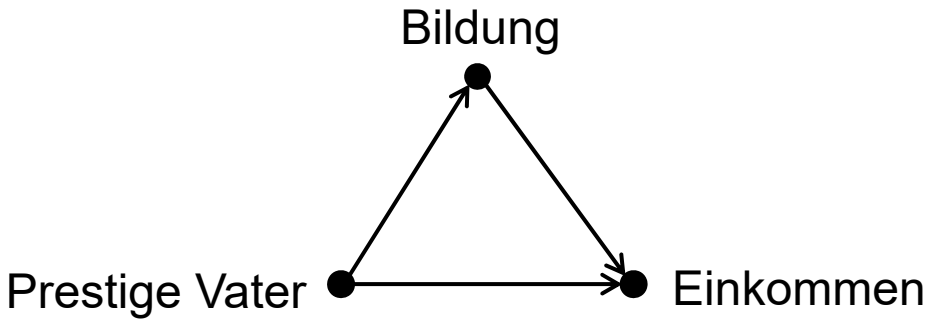
- Maßgeschneiderte Regression

$$\ln(\text{Wage}) = \alpha + \beta\text{Exp} + \gamma\text{Educ} + \delta\text{Female}$$

  - γ und δ sind keine (totalen) Kausaleffekte!

Regel II: Mediationsanalyse

- Wenn man einen Kausaleffekt gefunden hat, kann man im nächsten Schritt fragen: Was ist der Mechanismus?
  - Statistisch ist dies die Frage nach der Mediation: Welche Mediatorvariable(n) kann den Effekt erklären?
- Bsp. Statuszuweisung: Kann die Bildung den Effekt der Herkunft erklären?



- Vorgehen: man fügt der Regression die Mediatorvariable(n) hinzu
  - Reduziert sich der Kausaleffekt (signifikant)?

Beispiel: Statuszuweisungsmodell

	(1)	(2)
prestv	9.61***	1.99
	(2.00)	(2.00)
bild		180.23***
		(17.74)
_cons	1343.23***	-585.47**
	(128.51)	(224.37)
N	670	670
R-sq	0.034	0.163

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Der signifikante Kausaleffekt von „Prestige Vater“ wird unter Kontrolle des Mediators „Bildung“ deutlich kleiner und ist nicht mehr signifikant.
Oft wird in so einem Fall von „signifikanter Mediation“ gesprochen.
Das ist aber voreilig, denn dafür braucht es einen Signifikanztest für den indirekten Effekt. Diesen liefert der Sobel-Test.
In unserem Fall werden 79% des totalen Effektes signifikant mediiert.

Sobel-Goodman Mediation Tests

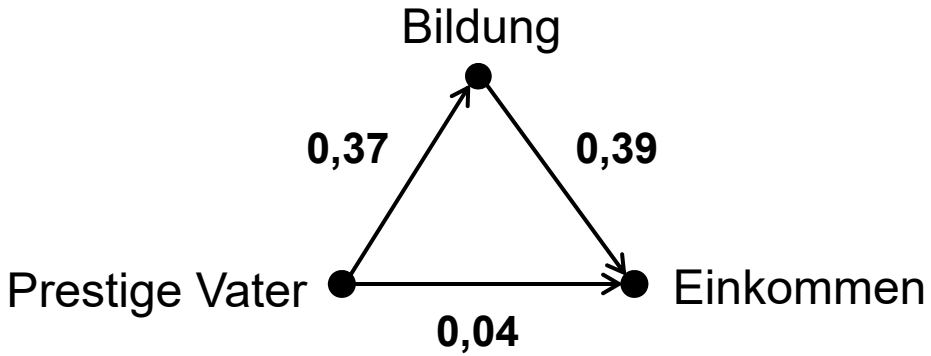
	Coef	Std Err	Z	P>|Z|
Indirect effect =	7.61593	1.04689	7.27483	3.5e-13
Direct effect =	1.99162	2.00446	.993595	.32042
Total effect =	9.60755	1.99636	4.81255	1.5e-06

Proportion of total effect that is mediated: .79270271

Daten: ALLBUS 2002
Do-File: 2 LinReg Modell.do

Beispiel: Statuszuweisung

- Das gesamte Kausalmodell mit den standardisierten Regressionskoeffizienten
  - Origin, Education, Destination: OED-Dreieck
  - Der direkte Herkunftseffekt ist schwach (und nicht signifikant)
  - Es gibt einen starken indirekten Effekt über Bildung

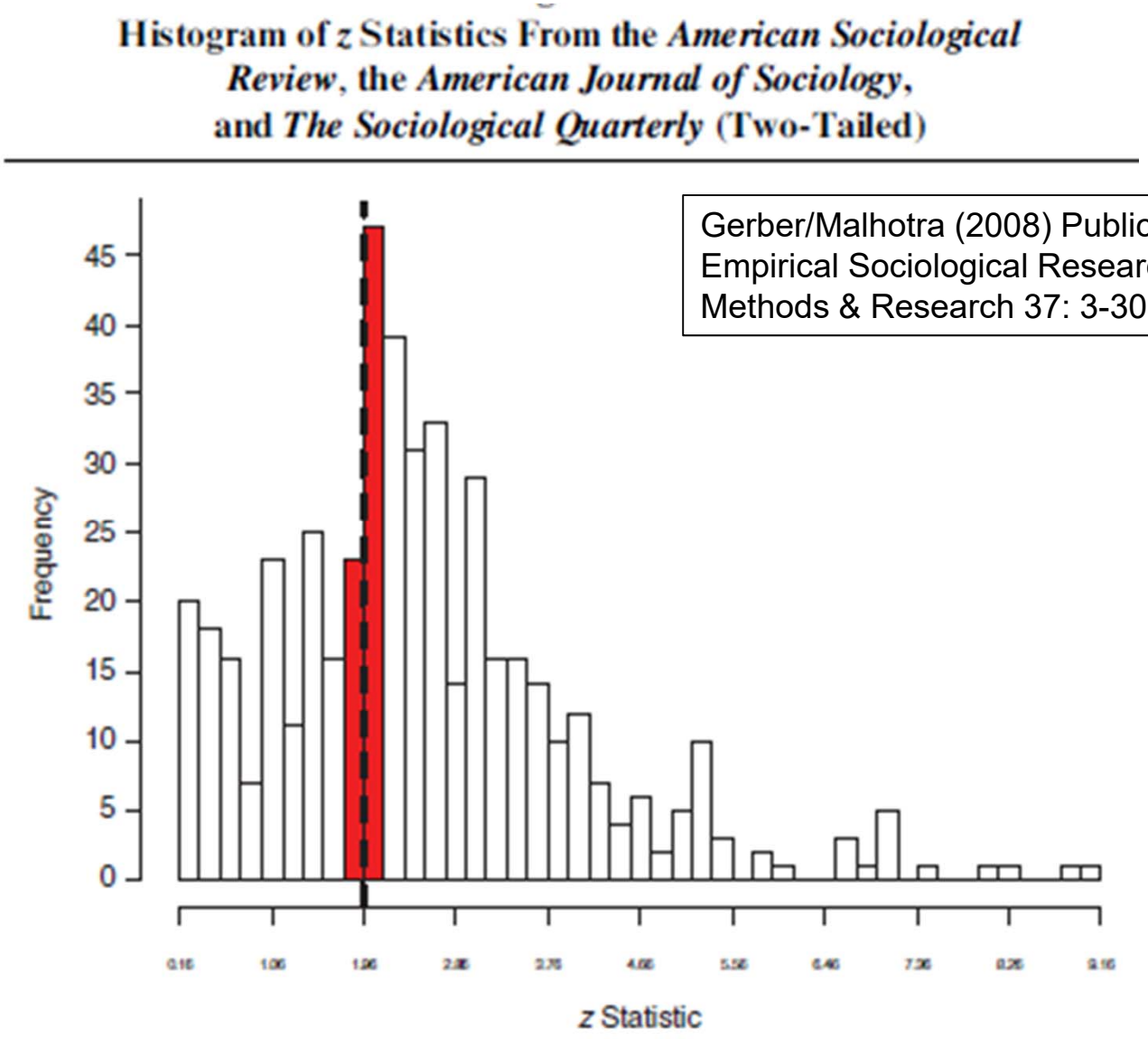


Daten: ALLBUS 2002
Do-File: 2 LinReg Modell.do

Das Signifikanztest-Ritual

- Unsinnige Anwendung der Signifikanztests in der Praxis
  - Multiples Testen und Publikations-Bias
    - Auch wenn in Wirklichkeit kein relevanter Effekt vorliegt, so werden von 100 Forschern irrtümlicherweise 5 einen „signifikanten“ Effekt finden und genau diese 5 „signifikanten“ Effekte werden publiziert
    - Analog: Variablen-/Modellselektion anhand von t-Tests
    - Nicht-signifikante Ergebnisse werden nicht publiziert (s. nächste Folie)
    - Folge: viele publizierte Ergebnisse sind zufällig zustande gekommen (also falsch, obwohl sie „signifikant“ sind)
  - Viele Forscher schauen nur noch auf die „Sternchen“
    - Aber: „Signifikanz ist nicht gleich Relevanz“
- Das Signifikanztestritual hat in den Sozialwissenschaften eine Menge an unsinnigen Ergebnissen produziert. Die Jagd nach Signifikanzen hat die Jagd nach der Realität verdrängt. Signifikanztests sollten deshalb abgeschafft werden!
  (Ziliak/McCloskey, 2008)

Publikations-Bias



Gerber/Malhotra (2008) Publication Bias in Empirical Sociological Research. Sociological Methods & Research 37: 3-30.

Regel III: Welches Signifikanzniveau?

- Statt Abschaffung, Verbesserung der Praxis (s.a. Krämer 2011)
  - Kein „Sternchenquetschen“: kein 10%-Signifikanzniveau
  - Achte mehr auf Effektstärke (und Effektrichtung!)
  - Keine Variablenselektion anhand von Signifikanztests
  - Auch nicht-signifikante Ergebnisse sind wichtig!
- Statt R^2-Zentrierung hin zur X-Zentrierung der Sozialforschung
  - Konzentriere dich auf einen Effekt und versuche den mittels harter Spezifikationstests zu widerlegen (Falsifikationismus!)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 5:
Interpretation von Regressionskoeffizienten

Interpretation von Regressionskoeffizienten

- Das multiple Regressionsmodell (ohne Personenindex i)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- Daraus ergibt sich der bedingte Erwartungswert

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- Verschiedene Interpretationsmöglichkeiten
  - Marginaleffekt (marginal effect, **ME**)

$$\frac{\partial E(y|\boldsymbol{x})}{\partial x_j} = \beta_j$$

  - Effekt der Veränderung von X_j um eine Einheit (discrete change, **DC**)

$$\frac{\Delta E(y|\boldsymbol{x})}{\Delta x_j} = E\left(y\middle|\boldsymbol{x}, x_j + 1\right) - E(y|\boldsymbol{x}, x_j) = \beta_j$$

  - Fazit: Im linearen Modell sind DC und ME identisch gleich dem Regressionskoeffizienten!

Graphische Präsentation von Regressionsergebnissen

- Anstatt einer Tabelle mit vielen Zahlen: Regressionsgraphen
  (Bauer, 2015)
- Im Prinzip drei Typen:
  1. Plotten der Marginaleffekte von X: (Effektplot) **Koeffizientenplot**
     - ME für metrische Variablen, DC für Dummies
  2. Plotten der vorhergesagten Werte von Y: **Profile-Plot**
     - Welcher Wert von Y ergibt sich für verschiedene X-Werte?
     - Für alle Fälle in den Daten werden die beobachteten Werte eingesetzt, $\widehat{y_i}$ berechnet und dann gemittelt (predictive margins)
  3. Plotten der Effekte von X gegeben Z: **konditionaler Effektplot**
     - Wie verändert sich der ME von X mit Z?
     - Hilfreich für Interaktionen

Präsentation der Regressionskoeffizienten

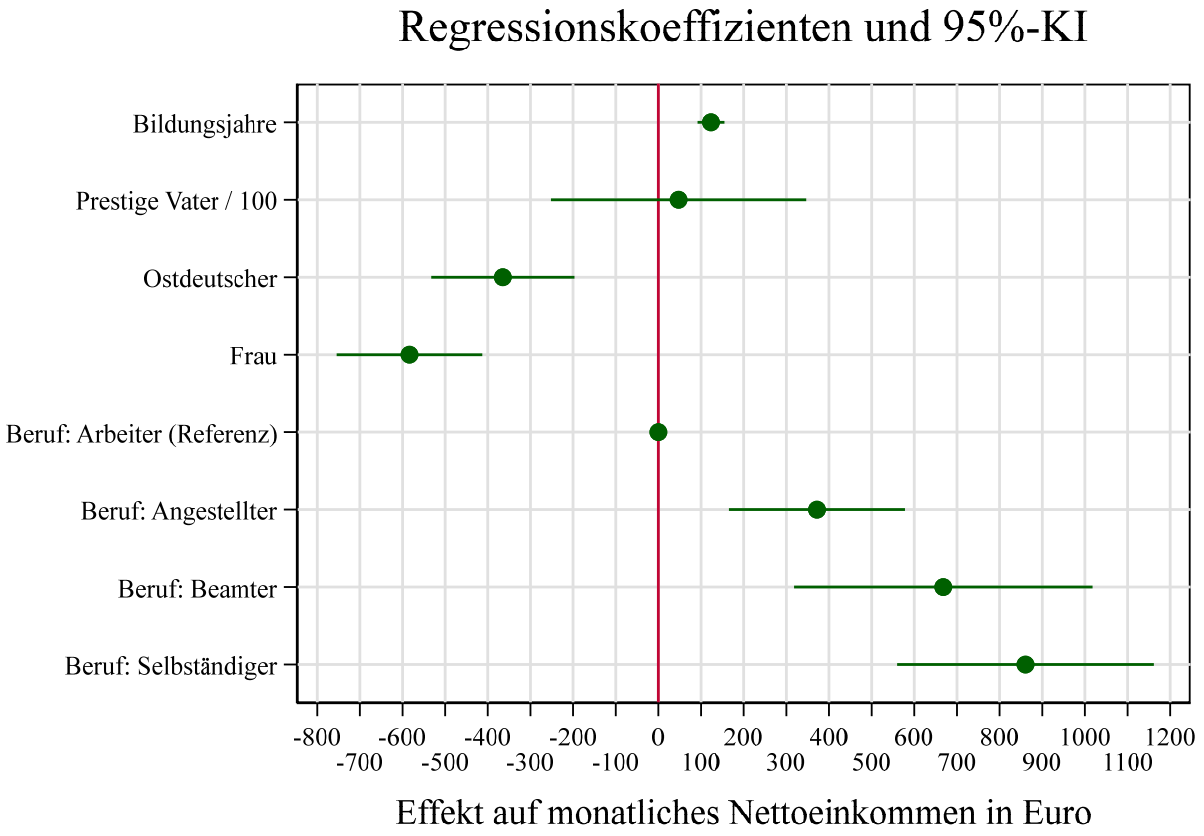
- Bsp. Einkommensregression
  - Monatliches Nettoeinkommen in Euro (nur Vollzeiterwerbstätige)
  - Bildungsjahre, Prestige Vater/100, Ostdeutscher, Frau, berufliche Stellung

```
regress eink bild prestv ost frau i.beruf
esttab using "RegTabelle.rtf", r2 b(%6.1f)
```

```
coefplot, drop(_cons) xline(0) base
```

Bildungsjahre	123.3***
	(7.67)
Prestige Vater / 100	47.4
	(0.31)
Ostdeutscher	-364.7***
	(-4.26)
Frau	-583.5***
	(-6.71)
Ber.: Arbeiter (Ref.)	
Ber.: Angestellter	371.8***
	(3.54)
Ber.: Beamter	667.9***
	(3.75)
Ber.: Selbständiger	860.7***
	(5.62)
Konstante	163.2
N	948
R^2	0.214

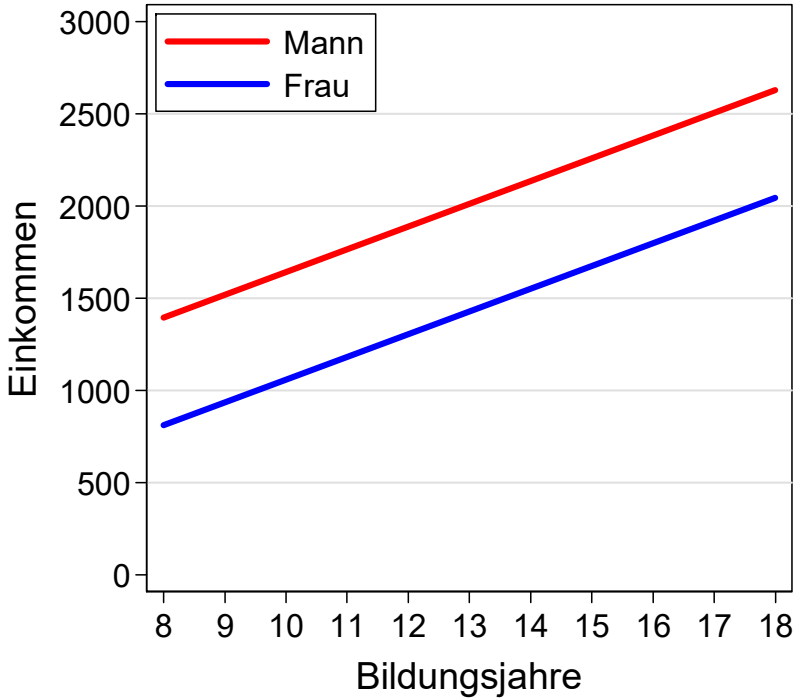
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



Daten: ALLBUS 2002
Do-File: 3 LinReg Interpretation.do

Profile-Plot

- Hilfreich zur Veranschaulichung von Regressionskoeffizienten
  - Man plottet die vorhergesagten Werte der Outcome-Variable
    - Z.B. die geschätzte Regressionsgerade für eine metrische Variable
    - Evtl. für verschiedene inhaltlich interessierende Gruppen
  - Predictive Margins
    - Für jede Beobachtung wird mit ihren Kovariatenwerten ein Vorhersagewert berechnet
    - Nur die „marginvars“ werden auf fixierte Werte gesetzt
    - Anschließend wird über alle Vorhersagewerte gemittelt
    - Dies ist eine Art „kontrafaktisches“ Vorgehen!



Veranschaulichung des Bildungs- und des Geschlechtseffektes

„frau“ und „bild“ sind hier die marginvars

```
margins frau, at(bild=(8 18))
marginsplot,   noci
```

$$\hat{y}_M = 409 + 123 \times \text{Bild}$$
$$\hat{y}_F = 409 + 123 \times \text{Bild} - 584$$

Daten: ALLBUS 2002
Do-File: 3 LinReg Interpretation.do

Interpretation einer Polynomregression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \text{Exp} + \beta_3 \text{Exp}^2 + \epsilon$$

Daten: ALLBUS 2002
Do-File: 3 LinReg Interpretation.do

$$\frac{\partial E(y)}{\partial \text{Exp}} = \beta_2 + 2 \times \beta_3 \text{Exp}, \qquad \text{Exp}_{\text{max/min}} = -\frac{\beta_2}{2 \times \beta_3}$$

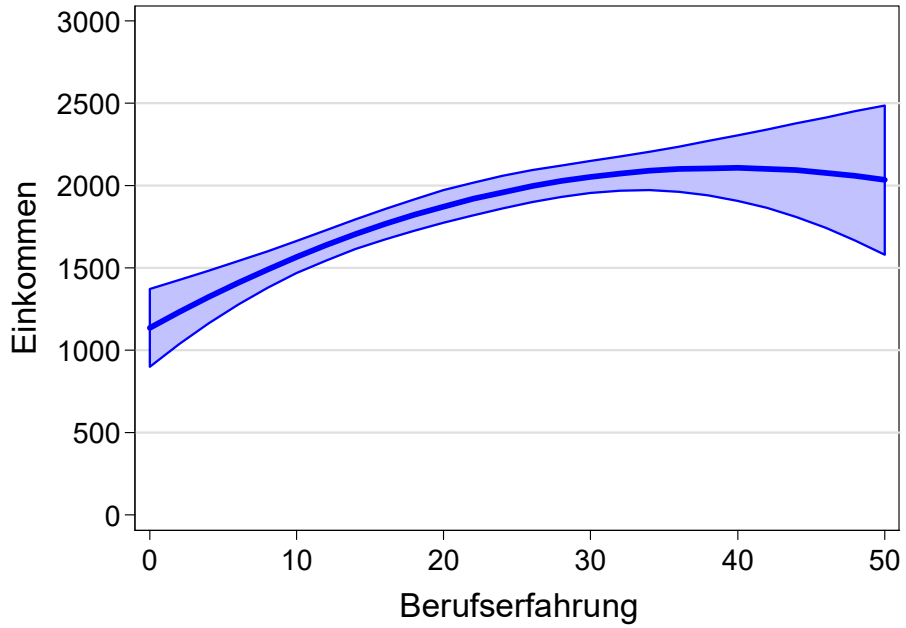
```
. regress   eink bild ost frau c.exp c.exp#c.exp

      Source |       SS       df       MS              Number of obs =    1118
-------------+------------------------------           F(  5,  1112) =   65.83
       Model |   474909648     5  94981929.7           Prob > F      =  0.0000
    Residual |  1.6045e+09  1112  1442907.24           R-squared     =  0.2284
-------------+------------------------------           Adj R-squared =  0.2249
       Total |  2.0794e+09  1117  1861613.69           Root MSE      =  1201.2

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |      180.95      12.36    14.64   0.000       156.69      205.20
         ost |    -448.92       77.23    -5.81   0.000      -600.45     -297.39
        frau |    -435.94       75.63    -5.76   0.000      -584.33     -287.55
         exp |      49.53       12.57     3.94   0.000        24.88       74.19
 c.exp#c.exp |      -0.63        0.29    -2.19   0.029        -1.20       -0.07
       _cons |    -971.85     202.76    -4.79   0.000     -1369.68     -574.01
------------------------------------------------------------------------------
```

Interpretation einer Polynomregression

```
. margins, dydx(exp) at(exp=(0 10 20 30 40 50))
------------------------------------------------
    |            Delta-method
    |      dy/dx   Std. Err.      z    P>|z|
----+-------------------------------------------
 0  |      49.53      12.57     3.94   0.000
10  |      36.91       7.18     5.14   0.000
20  |      24.28       3.38     7.18   0.000
30  |      11.66       6.13     1.90   0.057
40  |      -0.97      11.41    -0.08   0.932
50  |     -13.59      17.00    -0.80   0.424
------------------------------------------------
```

```
. test exp exp#exp

( 1)  exp = 0
( 2)  c.exp#c.exp = 0

 F(  2,  1112) =   26.21
      Prob > F =    0.0000
```

$$\text{Exp}_{\text{max}} = -\frac{49{,}53}{2 \times -0{,}63} = 39{,}3$$



Profile-Plot

```
margins, at(exp=(0(2)50))
marginsplot
```

Daten: ALLBUS 2002
Do-File: 3 LinReg Interpretation.do

Das semi-logarithmische Regressionsmodell

$$\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$\leftrightarrow y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon)$$

$$\frac{\partial E(y)}{\partial x_1} = E(y)\beta_1$$

Marginaleffekt

$$\frac{\Delta E(y)}{\Delta x_1} = E(y)\left(e^{\beta_1} - 1\right) \quad \Rightarrow \quad \frac{\frac{\Delta E(y)}{\Delta x_1}}{\mathrm{E(y)}} = e^{\beta_1} - 1$$

Discrete Change prozentuale Veränderung

$(e^{\beta_1} - 1) \times 100$ ist die prozentuale Veränderung von Y bei Erhöhung von X um eine Einheit.

$\beta_1 \times 100$ ist eine gute Näherung, falls $|\beta_1| < 0{,}1$.

Das semi-logarithmische Regressionsmodell

```
. regress  lneink bild ost frau c.exp c.exp#c.exp

      Source |       SS       df       MS              Number of obs =    1118
-------------+------------------------------           F(  5,  1112) =  145.12
       Model |  138.195766     5  27.6391532           Prob > F      =  0.0000
    Residual |   211.79019  1112  .190458804           R-squared     =  0.3949
-------------+------------------------------           Adj R-squared =  0.3921
       Total |  349.985956  1117  .313326729           Root MSE      =  .43642

------------------------------------------------------------------------------
      lneink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |      0.0863     0.0045    19.21   0.000       0.0775      0.0951
         ost |     -0.2496     0.0281    -8.90   0.000      -0.3047     -0.1946
        frau |     -0.2317     0.0275    -8.43   0.000      -0.2856     -0.1777
         exp |      0.0431     0.0046     9.43   0.000       0.0341      0.0520
 c.exp#c.exp |     -0.0006     0.0001    -6.13   0.000      -0.0008     -0.0004
       _cons |      5.8124     0.0737    78.90   0.000       5.6678      5.9569
------------------------------------------------------------------------------
```

Die exakten „Discrete Change“ %-Effekte

- Bildungsrendite: +9,1%
- Ostdeutscher: -22,1%
- Frau: -20,7%

Daten: ALLBUS 2002
Do-File: 3 LinReg Interpretation.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 6:
Regression mit Dummies

Regression mit Dummy

Einkommen in Abhängigkeit von Bildung und Geschlecht
Dummy-Kodierung: 0 = Mann, 1 = Frau

```
. regress eink bild frau

      Source |       SS       df       MS              Number of obs =    1118
-------------+------------------------------           F(  2,  1115) =  116.00
       Model |   358146421     2   179073211           Prob > F      =  0.0000
    Residual |  1.7213e+09  1115  1543745.36           R-squared     =  0.1722
-------------+------------------------------           Adj R-squared =  0.1707
       Total |  2.0794e+09  1117  1861613.69           Root MSE      =  1242.5

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |   169.2551   12.44852    13.60   0.000     144.8299    193.6803
        frau |  -511.2477   77.71202    -6.58   0.000    -663.7259   -358.7694
       _cons |  -264.8927   172.6304    -1.53   0.125    -603.6097    73.82433
------------------------------------------------------------------------------
```

Interpretation: Bei gleicher Bildung verdienen Frauen im Schnitt 511 Euro weniger als Männer.

Daten: ALLBUS 2002
Do-File: 4 LinReg Dummies.do

Kategoriale uV mit mehr als 2 Ausprägungen

- Auch kategoriale Variablen sind in der multiplen Regression möglich
  - Aber erst nach Dummy-Bildung!
  - Grundprinzip: Für jede Ausprägung wird eine Dummy gebildet
  - In die Regression werden alle Dummies außer einer (Referenzkategorie) aufgenommen
- Beispiel: berufliche Stellung

berufliche Stellung	D1	D2	D3	D4
Arbeiter	1	0	0	0
Angestellter	0	1	0	0
Beamter	0	0	1	0
Selbständiger	0	0	0	1

  - Werden keine weiteren unabhängigen Variablen berücksichtigt, so entspricht die Konstante β_0 dem Mittelwert der Referenzkategorie
  - Die Koeffizienten β_j der Dummies geben den Mittelwertunterschied der betreffenden Kategorie zur Referenzkategorie an

Generierung der Dummies

```
. tabulate beruf, gen(d)

       beruf |      Freq.     Percent        Cum.
-------------+-----------------------------------
    Arbeiter |        360       29.22       29.22
 Angestellter |        622       50.49       79.71
     Beamter |         90        7.31       87.01
Selbständiger |        160       12.99      100.00
-------------+-----------------------------------
       Total |      1,232      100.00

. tabulate beruf d1
               |    beruf==Arbeiter
       beruf |         0          1 |     Total
-------------+----------------------+----------
    Arbeiter |         0        360 |       360
 Angestellter |       622          0 |       622
     Beamter |        90          0 |        90
Selbständiger |       160          0 |       160
-------------+----------------------+----------
       Total |       872        360 |     1,232
```

Daten: ALLBUS 2002
Do-File: 4 LinReg Dummies.do

Interpretation der Dummies

```
. table beruf, contents(sum d1 sum d2 sum d3 sum d4)

------------------------------------------------------------
        beruf |    sum(d1)     sum(d2)     sum(d3)     sum(d4)
--------------+---------------------------------------------
     Arbeiter |         360           0           0           0
 Angestellter |           0         622           0           0
      Beamter |           0           0          90           0
Selbständiger |           0           0           0         160
------------------------------------------------------------
```

```
. table beruf, contents(mean eink)

--------------------------
        beruf | mean(eink)
--------------+-----------
     Arbeiter |   1332.902
 Angestellter |   1894.345
      Beamter |   2480.987
Selbständiger |   2714.033
--------------------------
```

```
. regr eink d2 d3 d4

----------------------------
        eink |      Coef.
-------------+--------------
          d2 |   561.4422
          d3 |   1148.084
          d4 |   1381.131
       _cons |   1332.903
----------------------------
```

Daten: ALLBUS 2002
Do-File: 4 LinReg Dummies.do

Regression mit kategorialer uV

```
. regress eink bild i.beruf

      Source |       SS       df       MS              Number of obs =    1072
-------------+------------------------------           F(  4,  1067) =   52.89
       Model |   330638913     4  82659728.2            Prob > F      =  0.0000
    Residual |  1.6674e+09  1067  1562726.39            R-squared     =  0.1655
-------------+------------------------------           Adj R-squared =  0.1624
       Total |  1.9981e+09  1071  1865609.69            Root MSE      =  1250.1

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------------------------------
        bild |   130.5443   14.63054     8.92   0.000
             |
       beruf |
          2  |   216.6888   95.98289     2.26   0.024
          3  |   562.1304   171.8194     3.27   0.001
          4  |   919.5047   143.5743     6.40   0.000
             |
       _cons |  -141.8791   179.5353    -0.79   0.430
------------------------------------------------------------------------------
```

Bivariate Effekte

bild	165
angest	561
beamt	1148
selbst	1381

`i.beruf` sagt Stata, dass „beruf“ eine Indikatorvariable ist. Stata bildet dann die „virtuellen“ Dummies „2.beruf“, „3.beruf“ und „4.beruf“. Referenzkategorie ist automatisch der kleinste Wert 1=Arbeiter.

Daten: ALLBUS 2002
Do-File: 4 LinReg Dummies.do

Regression mit kategorialer uV

```
. testparm i.beruf

 ( 1)  2.beruf = 0
 ( 2)  3.beruf = 0
 ( 3)  4.beruf = 0

       F(  3,  1067) =   15.01
            Prob > F =    0.0000
```

Signifikanz des Berufs?

```
. regress eink bild ib3.beruf
```

Beamte als Referenzkategorie

eink	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
bild	130.5443	14.63054	8.92	0.000	101.8364	159.2522
beruf						
1	-562.1304	171.8194	-3.27	0.001	-899.2727	-224.9881
2	-345.4416	154.3745	-2.24	0.025	-648.3538	-42.52949
4	357.3743	183.0629	1.95	0.051	-1.829857	716.5784
_cons	420.2513	271.3571	1.55	0.122	-112.2028	952.7055

Daten: ALLBUS 2002
Do-File: 4 LinReg Dummies.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 7:
Interaktionseffekte

Berücksichtigung von Interaktionseffekten

- Der Effekt von X hängt vom Wert von Z ab
  - Z nennt man auch Moderator-Variable (Interaktion = Moderation)
- Berücksichtigung in einer Regression
  - Nimm die Hauptterme in die Regression (X und Z)
  - Füge einen Interaktionsterm hinzu
    - Das Produkt von X und Z (deshalb auch „Produktterm“)
    - Hierbei unterstellt man eine multiplikative Interaktion

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 (x * z) + \epsilon$$

Der (konditionale) Marginaleffekt von X

$$\frac{\partial E(y)}{\partial x} = \beta_1 + \beta_3 z$$

$$z = 0: \frac{\partial E(y)}{\partial x} = \beta_1$$

$$z = 1: \frac{\partial E(y)}{\partial x} = \beta_1 + \beta_3$$

Der (konditionale) Marginaleffekt von Z

$$\frac{\partial E(y)}{\partial z} = \beta_2 + \beta_3 x$$

$$x = 0: \frac{\partial E(y)}{\partial z} = \beta_2$$

$$x = 1: \frac{\partial E(y)}{\partial z} = \beta_2 + \beta_3$$

Der Interaktionseffekt

$$\frac{\partial^2 E(y)}{\partial x \, \partial z} = \beta_3$$

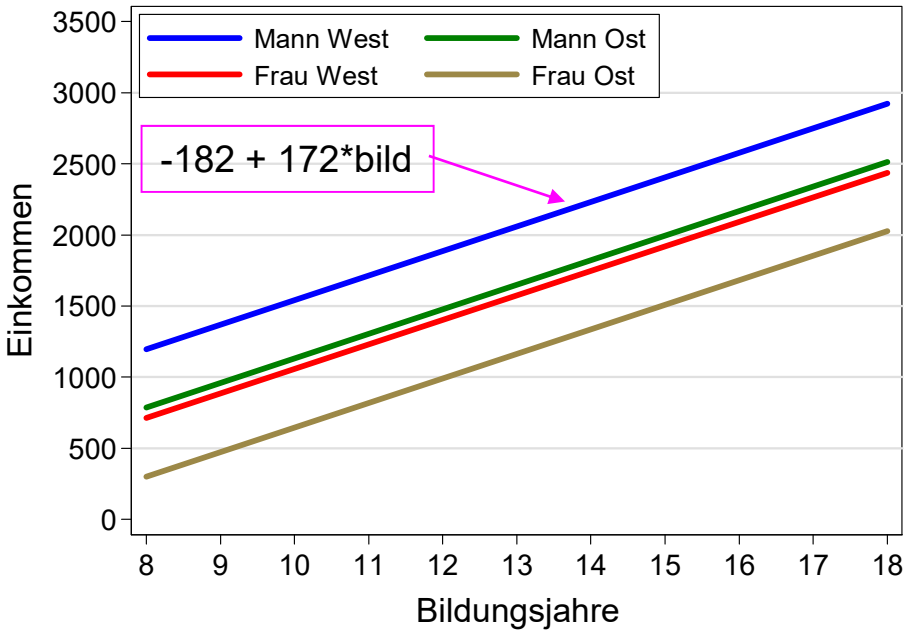
Dummy-Interaktion: Geschlecht/Wohnort

```
* Ohne Interaktion
. regress eink bild frau ost
------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |   172.4277   12.31945    14.00   0.000     148.2557    196.5996
        frau |  -484.5974   76.98204    -6.29   0.000    -635.6436   -333.5513
         ost |  -410.0585   78.53404    -5.22   0.000    -564.1498   -255.9672
       _cons |  -182.3318   171.3636    -1.06   0.288    -518.5635       153.9
------------------------------------------------------------------------------

. * Mit Interaktion
. regress eink bild i.frau i.ost frau#ost
                                                Number of obs =    1118
                                                R-squared     =  0.1947

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |   171.3628    12.3170    13.91   0.000     147.1957    195.5299
        frau |  -592.1200    95.0501    -6.23   0.000    -778.6176   -405.6225
         ost |  -526.8585    99.1837    -5.31   0.000    -721.4665   -332.2504
    frau#ost |   311.5265   161.9030     1.92   0.055      -6.1431    629.1960
       _cons |  -132.5139   173.1033    -0.77   0.444    -472.1594    207.1317
------------------------------------------------------------------------------
```

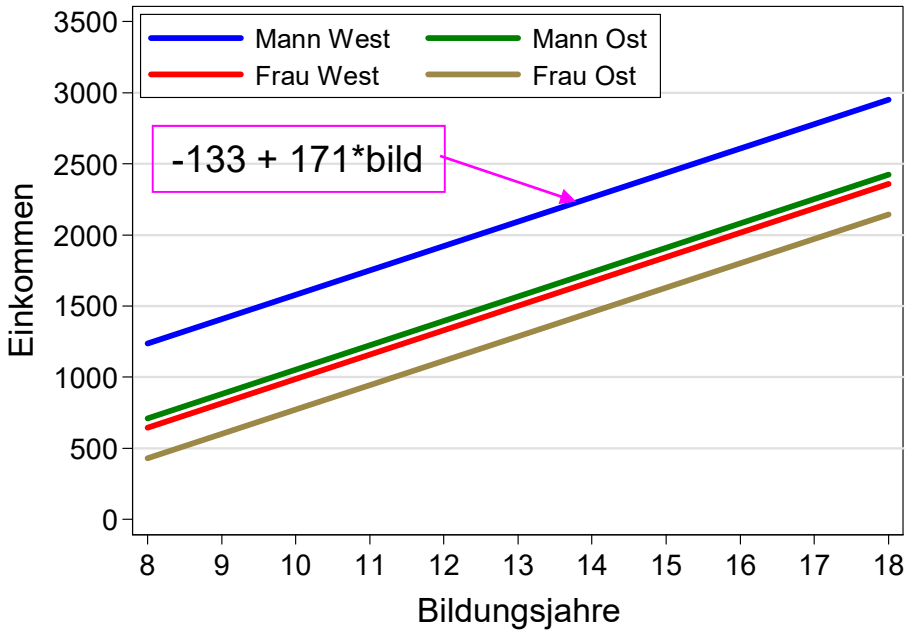
Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Dummy-Interaktion: Geschlecht/Wohnort

ohne Interaktion



mit Interaktion



Designmatrix	Frau -592	Ost -527	Ofrau 312	Einkommens- unterschied
Mann West	0	0	0	0
Mann Ost	0	1	0	-527
Frau West	1	0	0	-592
Frau Ost	1	1	1	-807

Referenzgruppe

Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Slope-Interaktion: Geschlecht/Bildung

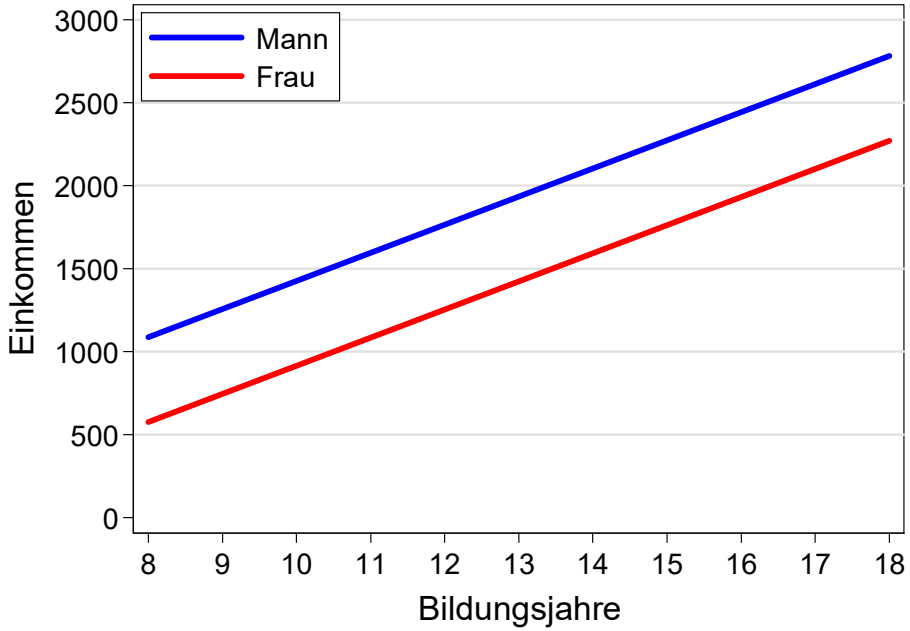
```
. * Ohne Interaktion
. regress eink c.bild i.frau

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |   169.2551   12.44852    13.60   0.000     144.8299    193.6803
        frau |  -511.2477   77.71202    -6.58   0.000    -663.7259   -358.7694
       _cons |  -264.8927   172.6304    -1.53   0.125    -603.6097    73.82433
------------------------------------------------------------------------------

. * Mit Interaktion
. regress eink c.bild i.frau   i.frau#c.bild

                                                Number of obs =    1118
                                                R-squared     =  0.1765

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |   190.2989   15.17511    12.54   0.000     160.5239    220.0739
        frau |   335.2419   359.1157     0.93   0.351    -369.3776    1039.861
 frau#c.bild |  -63.77517   26.41776    -2.41   0.016    -115.6094   -11.94099
       _cons |  -546.0541   207.9355    -2.63   0.009    -954.0433   -138.0648
------------------------------------------------------------------------------
```

Frauen verdienen mehr als Männer!?

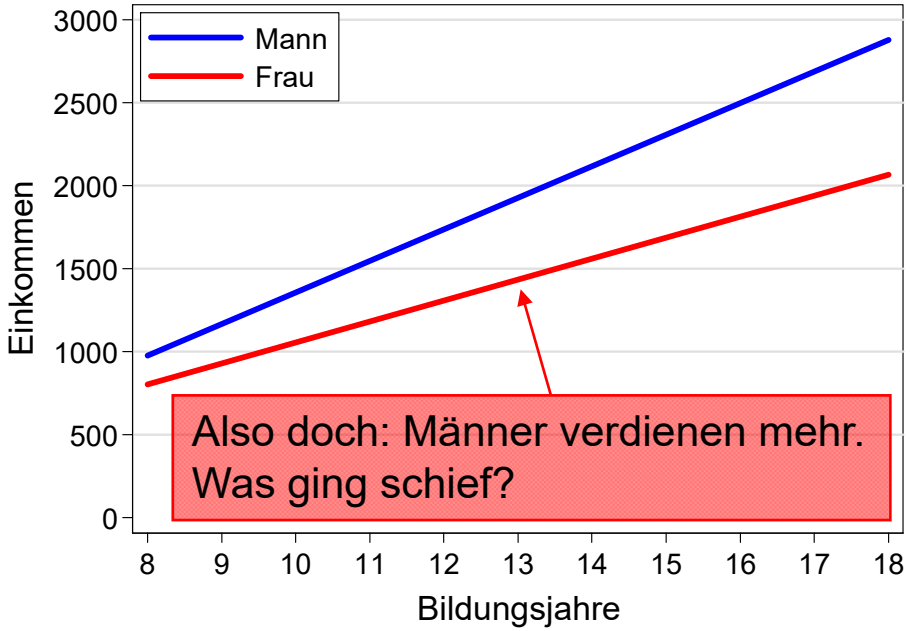
Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Slope-Interaktion: Geschlecht/Bildung

ohne Interaktion



mit Interaktion



Designmatrix	Konstante -546	Frau 335	Bild 190	Fbild -64	Regressions- gerade
Mann	1	0	1	0	-546 + 190•Bild
Frau	1	1	1	1	-211 + 126•Bild

Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Slope-Interaktion: Zentrierung

- Warum wurden wir in die Irre geführt?
  - Der Effekt von „Frau“ ist bei Bild=0 zu interpretieren
  - Dies ist offensichtlich eine sinnlose Interpretation
  - Problem tritt immer auf, wenn die metrische Interaktionsvariable keinen sinnvollen 0-Wert hat (z.B. auch bei Alter)
  - Abhilfe: Zentrieren der metrischen Variable (cbild = bild – mean(bild))

```
. regress eink c.cbild i.frau   i.frau#c.cbild

      Source |       SS       df       MS              Number of obs =    1118
-------------+------------------------------           F(  3,  1114) =   79.61
       Model |   367104408     3   122368136           Prob > F      =  0.0000
    Residual |  1.7123e+09  1114  1537089.85           R-squared     =  0.1765
-------------+------------------------------           Adj R-squared =  0.1743
       Total |  2.0794e+09  1117  1861613.69           Root MSE      =  1239.8

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cbild |   190.2989   15.17511    12.54   0.000     160.5239    220.0739
        frau |  -513.8864   77.55202    -6.63   0.000    -666.0509   -361.7219
frau#c.cbild |  -63.77517   26.41776    -2.41   0.016    -115.6094   -11.94099
       _cons |   1987.661   46.14574    43.07   0.000     1897.119    2078.204
------------------------------------------------------------------------------
```

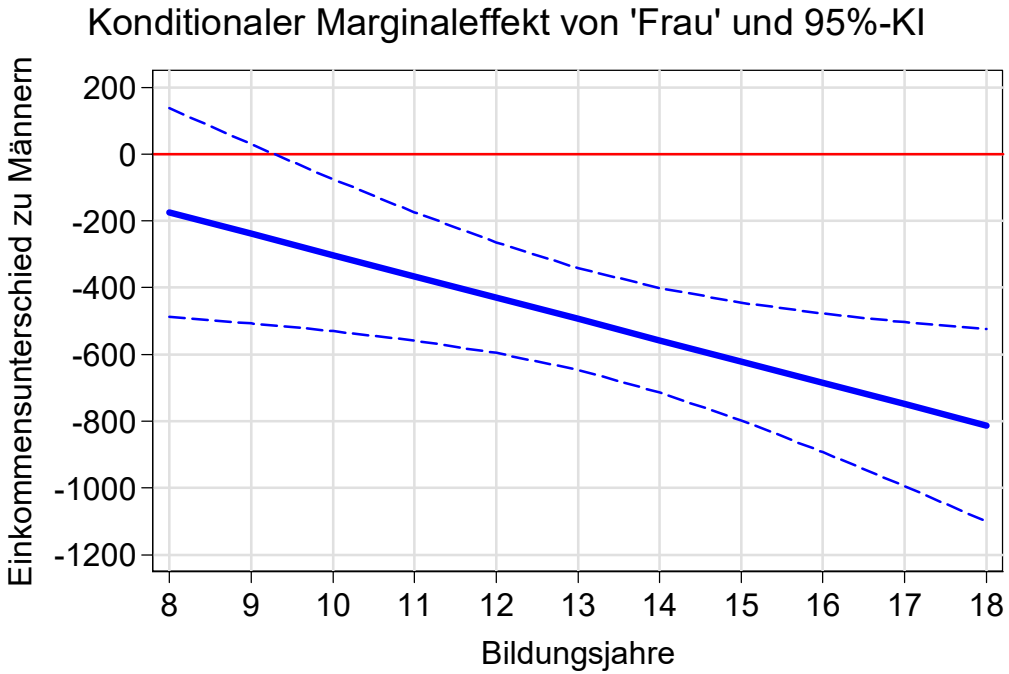
Berichte auch die konditionalen Marginaleffekte I

- Wir wissen nun, dass sich die Bildungsrenditen signifikant unterscheiden
- Oft wollen wir aber auch die Bildungsrendite in den beiden Gruppen wissen
  - Dies ist die Frage nach den konditionalen Marginaleffekten
  - `margins frau, dydx(cbild)`
- Man erhält sie auch über eine alternative Parametrisierung (**nested effects**)
- Im Beispiel: Bildungsrendite für Männer und Frauen getrennt
  - „Bild“ aus dem Modell nehmen, ersetzen durch:
    „Bild_F“: Bildung für Frauen, 0 sonst (zentriert)
    „Bild_M“: Bildung für Männer, 0 sonst (zentriert)

```
. regress eink  cbild_m  cbild_f  frau
      Source |       SS       df       MS              Number of obs =    1118
-------------+------------------------------           F(  3,  1114) =   79.61
       Model |   367104408     3   122368136           Prob > F      =  0.0000
    Residual |  1.7123e+09  1114  1537089.85           R-squared     =  0.1765
-------------+------------------------------           Adj R-squared =  0.1743
       Total |  2.0794e+09  1117  1861613.69           Root MSE      =  1239.8

------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cbild_m |   190.2989   15.17511    12.54   0.000     160.5239    220.0739
     cbild_f |   126.5237   21.62439     5.85   0.000     84.09456    168.9528
        frau |  -513.8864   77.55202    -6.63   0.000    -666.0509   -361.7219
       _cons |   1987.661   46.14574    43.07   0.000     1897.119    2078.204
------------------------------------------------------------------------------
```

Berichte auch die konditionalen Marginaleffekte II

- Symmetrie: Bildung moderiert den Effekt von „Frau“
  - Konditionale Marginaleffekte „Frau“: 335 – 64 * Bild
    - Ab Bild=6 ist der Effekt negativ. Dann geht die „Schere“ weiter auf
    - Achtung: der Interaktionseffekt (-64) ist zwar signifikant, das sagt uns aber nichts über die Signifikanz des Fraueneffekts!
    - Deshalb: ab welchem Bildungsniveau verdienen Frauen signifikant weniger als Männer?



Conditional-Effects-Plot

In Stata 12 leicht zu produzieren

```
margins, at(bild=(8(1)18)) dydx(frau)
marginsplot
```

Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Vollständige Interaktion mit Ost

- Will man Interaktionen aller Variablen zulassen
  - Interaktionsterme für alle Variablen (+ Hauptterm)
    - Test der Signifikanz der Interaktion für jede Variable einzeln
    - Signifikanztest für alle Interaktionseffekte (+Haupteffekt) (entspricht Signifikanztest für getrennte Modelle, Chow-Test)

```
. regress eink   i.ost##(i.frau c.cbild)
------------------------------------------------------------------------------
        eink |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------------------------------
       cbild |   181.2513   14.78765    12.26   0.000
        frau |  -588.6129   95.07484    -6.19   0.000
         ost |  -524.2035    99.1876    -5.28   0.000
   ost#cbild |  -32.26021   26.70952    -1.21   0.227
    ost#frau |   313.2981   161.8763     1.94   0.053
       _cons |   2148.898   54.73564    39.26   0.000
------------------------------------------------------------------------------
```

```
. * CHOW TEST
. contrast ost ost#i.frau ost#c.cbild, overall

------------------------------------------------
             |         df           F        P>F
-------------+----------------------------------
     Overall |          3       10.83     0.0000
------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Regeln für den Umgang mit Interaktionen

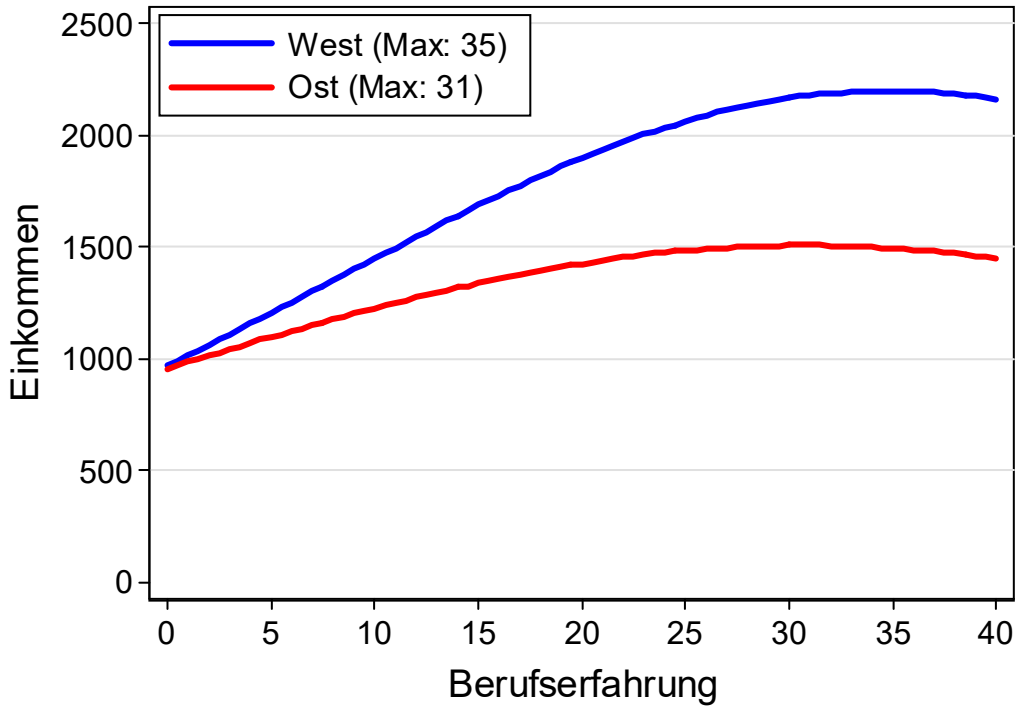
- Beide Hauptterme müssen im Modell sein
  - Ansonsten kaum realistische Restriktionen und schwer interpretierbar
  - Kollinearität zwischen Haupttermen und Interaktionstermen ist ein Datenproblem, kein Spezifikationsproblem! Bei hoher Kollinearität braucht man halt mehr Daten. Es macht aber keinen Sinn, den Hauptterm zu eliminieren.
- Zentriere metrische Interaktionsvariablen
  - Regressionskoeffizienten sind keine durchschnittlichen Marginaleffekte, sondern die Marginaleffekte an der Stelle Z = 0 (bzw. X = 0)
  - Bzw. wenn man Z zentriert hat: an der Stelle Z = mean(Z)
    - Zentrieren macht die Interpretation einfacher!
- Berichte inhaltlich bedeutsame Marginaleffekte (plus KI)
  - Z kategorial: berichte die Marginaleffekte von X in den Kategorien von Z
  - Z metrisch: plotte den Marginaleffekt von X gegen Z (Conditional-Effects-Plot)

Schließlich: Ein Humankapitalmodell getrennt für West/Ost

AV: logarithmiertes Einkommen

	West	Ost
bild	0.089***	0.082***
	(16.50)	(10.15)
exp	0.047***	0.029***
	(8.86)	(3.35)
exp^2	-0.001***	-0.000*
	(-5.56)	(-2.38)
frau	-0.246***	-0.186***
	(-7.25)	(-3.96)
_cons	5.723***	5.793***
	(64.68)	(43.08)
N	752	366
R^2	0.424	0.281

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



N.B.: Die Tabelle wurde direkt aus Stata mit dem Befehl „esttab“ erzeugt.

Daten: ALLBUS 2002
Do-File: 5 LinReg Interaktion.do

Ein komplexes Modell

Daten: ALLBUS 2002
Do-File: 3a LinReg Regressionsplots.do

```
. regress eink bild frau i.ost##(i.beruf c.exp##c.exp)
-----------------+------------------------------------------------------------
            bild |   142.4947    14.5792     9.77   0.000     113.8873    171.1021
            frau |  -470.5558    80.4638    -5.85   0.000    -628.4426   -312.6690
          1.ost |   458.5089   330.7956     1.39   0.166    -190.5811   1107.5989
                |
           beruf |
   Angestellter |   431.1194   112.9495     3.82   0.000     209.4889    652.7499
        Beamter |   387.5722   192.8944     2.01   0.045       9.0731    766.0714
  Selbständiger |  1138.4152   169.4858     6.72   0.000     805.8487   1470.9817
                |
             exp |    57.3432    15.9833     3.59   0.000      25.9807     88.7057
                |
     c.exp#c.exp |    -0.8235     0.3554    -2.32   0.021      -1.5209     -0.1261
                |
       ost#beruf |
 1#Angestellter |  -264.1148   176.2388    -1.50   0.134    -609.9322     81.7026
      1#Beamter |   846.6789   345.1562     2.45   0.014     169.4104   1523.9474
1#Selbständiger |  -599.7550   263.7621    -2.27   0.023    -1.12e+03    -82.1988
                |
       ost#c.exp |   -65.5124    30.2477    -2.17   0.031    -124.8648     -6.1601
                |
 ost#c.exp#c.exp |     1.1592     0.6706     1.73   0.084      -0.1567      2.4751
-----------------+------------------------------------------------------------
```

Die Ergebnisse dieses Modells sind in Tabellenform praktisch nicht zu interpretieren

Systematik der Regressionsplots

*** I) Koeffizientenplots**

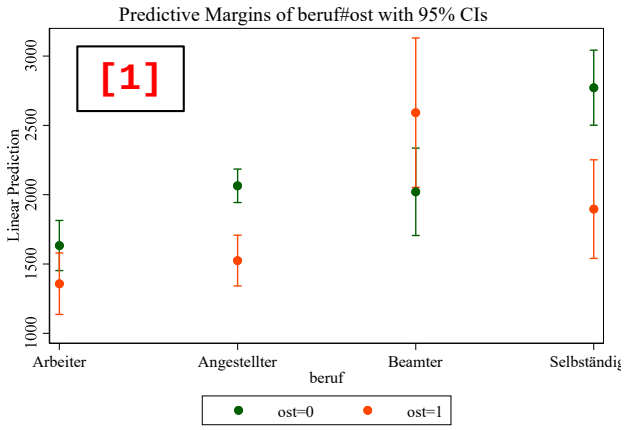
```
coefplot, drop(_cons) xline(0) base //hier nicht sehr hilfreich

margins, dydx(*)   // Plot der "Average Marginal Effects" (AME)
marginsplot, horizontal xline(0) plotopts(connect(i))  // auch
nicht sehr informativ
```

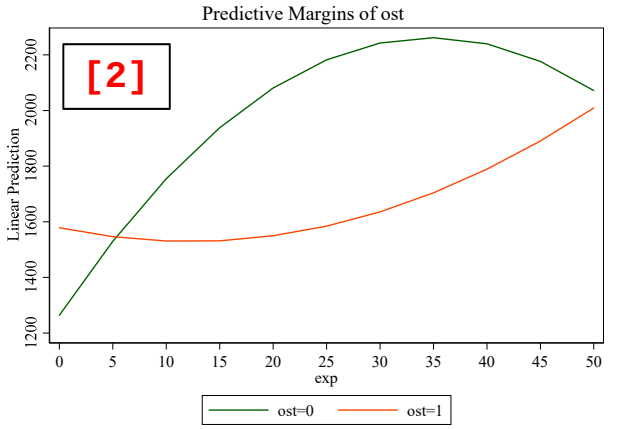
*** II) Profile Plot (PP)**

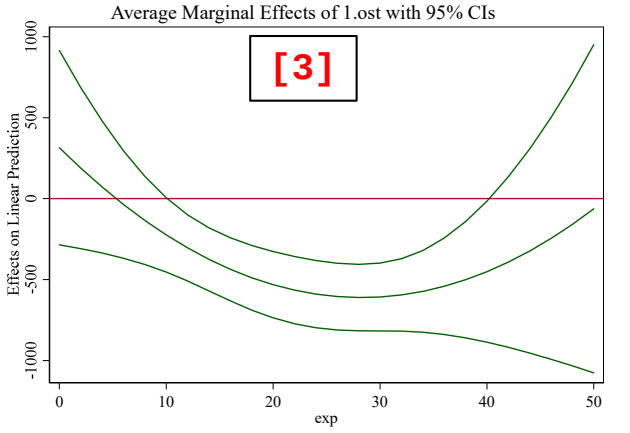
```
margins beruf                    //PP mit kategorialer Variable
marginsplot, plotopts(connect(i))

margins, at(exp=(0(5)50))        //PP mit metrischer Variable
marginsplot, recast(line) recastci(rarea)

margins beruf#ost                //Interaktion kategor. Variablen
marginsplot, plotopts(connect(i))  [1]

margins ost, at(exp=(0(5)50))  //Interaktion metr./kateg. Var.
marginsplot, noci recast(line)     [2]
```

*** III) Conditional-Effects Plot (CEP)**

```
margins ost, dydx(exp)                //AMEs Exp nach Ost (CEP I)
marginsplot, horizontal xline(0) plotopts(connect(i))

margins, dydx(ost) at(exp=(0(2)50)) //AMEs Ost nach Exp (CEPII)
marginsplot, recast(line) recastci(rline) yline(0)    [3]
```







LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 8:
Regressionsdiagnostik

- Linearität
- Homoskedastizität
- Normalverteilung
- Ausreißer

Regressionsdiagnostik

- Die Schätzung der Regressionskoeffizienten und die Tests auf ihre Signifikanz sind von Annahmen abhängig
- Deshalb sollte auch immer überprüft werden, ob diese Annahmen gerechtfertigt sind. Im Folgenden:
  - Multikollinearität
  - Linearitätsannahme A1/A2
  - Homoskedastizitäts-Annahme A3
  - Normalverteilungsannahme A6
  - zusätzlich: Ausreißerdiagnostik
- Meist analysiert man dazu die Residuen (Residuenanalyse)
  - Die Residuen sind Schätzer für die Fehlerterme

$$\hat{\varepsilon}_i = \hat{\varepsilon}(y_i|x_i) = y_i - \hat{y}_i$$

Multikollinearität

- Perfekte Kollinearität: lineare Abhängigkeit unter den Variablen
  - Modell nicht schätzbar (ab $r \geq 0{,}99$ wird es kritisch)
  - Stata lässt automatisch Variablen weg, um Kollinearität zu beheben
- „Mäßige“ Multikollinearität ($r < 0{,}99$)
  - OLS schätzbar und konsistent
  - Aber S.E.s der betroffenen Variablen größer (Schätzung unpräzise)
    - Das sehen viele Forscher als Problem („Sternchenjagd“)
- Diagnose: variance-inflation-factor (VIF)

$$\hat{V}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{(n-1){s_{x_j}}^2} \frac{1}{1-{R_j}^2} \quad \rightarrow \quad VIF = \frac{1}{1-{R_j}^2}$$

  - Die S.E.s sind „inflationiert“ mit Faktor $\sqrt{VIF}$
  - Faustregel: problematisch falls $VIF > 30$
- Maßnahme
  - Spezifikation überprüfen (aber nicht: einfach Variable weglassen)
  - Mehr Daten erheben, damit die Schätzung präziser wird
  - Index bilden (betroffene Variablen messen dasselbe?)

Multikollinearität

Interaktionsterme korrelieren stark

```
regress eink   c.bild##frau
               c.exp##c.exp

. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
        bild |      1.54    0.649513
      1.frau |     21.48    0.046559
 frau#c.bild |     21.83    0.045815
         exp |     14.55    0.068728
 c.exp#c.exp |     14.72    0.067929
-------------+----------------------
    Mean VIF |     14.82
```

Mit zentrierten Variablen ist das Problem geringer!

```
regress eink   c.bild_cen##frau
               c.exp##c.exp

. estat vif

    Variable   |       VIF       1/VIF
---------------+----------------------
    bild_cen   |      1.54    0.649513
      1.frau   |      1.01    0.991694
frau#c.bild_cen|      1.49    0.669306
         exp   |     14.55    0.068728
 c.exp#c.exp   |     14.72    0.067929
---------------+----------------------
    Mean VIF   |      6.66
```

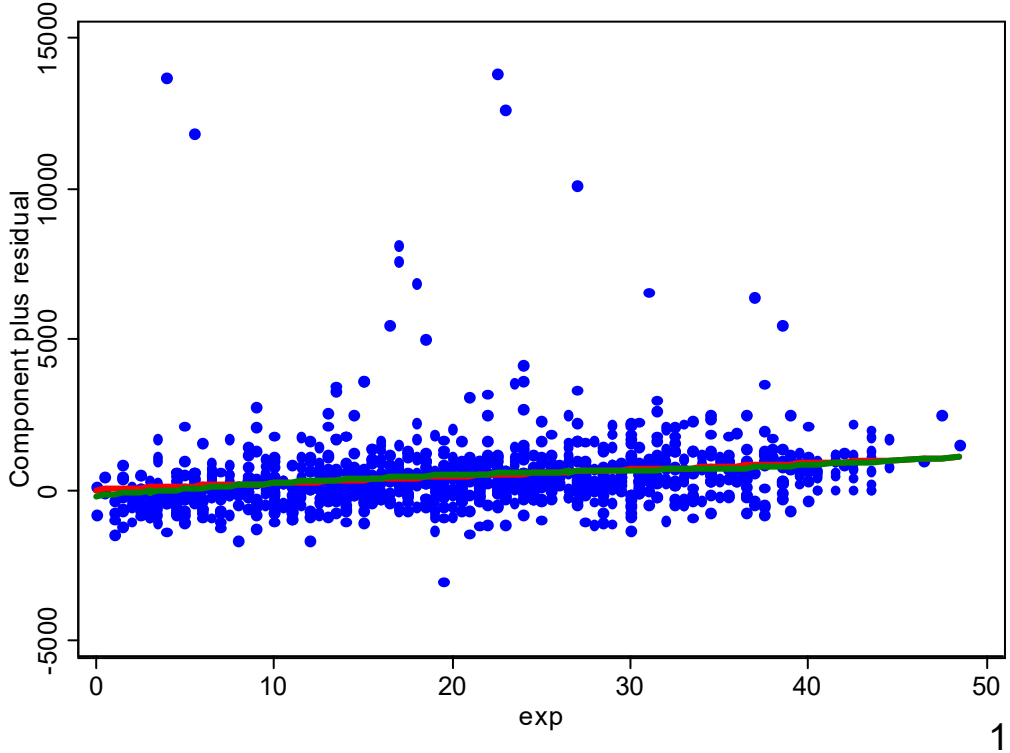
Linearität

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

- Nicht-Linearität erkennt man in einem Residuen-Plot
  - Residuen gegen die uV auftragen
    - Abweichungen von der Null-Linie Anzeichen von Nicht-Linearität
  - In STATA: component-plus-residual plot (cprplot)
    - hilfreich: nicht-parametrischer Smoother (lowess)
    - weicht der Lowess von Regressionsgerade ab, dann Nicht-Linearität

```
* Beispiel: Berufserfahrung

regress eink bild exp frau
cprplot exp, lowess
```

Der Lowess (grün) zeigt nur geringfügige Abweichungen von der Gerade.
Es liegt also keine Nicht-Linearität vor.

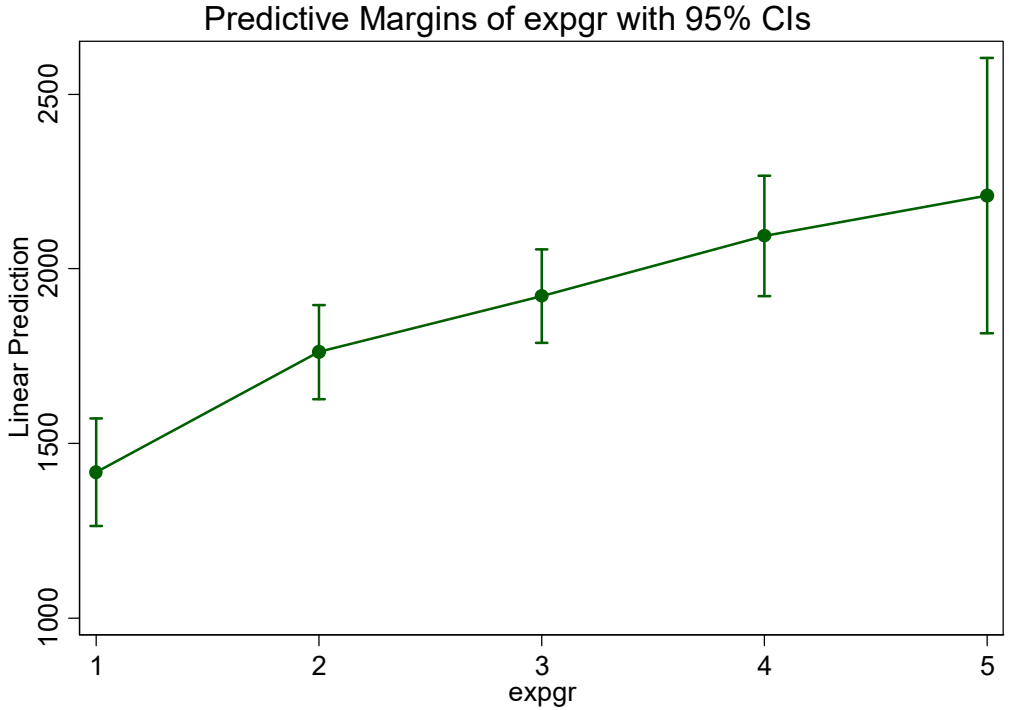


Linearität

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

- Alternative: Man gruppiert die Variable und plottet die für die Gruppen vorhergesagten Werte
  - So kann man graphisch die Linearität beurteilen
  - Man kann auch testen, welche Polynomterme signifikant sind

```
recode expgr 0/10=1 10/20=2 20/30=3 30/40=4 40/50=5
regress eink bild frau i.expgr
margins expgr                      //Vorhergesagtes Einkommen in den Gruppen
marginsplot                        //Plot der vorhergesagten Werte
contrast p.expgr, asobserved       //Test, ob Polynomterme signifikant sind
```




```
------------------------------------------
             | df         F       P>F
-------------+----------------------------
       expgr |
    (linear) |  1     18.08    0.0000
 (quadratic) |  1      0.92    0.3371
     (cubic) |  1      0.17    0.6825
   (quartic) |  1      0.18    0.6715
       Joint |  4     10.44    0.0000
             |
    Residual | 1111
------------------------------------------
```

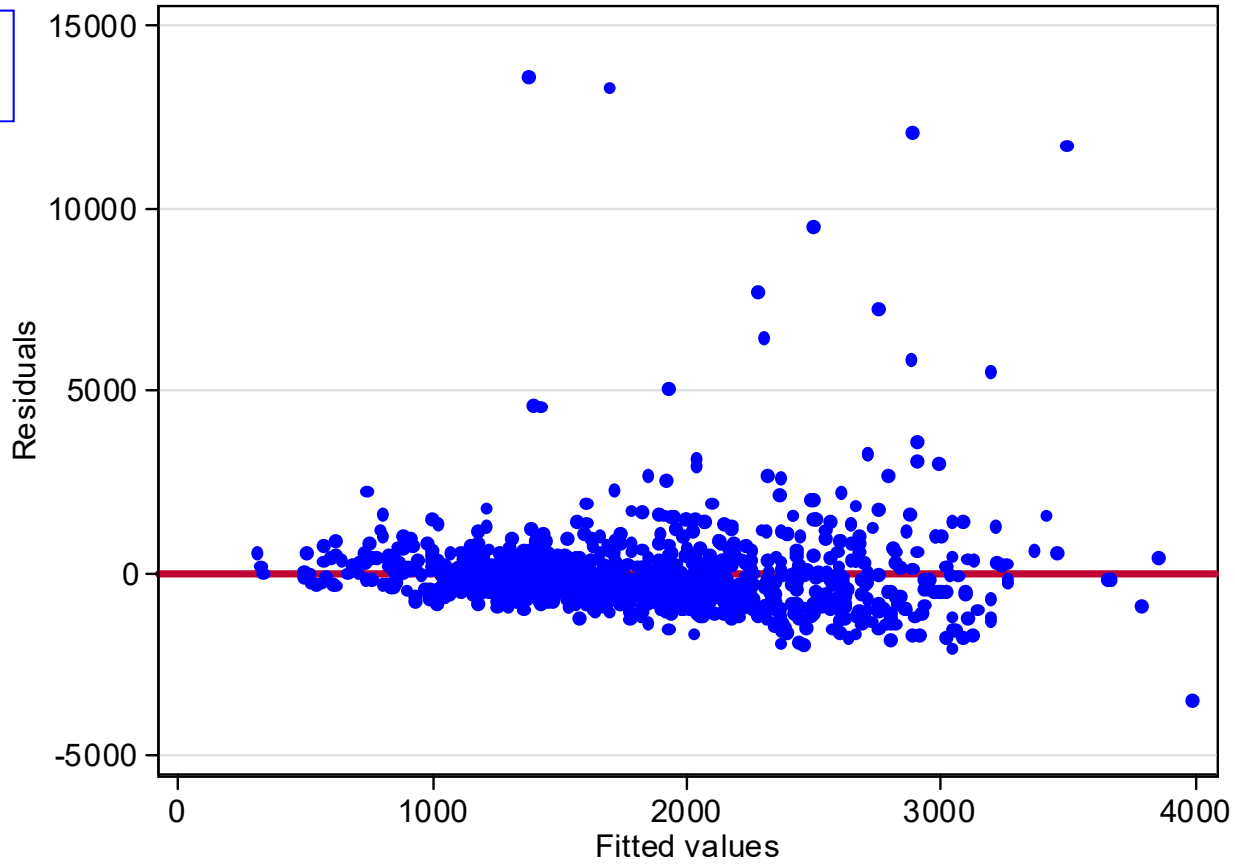
Homoskedastizität

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

- Heteroskedastizität: Residuen streuen unterschiedlich
  - STATA: residual-versus-fitted-values Plot (rvfplot)

```
regress eink bild exp frau
rvfplot, yline(0)
```

Deutlicher Trichter erkennbar: Streuung der Residuen bei großen Werten von y-Dach höher.
Grund: rechtsschiefe Einkommensverteilung.
Abhilfe: Transformation (s.u.)

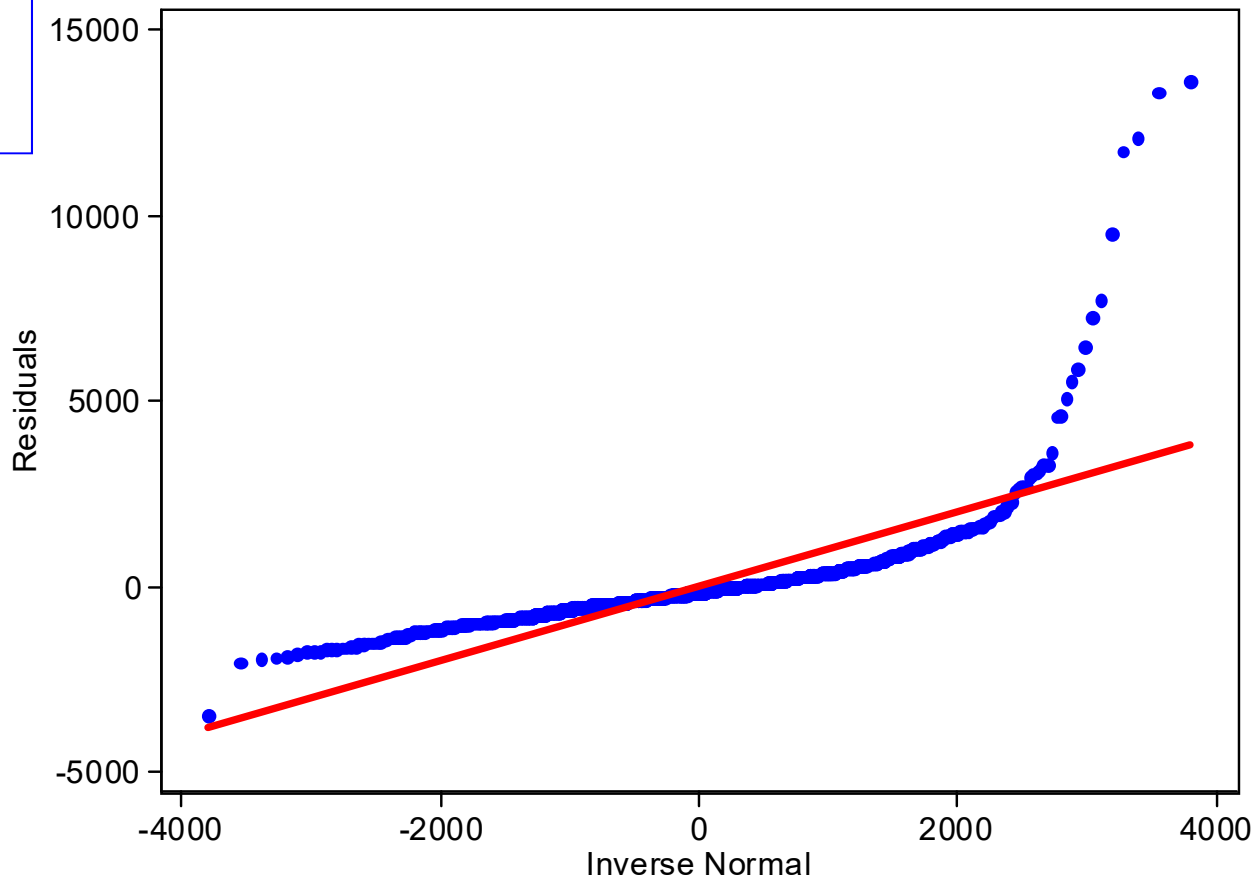


Normalverteilungsannahme

- Folgen die Residuen einer Normalverteilung?
  - STATA: Normal-Probability Plot (qnorm)

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

```
regress eink bild exp frau
predict res1, residual
qnorm res1
```

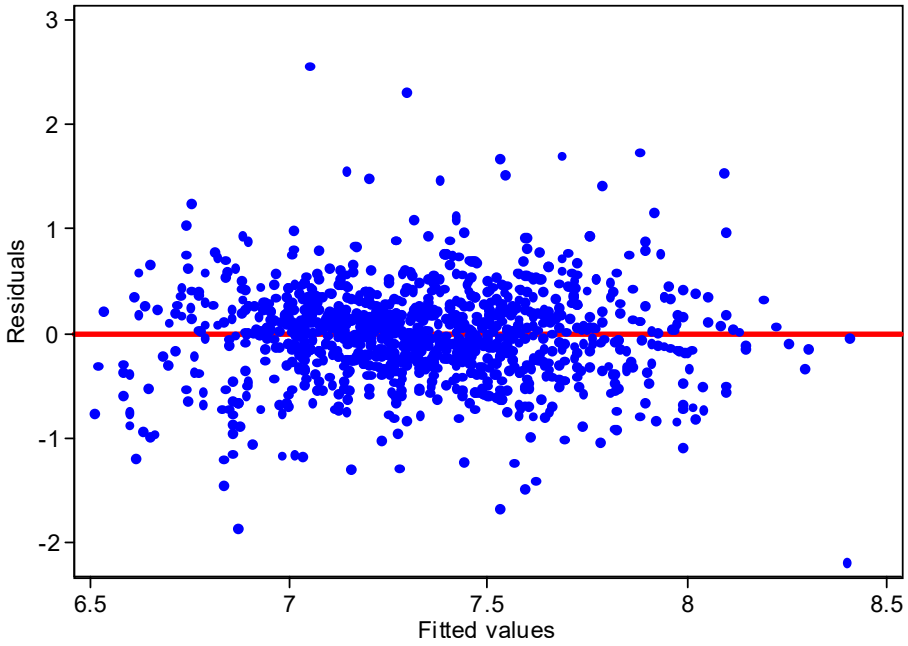


Die Residuen weichen deutlich von der roten Referenzlinie ab. Die Normalverteilungsannahme ist verletzt.
Grund: rechtsschiefe Einkommensverteilung
Abhilfe: logarithmische Transformation

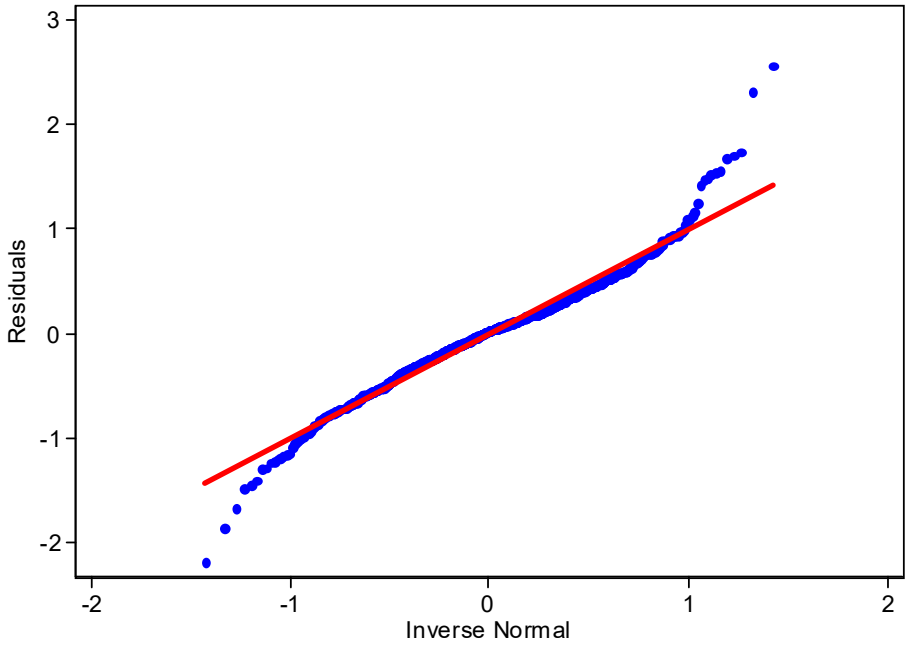
Semi-logarithmische Einkommensregression

```
. * logarithmische Transformation der aV
. generate lneink = ln(eink)
. regress lneink bild exp frau
```

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do



Kein augenfälliges Muster erkennbar: Homoskedastie



Nur mehr geringe Abweichungen an den Ränder: Residuen fast normalverteilt

Robuste Standardfehler

- Man kann die S.E.s „robust“ berechnen, d.h. sie sind robust gegen Verletzungen von A3
  - Huber-White-Sandwich-Estimator (option: `vce(robust)`)

	(1) normale S.E.s	(2) robuste S.E.s
bild	181.58***	181.58***
	(12.36)	(18.75)
exp	22.15***	22.15***
	(3.38)	(3.18)
frau	-474.57***	-474.57***
	(76.49)	(70.69)
N	1118	1118
R-sq	0.203	0.203

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Normaler Standardfehler:

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1},$$

$$\text{wobei} \quad \hat{\sigma}^2 = \frac{\sum_i \hat{\epsilon}_i^2}{n-k}.$$

Huber-White-Sandwich-Estimator:

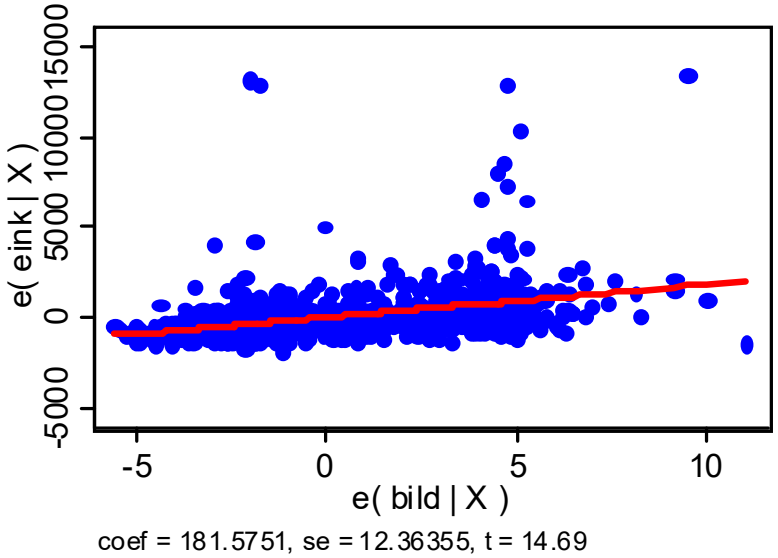
$$\hat{V}_{\mathrm{W}}(\hat{\beta}) = (X'X)^{-1} X'DX (X'X)^{-1},$$

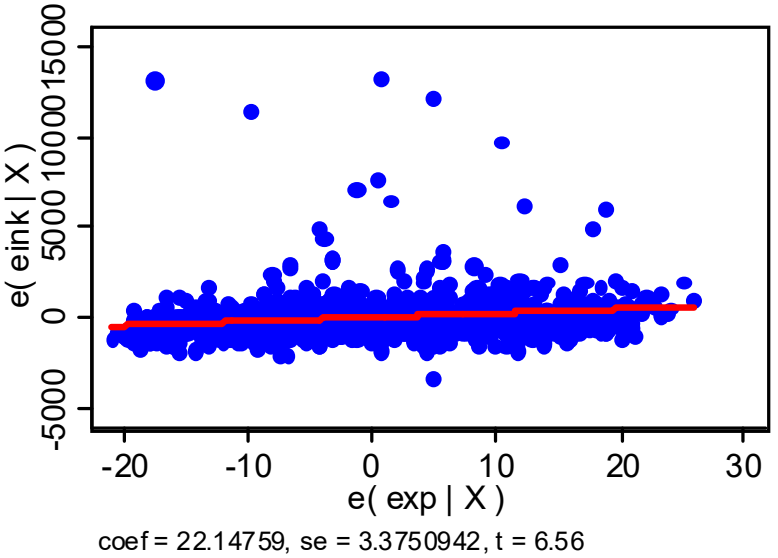
$$\text{wobei } D = \operatorname{diag}(\hat{\epsilon}_1^2, \ldots, \hat{\epsilon}_n^2)$$

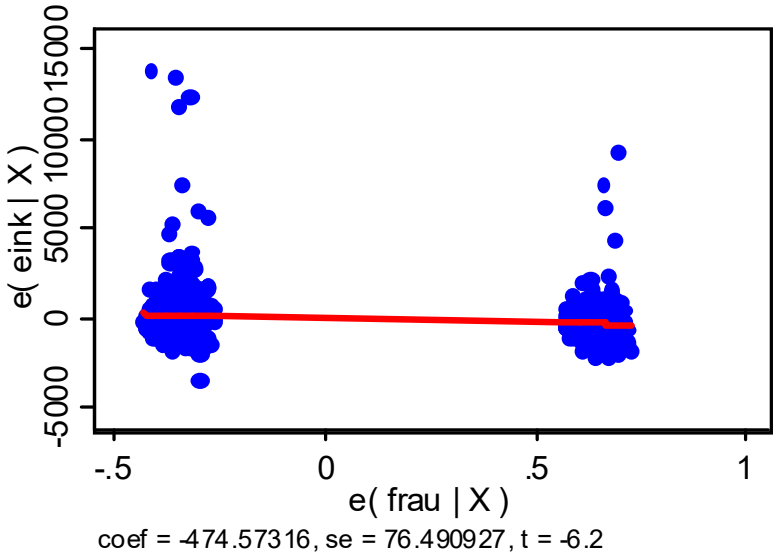
Ausreißerdiagnostik

- Ein Datenpunkt ist einflussreich, wenn seine Beseitigung die Ergebnisse der Regression deutlich verändert
  - Fälle mit ungewöhnlichem X- und Y-Wert (Ausreißer) haben Einfluss
  - Problem: das Ergebnis repräsentiert evtl. nur wenige Ausreißer
- Einflussdiagnostik
  - Im Streudiagramm erkennt man einflussreiche Datenpunkte
  - Im multiplen Fall: Partielles-Regressions Streudiagramm
  - Cook‘s D: Veränderung der Regressionskoeffizienten, wenn man einen Fall weglässt. Fälle mit besonders hohem D haben starken Einfluss.
- Abhilfe
  - Ist der einflussreiche Datenpunkt korrekt vercodet?
  - Fehlspezifikation? Was haben die einflussreichen Datenpunkte gemeinsam?
  - Weglassen ist keine Lösung, das ist Manipulation!

Partielle-Regressions Streudiagramme





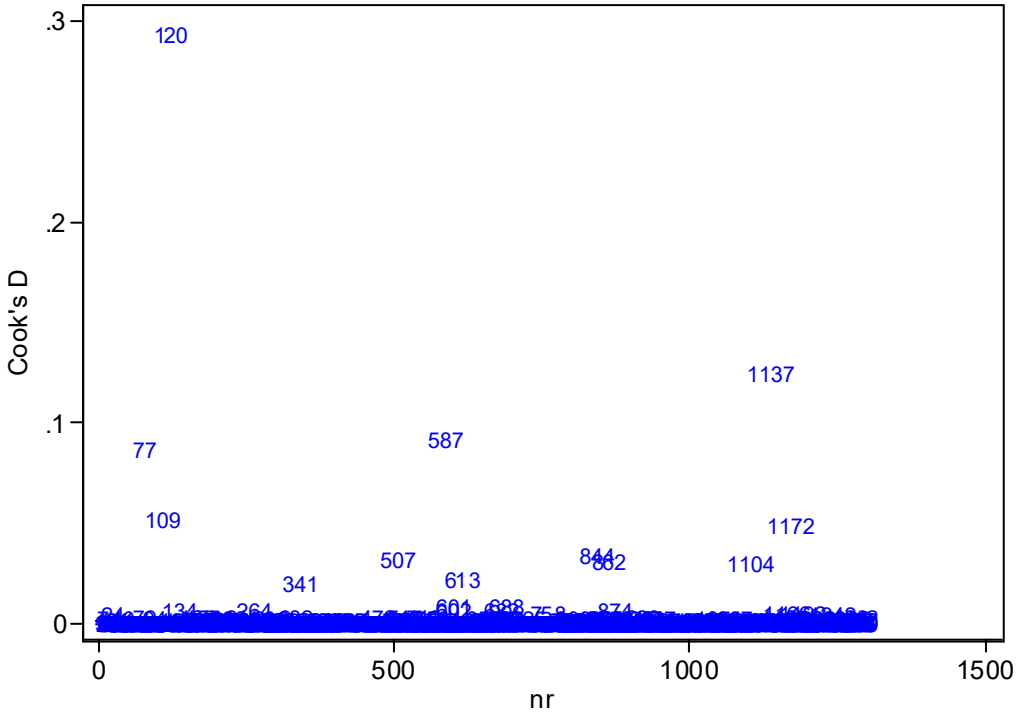


Plottyp: Added-variable plots

`avplots`

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

Einflussreiche Datenpunkte




```
* Indexplot von Cook's D
gen nr=_n
scatter D nr,  msymbol(i) mlabel(nr)
mlabposition(0)
```

Besonderen Einfluss hat Fall 120.
Wir schauen uns die Fälle über 0,1 an.

```
. list eink bild exp frau if D>0.1 & D~=.

     +---------------------------+
     |  eink   bild   exp   frau |
     |---------------------------|
 120. | 15200   23.5   5.5      0 |
1137. | 15000     12     4      0 |
     +---------------------------+
```

Es handelt sich um „Großverdiener“.
Überprüfen, ob deren Einkommen richtig vercodet wurde.

Daten: ALLBUS 2002
Do-File: 6 LinReg Diagnostik.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 9:
Maximum-Likelihood Schätzung

Maximum-Likelihood Schätzung

- Die Parameter nicht-linearer Regressionsmodelle werden meist mit Maximum-Likelihood (ML) geschätzt
- Kurze Darstellung des Maximum-Likelihood Prinzips
  - Gegeben Daten: $(y_i, \boldsymbol{x}_i)$
  - Gegeben Regressionsmodell: $f(Y = y_i | \boldsymbol{x}_i, \boldsymbol{\beta})$
  - Schätzprinzip: Bestimme $\boldsymbol{\beta}$ so, dass die Wahrscheinlichkeit diese Daten zu beobachten, maximal wird
  - Die Wahrscheinlichkeit (Likelihood) der Daten unter dem gegebenen Modell und unabhängiger Stichprobenziehung ist

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i, x_i; \boldsymbol{\beta})$$

  - Für die Berechnung ist es vorteilhaft, die Log-Likelihood zu maximieren

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln f(y_i, x_i; \boldsymbol{\beta})$$

  - Ableiten und Null-Setzen liefert die ML-Schätzer

Eigenschaften der ML-Schätzer

- ML-Schätzer haben einige wünschenswerte Eigenschaften (asymptotisch!)
  - Konsistent (unverzerrt)

$$E(\hat{\boldsymbol{\beta}}_{ML}) = \boldsymbol{\beta}$$

  - Normalverteilt

$$\hat{\boldsymbol{\beta}}_{ML} \sim N(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1}), \quad \text{wobei } I(\boldsymbol{\beta}) = -E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'}\right)$$

  - Effizient
    - Die ML-Schätzer haben minimale Varianz (unter den konsistenten Schätzern)
    - Sie erreichen die Rao-Cramer Schranke

ML-Schätzer des binären Logit-Modells

- Anwendung auf das binäre Logit-Modell (McFadden 1972)

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ P(Y = 1|\boldsymbol{x}_i, \boldsymbol{\beta})^{y_i} \cdot P(Y = 0|\boldsymbol{x}_i, \boldsymbol{\beta})^{(1-y_i)} \right\}$$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \left[\frac{e^{\boldsymbol{\beta}'\mathbf{x_i}}}{1 + e^{\boldsymbol{\beta}'\mathbf{x_i}}} \right]^{y_i} \cdot \left[\frac{1}{1 + e^{\boldsymbol{\beta}'\mathbf{x_i}}} \right]^{(1-y_i)} \right\}$$

  - Logarithmieren, Ableiten und Null-Setzen liefert die Schätzgleichungen

$$\sum_{i=1}^{n} y_i \mathbf{x_i} = \sum_{i=1}^{n} \frac{e^{\hat{\boldsymbol{\beta}}'\mathbf{x_i}}}{1 + e^{\hat{\boldsymbol{\beta}}'\mathbf{x_i}}} \mathbf{x_i}$$

  - Dies ist ein nicht-lineares Gleichungssystem,
    Lösung deshalb mittels iterativer numerischer Algorithmen

Signifikanztests

- Test eines einzelnen Regressionskoeffizienten
  - Nullhypothese: X_j hat keinen Einfluss auf Y (kein Zusammenhang)

    $H_0: \beta_j = 0$
  - Die Teststatistik (z-Wert) ist $Z = \frac{\hat{\beta}_j}{\hat{\sigma}_j} \overset{a}{\sim} N(0,1)$
  - Die H_0 wird abgelehnt, falls $|Z| > z_{1-\alpha/2}$
    - Ab n>100 sinnvoll (Faustregel für α=5%: |Z| > 2)
- Test des gesamten Modells: Likelihood-Ratio (LR) Test
  - Nullhypothese: keine X-Variable hat einen Einfluss auf Y

    $H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0$
  - Die Teststatistik (LR-Wert) ist $\chi^2 = -2\ln\left(\frac{L_0}{L_1}\right) = 2\left(\ln L_1 - \ln L_0\right)$
    - L_0: Likelihood des Modells nur mit Konstante (Nullmodell)
    - L_1: Likelihood des Gesamtmodells
  - Die H_0 wird verworfen, falls: $\chi^2 > \chi^2_{1-\alpha}$ (p)

Modellfit: Pseudo-R^2

- R^2 nicht sinnvoll, da keine sinnvolle Streuungszerlegung
- In Analogie: Pseudo-R^2 Maße
  - Wie viel von der Likelihood des Nullmodells wird durch das Gesamtmodell „erklärt“
    - Null, wenn die weiteren X-Variablen nichts erklären
    - Maximum allerdings kleiner Eins
  - McFadden‘s Pseudo-R^2

$$R^2_{MF} = \frac{\ln L_0 - \ln L_1}{\ln L_0}$$

    - nicht: Anteil erklärter Varianz
    - Relative Log-Likelihood Verbesserung (im Vergleich zum Nullmodell)
    - Fällt kleiner aus, als das R^2 des Linearen Wahrscheinlichkeitsmodells
  - Weitere Pseudo-R^2 s. Long/Freese (2006) S. 109 ff
    - Manche Autoren präferieren McKelvey/Zavoina‘s R^2

ML-Output

- Beispiel: Logit Modell der Arbeitslosigkeit
  - Im ALLBUS 2002 wurden Erwerbstätige gefragt, ob sie die letzten 10 Jahre arbeitslos waren (1=Arbeitslosigkeit)
    - Hinzu: die gegenwärtig Arbeitslosen
    - Sinnvoll: Einschränkung auf 30-65 Jährige

Daten: ALLBUS 2002
Do-File: 7 ML.do

Das Iterationsprotokoll

```
. logit arblos bild alter frau ost

Iteration 0:   log likelihood =  -850.0254
Iteration 1:   log likelihood = -776.09162
Iteration 2:   log likelihood = -774.80455
Iteration 3:   log likelihood = -774.80199
Iteration 4:   log likelihood = -774.80199
```

Protokoll der Iterationsschritte

- Hohe, negative Werte der Log-Likelihood, da Likelihood fast 0
- „Iteration 0“ ist der Startwert: **$\ln(L_0)$**
- Steigen an, d.h. die L der Daten nimmt zu
- Zum Schluss werden die Schritte kleiner
- Mit der 4. Iteration ist die Konvergenz erreicht: **$\ln(L_1)$**

Header

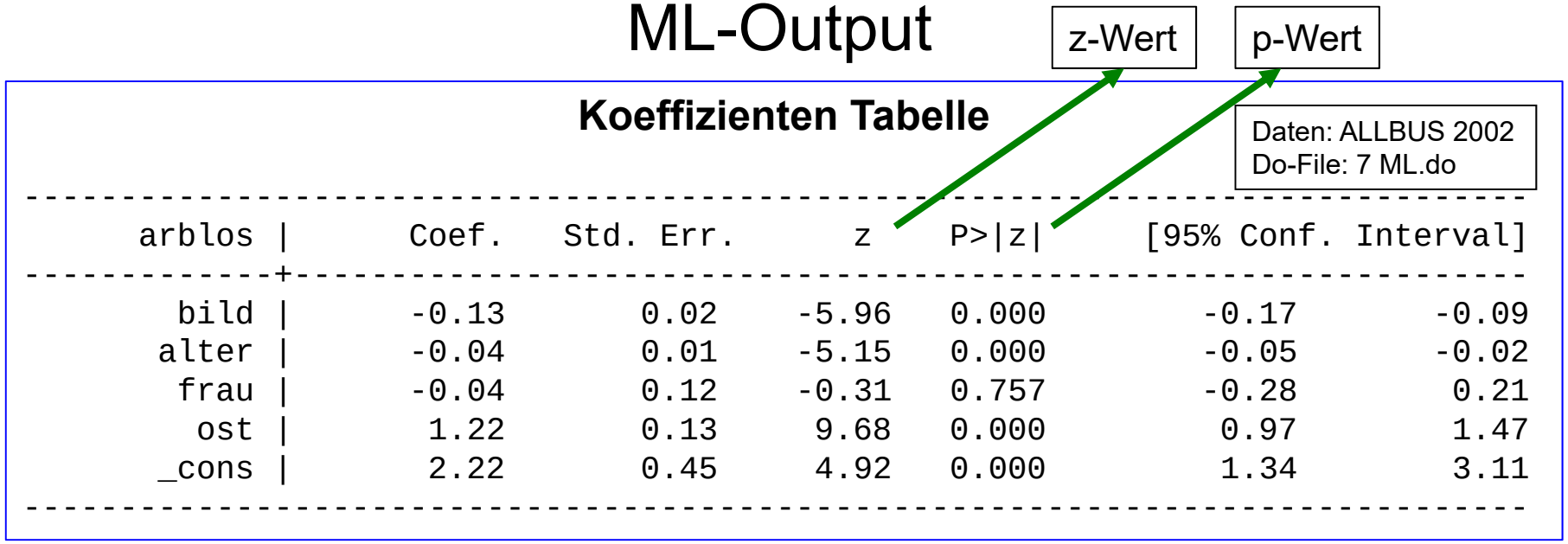
```
Logistic regression       Number of obs =   1329
                          LR chi2(4)    = 150.45
                          Prob > chi2   = 0.0000
Log likelihood = -774.802 Pseudo R2     = 0.0885
```

Schätzsample nach „listwise deletion“

LR−Teststatistik: $2(-775 + 850)$

R^2−McFadden: $\frac{-850 + 775}{-850}$

ML-Output



Koeffizienten Tabelle

arblos	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
bild	-0.13	0.02	-5.96	0.000	-0.17	-0.09
alter	-0.04	0.01	-5.15	0.000	-0.05	-0.02
frau	-0.04	0.12	-0.31	0.757	-0.28	0.21
ost	1.22	0.13	9.68	0.000	0.97	1.47
_cons	2.22	0.45	4.92	0.000	1.34	3.11

- Zahl der Nachkommastellen in der Tabelle
  - `set cformat %9.2f  //Format der Koeffizienten, S.E., KI`
  - `set pformat %5.3f  //Format des p-Wert`
- Schätzoptionen
  - `level(90)` 90%-Signifikanzniveau
  - `vce(robust)` robuste Standardfehler (Huber-White-Sandwich)
  - `vce(cluster` *VC*`)` Abhängigkeit in den durch *VC* definierten Gruppen
- Weitere Schätzstatistiken
  - `estat vce` Varianz-Kovarianzmatrix der ML-Schätzer

Standardisierte Koeffizienten

- Manchmal will man Koeffizienten vergleichen: Standardisierung
  - X-Standardisierung: Koeffizienten multipliziert mit Standardabw. von X
  - Volle-Standardisierung: Koeffizienten multipliziert mit Standardabw. von X und dividiert mit Standardabweichung von Y

```
. listcoef, help          //funktioniert nur, wenn SPost Ados geladen sind

------------------------------------------------------------------------------
      arblos |      b         z     P>|z|    e^b    e^bStdX      SDofX
-------------+----------------------------------------------------------------
        bild |  -0.13053   -5.959   0.000   0.8776   0.6722     3.0425
       alter |  -0.03801   -5.154   0.000   0.9627   0.7199     8.6480
        frau |  -0.03851   -0.309   0.757   0.9622   0.9811     0.4942
         ost |   1.22097    9.684   0.000   3.3905   1.7956     0.4794
------------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

Daten: ALLBUS 2002
Do-File: 7 ML.do

Tests für komplexere Hypothesen

Sind Alter und Alter^2 gemeinsam signifikant?

Daten: ALLBUS 2002
Do-File: 7 ML.do

```
. logit arblos bild frau ost
  (output omitted)
. estimates store null
.
. logit arblos bild frau ost alter c.alter#c.alter
------------------------------------------------------------------------------
      arblos |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bild |      -0.13       0.02    -5.96   0.000        -0.17       -0.09
        frau |      -0.03       0.13    -0.21   0.830        -0.27        0.22
         ost |       1.23       0.13     9.71   0.000         0.98        1.47
       alter |      -0.13       0.07    -1.79   0.073        -0.28        0.01
c.alter#c.alter |       0.00       0.00     1.29   0.197        -0.00        0.00
       _cons |       4.28       1.66     2.57   0.010         1.02        7.54
------------------------------------------------------------------------------
. estimates store full
.
. test alter alter#alter     //Wald Test

 ( 1)  [arblos]alter = 0
 ( 2)  [arblos]c.alter#c.alter = 0

           chi2(  2) =   28.66
         Prob > chi2 =    0.0000

. lrtest null full           //LR-Test

Likelihood-ratio test                                 LR chi2(2)  =     29.18
(Assumption: null nested in full)                     Prob > chi2 =    0.0000
```

Fitmaße

```
. logit arblos bild alter frau ost
  (output omitted)

. fitstat                  //funktioniert nur, wenn SPost Ados geladen sind

Measures of Fit for logit of arblos

Log-Likelihood Maße
Log-Lik Intercept Only:       -850.025   Log-Lik Full Model:           -774.802
                                         LR(4):                        150.447
                                         Prob > LR:                      0.000

Pseudo R2 Maße (zum Vergleich: R2 = 10,9%)
McFadden's R2:                   0.088   McFadden's Adj R2:              0.083
ML (Cox-Snell) R2:               0.107   Cragg-Uhler(Nagelkerke) R2:     0.148
McKelvey & Zavoina's R2:         0.149   Efron's R2:                     0.110

Maße aus der Vorhersagetabelle
Count R2:                        0.698   Adj Count R2:                   0.105

Informations Maße
AIC:                             1.174   AIC*n:                       1559.604
BIC:                         -7972.845   BIC':                        -121.678
```

Daten: ALLBUS 2002
Do-File: 7 ML.do

Modellvergleich mit BIC

- Vorschlag von Raftery (1996, Sociological Methodology)
  - BIC (Bayesian Information Criterion) zum Vergleich auch von nicht-verschachtelten Modellen. Mehrere Varianten: hier BIC‘

$$\mathrm{BIC}' = -LR + p \ln N$$

    - LR: Likelihood-Ratio Teststatistik, p: Zahl Regressoren, N: Fallzahl
  - Das Modell mit dem negativeren BIC‘ ist besser

Daten: ALLBUS 2002
Do-File: 7 ML.do

```
. logit arblos bild frau ost
. fitstat, saving(mod1)         //funktioniert nur, wenn SPost Ados geladen sind
. logit arblos bild frau ost alter c.alter#c.alter
. fitstat, using(mod1)          //funktioniert nur, wenn SPost Ados geladen sind

                              Current             Saved        Difference
Model:                          logit             logit

LR                          152.092(5)        122.908(3)         29.184(2)
Prob > LR                     0.000             0.000              0.000

McFadden's R2                 0.089             0.072              0.017

BIC'                       -116.131          -101.332           -14.800

Difference of   14.800 in BIC' provides very strong support for current model.

Note: p-value for difference in LR is only valid if models are nested.
```

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 10:
Logistische Regression

- Das lineare Wahrscheinlichkeitsmodell
- Das logistische Modell
- Multiple logistische Regression
- Interpretation der Koeffizienten
- Interaktionseffekte
- Das Skalierungsproblem
- Logit vs. Probit

Regression mit kategorialen Outcomes

- Bei kategorialen Outcomes kann man lineare Wahrscheinlichkeitsmodelle einsetzen (linear probability model, LPM)
- Meist werden aber nicht-lineare Modelle verwendet (nonlinear probability models, NLPM)
  - Binäres Outcome
    - Logit (logistische Regression), Probit, etc.
  - Multinomiales Outcome
    - Multinomiales Logit, multinomiales Probit
  - Ordinales Outcome
    - Ordinales Logit, ordinales Probit

- Kodierung
  - Die abhängige Variable sollte 0, 1, … kodiert sein

Das lineare Wahrscheinlichkeitsmodell

- Man kann eine dichotome aV linear modellieren

$$E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = \boldsymbol{\beta}' \boldsymbol{x}$$

- Da gilt $E(y) = P(Y = 0) \cdot 0 + P(Y = 1) \cdot 1 = P(Y = 1)$
- Erhält man das lineare Wahrscheinlichkeitsmodell (LPM)

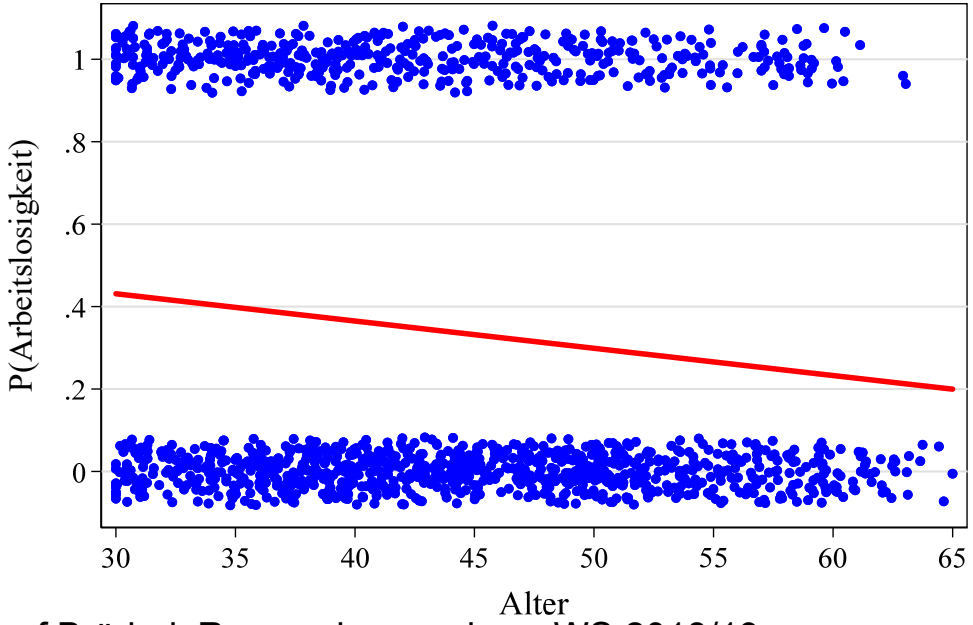
$$P(Y = 1) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = \boldsymbol{\beta}' \boldsymbol{x}$$

- Beispiel: Im ALLBUS 2002 wurden Erwerbstätige gefragt, ob sie die letzten 10 Jahre arbeitslos waren (1=Arbeitslosigkeit)
  - Hinzu: die gegenwärtig Arbeitslosen
  - Sinnvoll: Einschränkung auf 30-65 Jährige

Das lineare Wahrscheinlichkeitsmodell

```
. regress arblos alter

      Source |       SS           df       MS      Number of obs   =     1,330
-------------+----------------------------------   F(1, 1328)      =     19.68
       Model |  4.34325173         1  4.34325173   Prob > F        =    0.0000
    Residual |  293.077049     1,328  .220690549   R-squared       =    0.0146
-------------+----------------------------------   Adj R-squared   =    0.0139
       Total |  297.420301     1,329  .223792551   Root MSE        =    .46978

------------------------------------------------------------------------------
      arblos |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alter |  -.0066123   .0014905    -4.44   0.000    -.0095363   -.0036883
       _cons |   .6295199   .0670537     9.39   0.000     .4979772    .7610625
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do



Mit jedem Jahr, sinkt die Whs. arbeitslos gewesen zu sein um 0,7 Prozentpunkte.

Probleme LPM:

1) LPM macht Prognosen jenseits [0,1].
2) Fehler sind heteroskedastisch und nicht-normalverteilt.
3) Die lineare Funktion unterstellt konstante Effekte. Oft macht es Sinn, abnehmende Effekte zu modellieren, wenn P nahe 0 oder 1 ist.

Das Logit-Modell

- Aufgrund der Probleme des LPM verwendet man zur Modellierung von Wahrscheinlichkeiten oft Verteilungsfunktionen F(.), deren Wertebereich [0,1] ist:

$$P(Y = 1) = F(\boldsymbol{\beta}'\boldsymbol{x})$$

  - **Probit:** Standard-Normalverteilung $\Phi(.)$ [Varianz = 1]

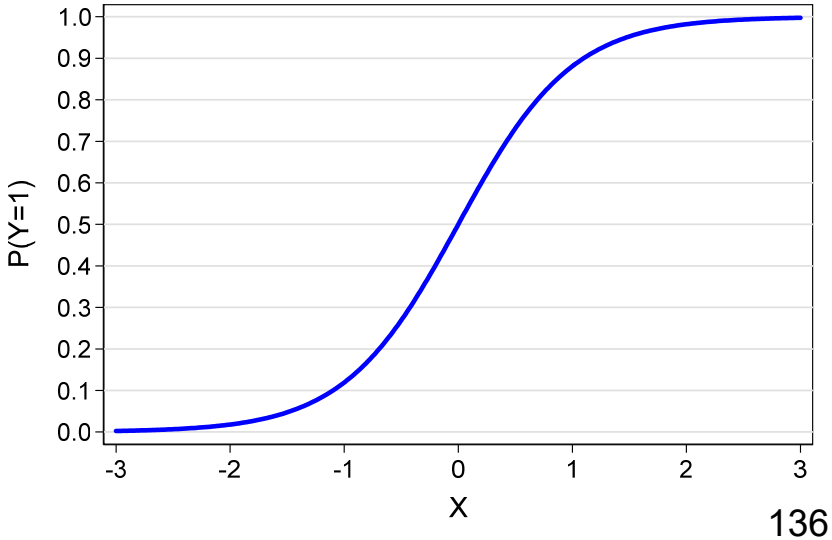
$$P(Y = 1) = \Phi(\boldsymbol{\beta}'\boldsymbol{x})$$

  - **Logit:** Standard-Logistische-Verteilung $\Lambda(.)$ [Varianz = $\pi^2/3 = 3{,}29$]

$$P(Y = 1) = \frac{e^{\boldsymbol{\beta}'\boldsymbol{x}}}{1 + e^{\boldsymbol{\beta}'\boldsymbol{x}}} = \frac{1}{1 + e^{-\boldsymbol{\beta}'\boldsymbol{x}}}$$

$$P(Y = 0) = 1 - P(Y = 1) = \frac{1}{1 + e^{\boldsymbol{\beta}'\boldsymbol{x}}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(2 \cdot x)}}$$



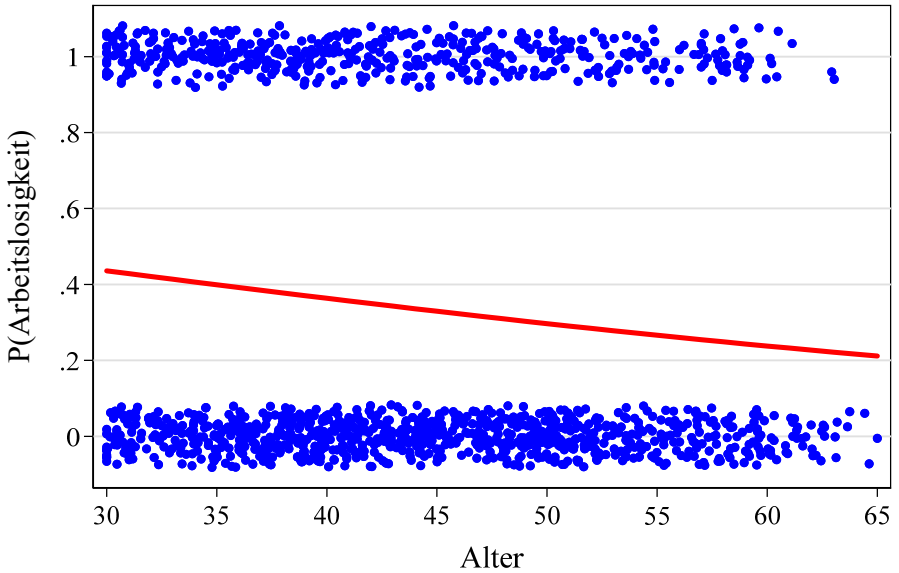
Das Logit-Modell

```
. logit arblos alter

Iteration 0:   log likelihood = -850.43747  [ln(L0)]
...
Iteration 3:   log likelihood = -840.59929  [ln(L1)]

Logistic regression                     Number of obs     =      1,330
                                        LR chi2(1)        =      19.68
                                        Prob > chi2       =     0.0000
Log likelihood = -840.59929             Pseudo R2         =     0.0116
------------------------------------------------------------------------------
      arblos |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alter |  -.0302412    .006902    -4.38   0.000    -.0437688   -.0167136
       _cons |   .6498028   .3051853     2.13   0.033     .0516507    1.247955
------------------------------------------------------------------------------
```

LR-Teststatistik:
$2(-840{,}6 + 850{,}4)$

$$R^2_{MF} = \frac{-850{,}4 + 840{,}6}{-850{,}4}$$

z-Wert

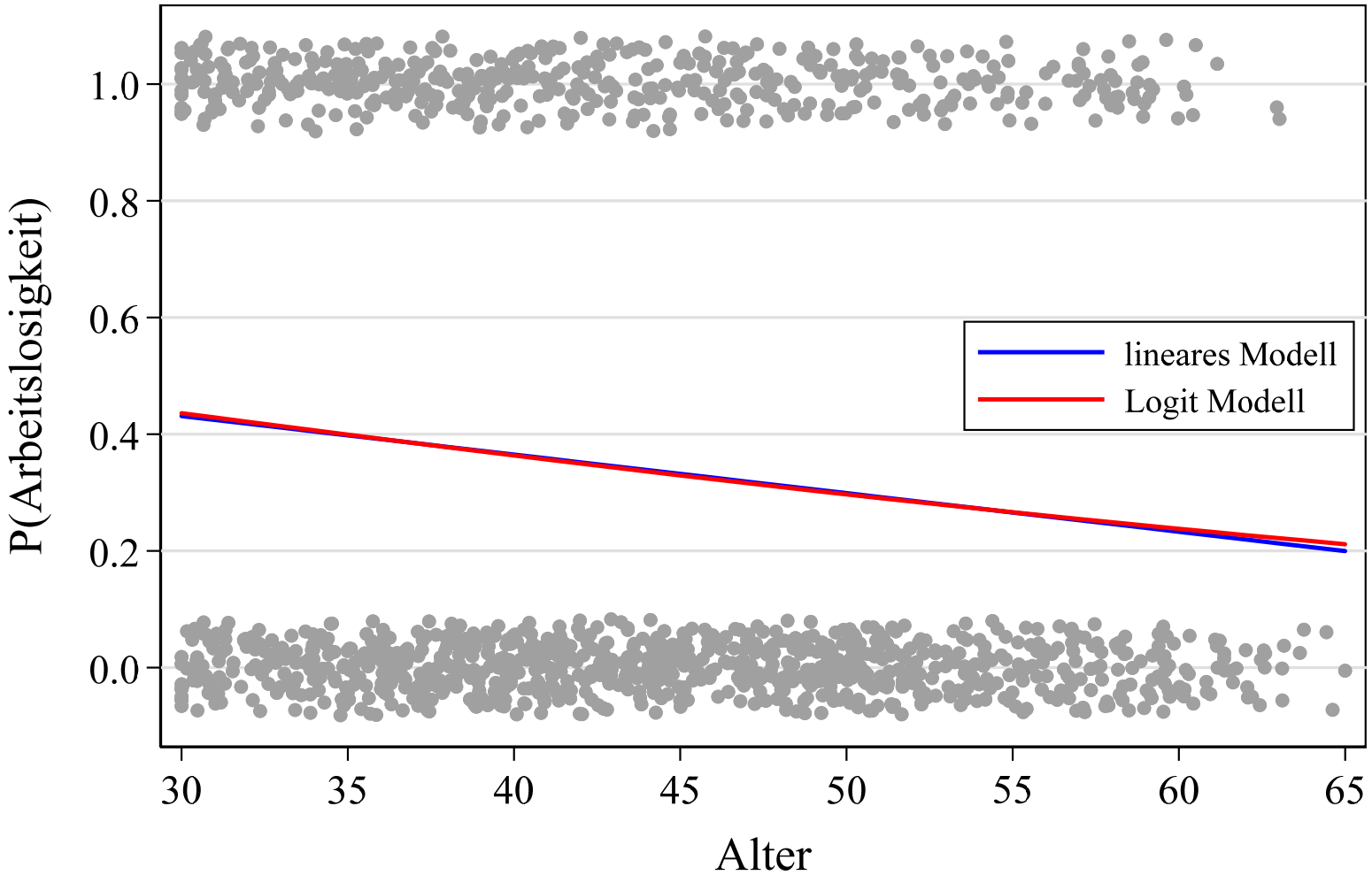


$$P(Y = 1) = \frac{1}{1 + e^{-(0.65 - 0.03 \cdot x)}}$$

Nur Vorzeicheninterpretation: Mit jedem Jahr, sinkt die Whs. arbeitslos gewesen zu sein.

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

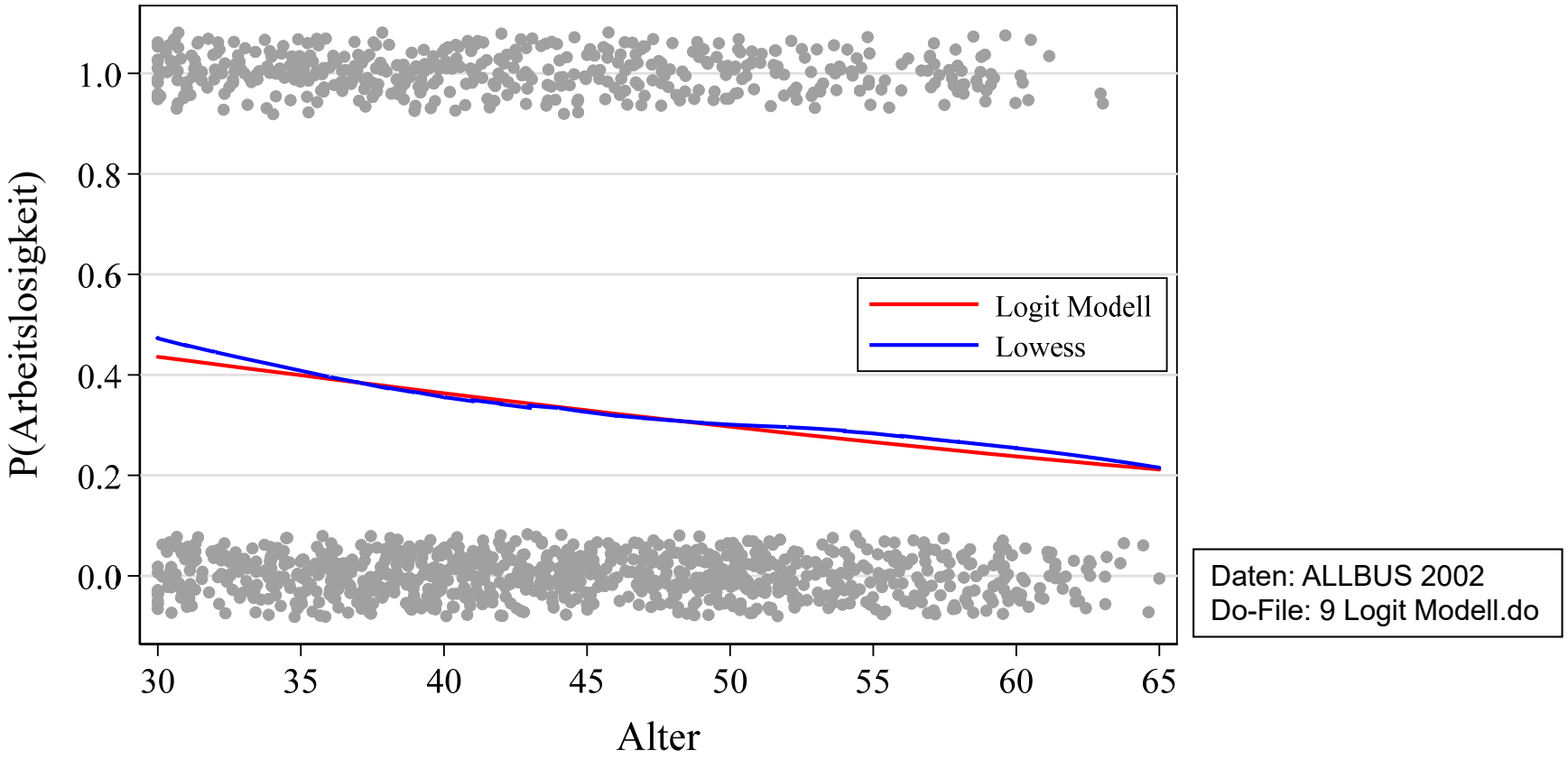
STATA-Beispiel: Vergleich LPM und Logit-Modell



- Liegt P(Y=1) im Intervall [0,2; 0,8], so kommen LPM und Logit-Modell zu identischen Ergebnissen

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

STATA-Beispiel: Diagnostik (funktionale Form)



Das Logit-Modell repräsentiert den Zusammenhang in den Daten ganz gut.
Der Zusammenhang ist annähernd linear.
Problem: Überprüfung ist hier nur bivariat!

STATA-Beispiel: diskretes X

```
. tab   arblos ost, col chi2

           |        ost
    arblos |         0          1 |     Total
-----------+----------------------+----------
         0 |       645        236 |       881
           |     75.44      49.68 |     66.24
-----------+----------------------+----------
         1 |       210        239 |       449
           |     24.56      50.32 |     33.76
-----------+----------------------+----------
     Total |       855        475 |     1,330
           |    100.00     100.00 |    100.00

 Pearson chi2(1) =  90.5715   Pr = 0.000
```

In diesem Fall reproduziert das logistische Modell exakt die Kreuztabelle (saturiertes Modell)

$$P(Y = 1|ost = 0) = \frac{1}{1 + e^{-(-1{,}12)}} = 0{,}246$$

$$P(Y = 1|ost = 1) = \frac{1}{1 + e^{-(-1{,}12+1{,}13)}} = 0{,}503$$

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

```
. logit arblos ost
                                                  LR chi2(1)      =      89.14

------------------------------------------------------------------------------
      arblos |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ost |   1.134775   .1213824     9.35   0.000      .8968694     1.37268
       _cons |  -1.122143   .0794499   -14.12   0.000     -1.277862   -.9664237
------------------------------------------------------------------------------
```

STATA-Beispiel: Multiple logistische Regression

```
. logit arblos bild alter frau ost[, or]

Logistic regression                                     Number of obs   =       1329
                                                        LR chi2(4)      =     150.45
                                                        Prob > chi2     =     0.0000
Log likelihood = -774.80199                             Pseudo R2       =     0.0885

------------------------------------------------------------------------------
      arblos |      Coef.   Std. Err.      z    P>|z|     Odds
-------------+----------------------------------------------------------------
        bild |  -.1305347   .0219059    -5.96   0.000     0.88
      alter |  -.0380065   .0073747    -5.15   0.000     0.96
       frau |  -.0385114   .1246573    -0.31   0.757     0.96
        ost |   1.220968   .1260793     9.68   0.000     3.39
      _cons |    2.22248   .4519836     4.92   0.000
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

Interpretation der Koeffizienten

- Wie bei linearer Regression: die Effekte der anderen uVs sind „herauspartialisiert“
- Logit-Effekte: nur das Vorzeichen ist interpretierbar
- Odds-Effekt: multiplikative Effekte auf die „Chance“ (z-Werte unverändert!)

Interpretation der Koeffizienten

- Das Logit-Modell hat drei äquivalente Formulierungen:

Whs.: $P(Y=1) = \frac{e^{\boldsymbol{\beta' x}}}{1+e^{\boldsymbol{\beta' x}}}$ „Wahrscheinlichkeit“

Odds: $\frac{P(Y=1)}{P(Y=0)} = e^{\boldsymbol{\beta' x}}$ „Odds / Chance“

Logit: $\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \boldsymbol{\beta' x}$ „log. Odds / log. Chance“

- Deshalb drei mögliche Interpretationen:
  - β_j ist der lineare, additive Effekt auf das Logit (unverständlich)
  - $\exp(\beta_j)$ ist der multiplikative Effekt auf die Odds (komplex)

$$\text{Odds}-\text{Ratio (OR)} := \frac{O_{x+1}}{O_x} = \frac{\exp(\beta(x+1))}{\exp(\beta x)} = \exp(\beta)$$

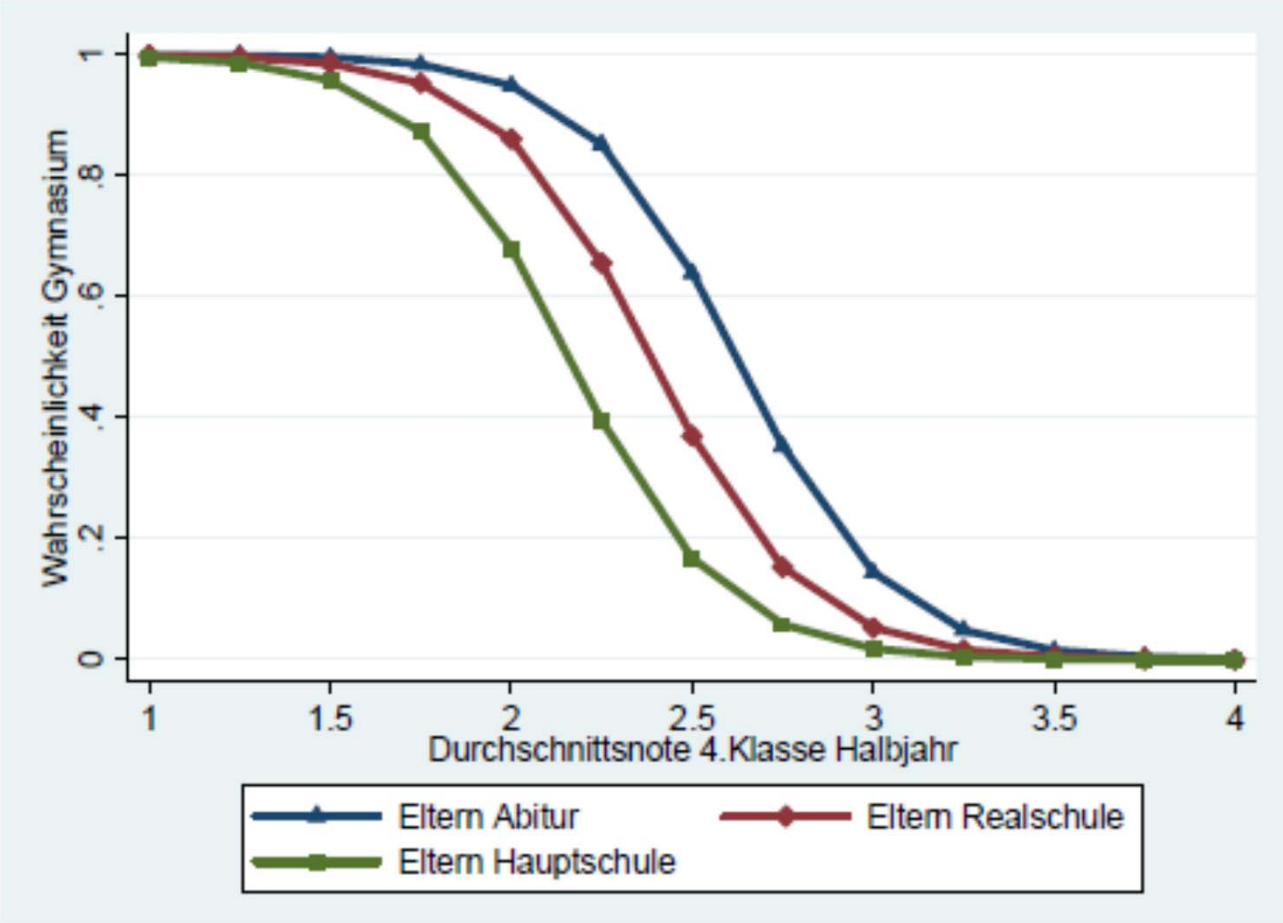
  - Wahrscheinlichkeitseffekte sind am anschaulichsten, müssen aber mit speziellen Routinen ausgerechnet werden

STATA-Beispiel

- Logit-Interpretation (Bildungs-Effekt)
  - Das Logit sinkt um 0,13 mit jedem zusätzlichen Bildungsjahr
  - Das ist sehr unanschaulich
- Odds-Interpretation (Ost-Effekt)
  - Das Odds (die „Arbeitslosigkeits-Chance“) ist im Osten um den Faktor exp(1,22) = 3,4 höher
  - Auch Prozentinterpretation: $(\exp(ß_j) - 1) \bullet 100$
    Das Odds ist im Osten um 240% höher
    - Falsch: die Whs. erhöht sich um 240%!
  - Die Odds-Ratio ist 3,4
    - Ca. Werte aus Kreuztabelle: P(Arblos|Ost)=0,5, P(Arblos|West)=0,25
    - $OR = {\frac{50}{50}}\Big/{\frac{25}{75}} = 3$
  - Odds („Chancen“) sind leider auch ziemlich unanschaulich

Achtung bei der Interpretation von Odds-Ratios

Odds-Ratios werden häufig falsch verstanden. Hierzu ein Beispiel.
Um wie viel höher ist die Wahrscheinlichkeit nach der 4. Klasse aufs Gymnasium zu gehen, für ein Kind von Abiturienten im Vergleich zu Kindern von Hauptschülern (bei gleicher Leistung, für die hier kontrolliert ist)?



Die Odds-Ratio in diesem Modell ist etwa 6.

Z.B. Note 2,5:

Abi	60%
Hauptschule	20%

→ OR = 60/40 / 20/80 = 6

Also ist die „Chance“ 6 mal so hoch.

Aber Achtung: das sind keine Whs.verhältnisse:
Bei 2,5 ist das 3
Bei 1,5 ist das ca. 1

Daten: DJI Kinderpanel
Quelle: Diplomarbeit Volker Roth

Wahrscheinlichkeitsinterpretation

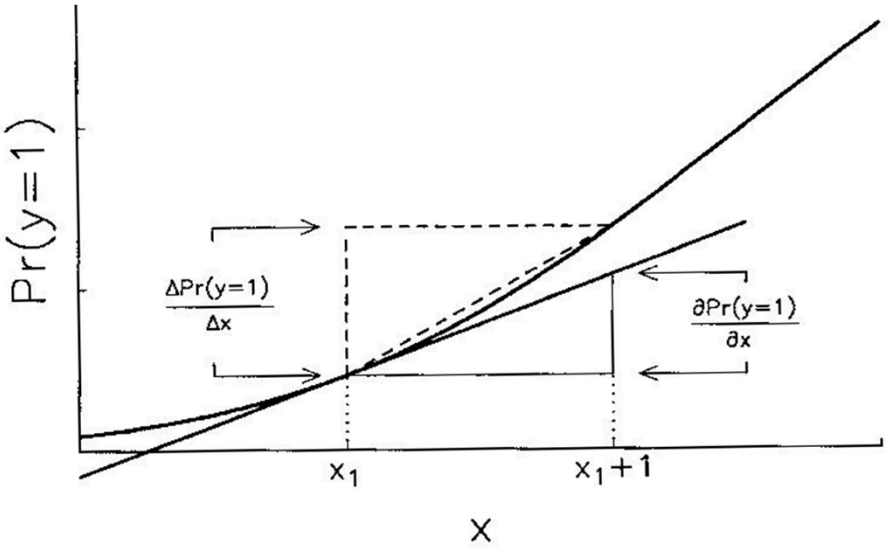
- „Discrete Change“ (DC) und „Marginaleffekt“ (ME)

$$\frac{\Delta P(Y=1)}{\Delta x_j} = P\left(Y=1|\overline{\boldsymbol{x}}, \overline{x_j}+1\right) - P\left(Y=1|\overline{\boldsymbol{x}}, \overline{x_j}\right)$$

$$\frac{\partial P(Y=1)}{\partial x_j} = F'(\boldsymbol{\beta}'\overline{\boldsymbol{x}}) = f(\boldsymbol{\beta}'\overline{\boldsymbol{x}})\beta_j$$

  - Man beachte: DC und ME hängen von den X ab! Weshalb man X-Werte setzen muss (hier: das Mittel, MEM und DCM)

- DC oder ME?
  - X kategorial: verwende DC
  - X metrisch: verwende ME
    - Long/Freese plädieren auch hier für DC, da der ME nur eine Näherung ist (s. Graphik)



Quelle: Long/Freese (2006) S. 169

Average Marginal Effect (AME)

- Marginaleffekte sind nicht eindeutig
  - Je nachdem, an welcher Stelle man sie berechnet, fallen sie anders aus
- Average marginal effects
  - Seit einigen Jahren die Lehrmeinung: um eine eindeutige Kennzahl zu erhalten, bilde den Durchschnitt der ME in der Stichprobe

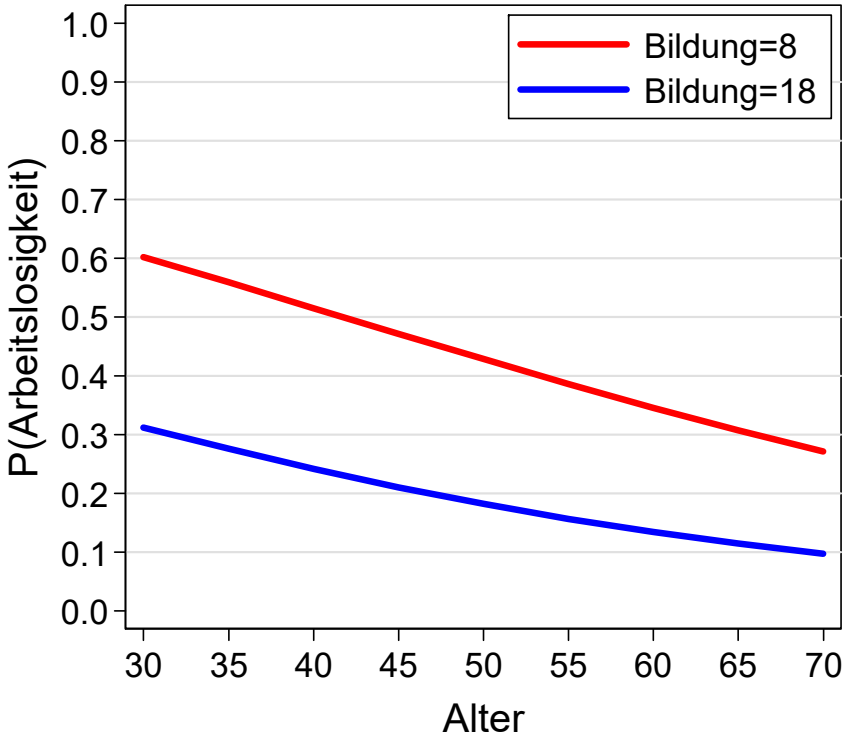
$$AME_j = \frac{1}{N}\sum_{i=1}^{N} ME_{ij} = \frac{1}{N}\sum_{i=1}^{N} f\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{x_i}\right)\hat{\beta}_j$$

  - In Stata: `margins, dydx(*)`
    - Für kategoriale X berechnet Stata ADC
  - Interpretation: P(Y=1) erhöht sich – im Mittel aller Beobachtungen der vorliegenden Stichprobe – um AME Prozentpunkte, wenn sich X um eine Einheit (bzw. marginal) erhöht

STATA-Bsp.: Vorhergesagte Wahrscheinlichkeiten

Alter und Bildung

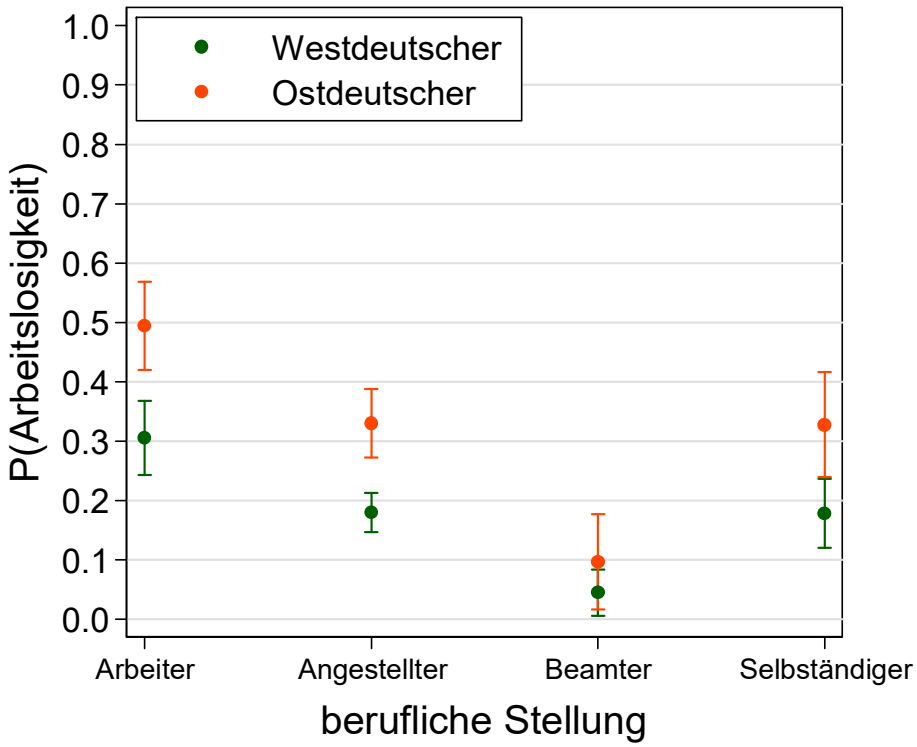
```
logit arblos bild alter frau ost

margins, at(alter=(30(5)70) bild=(8 18))
marginsplot, noci
```



Dies sind Profile-Plots

Berufliche Stellung und Ost

```
logit arblos bild alter frau  ///
                i.ost i.beruf
margins beruf#ost
marginsplot, plotopts(connect(i))
```

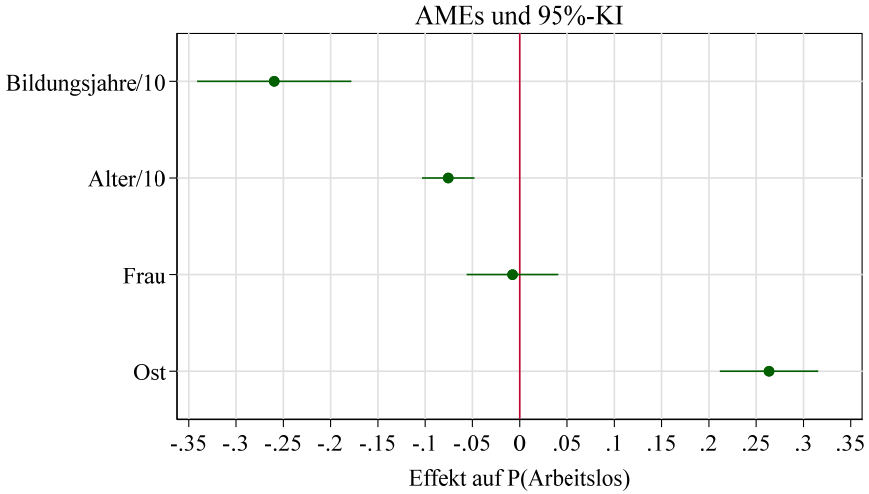


Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

STATA-Beispiel: AMEs

```
. logit arblos bild alter i.frau i.ost
. margins, dydx(*)

Average marginal effects                  Number of obs   =       1329

------------------------------------------------------------------------------
             |                 Delta-method
             |      dy/dx      Std. Err.       z    P>|z|
-------------+----------------------------------------------------------------
        bild |    -0.0260                     0.0042    -6.24   0.000
       alter |    -0.0076   AME              0.0014    -5.34   0.000
      1.frau |    -0.0077                     0.0247    -0.31   0.757
       1.ost |     0.2636   ADC              0.0266     9.91   0.000
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.
```

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do



Alter und Bildung sind hier durch 10 geteilt, damit die Skalierungen vergleichbar sind.

STATA-Beispiel: Vergleich mit LPM

	Logit	LPM
Bildung	AME -0,0260	-0,0248
Alter	AME -0,0076	-0,0074
Frau	ADC -0,0077	-0,0092
Ost	ADC 0,2636	0,2633

Das LPM approximiert die AMEs sehr gut

- Das Logit-Modell: ein umständlicher Weg, um im Endeffekt nur wieder die Effekte des LPM zu erhalten?
- Simulationen (Best/Wolf 2012) zeigen allerdings, dass das LPM bei nicht-normalverteilten (stark schiefen) X-Variablen zu verzerrten Ergebnissen führen kann
- Dennoch spricht Vieles (Interaktionseffekte, Gruppenvergleich, s.u.) dafür, das LPM zu verwenden. Deshalb erlebt das LPM in den letzten Jahren wieder eine Renaissance (Breen et al. 2018)

Interaktionseffekte

- Im linearen Modell $$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 (x \cdot z) + \varepsilon$$

$$\frac{\partial^2 E(y)}{\partial x \partial z} = \beta_3$$

  - Inteff ist konstant und gleich dem Koeffizienten des Produktterms
- Im Logit $$P(Y = 1) = F[\beta_0 + \beta_1 x + \beta_2 z + \beta_3 (x \cdot z)] = F(u)$$

$$\frac{\partial^2 P(Y = 1)}{\partial x \partial z} = \frac{\partial\, (\beta_1 + \beta_3 z) F'(u)}{\partial z} = \beta_3 F'(u) + (\beta_1 + \beta_3 z) F''(u)(\beta_2 + \beta_3 x)$$

  - Inteff $\neq \beta_3$
  - Inteff ist nicht konstant, sondern hängt von allen Kovariaten ab
    - Das Vorzeichen des Inteff kann sich von β_3 unterscheiden (F‘‘(u) kann kleiner null sein)
    - Je nach Kovariatenwert kann das Vorzeichen des Inteff anders sein (wir haben also positive und negative Inteff in den selben Daten)
  - Die Signifikanz des Inteff kann nicht durch Test von β_3 geprüft werden
  - Bereits das Modell ohne Produktterm modelliert einen Inteff

$$\left.\frac{\partial^2 P(Y = 1)}{\partial x \partial z}\right|_{\beta_3 = 0} = \beta_1 F''(u) \beta_2$$

Interaktionseffekte

Fazit: Produktterme in nicht-linearen Modellen erhöhen die Komplexität enorm. Um zu verstehen, was man hier eigentlich modelliert, muss man Profile-Plots zur Hilfe nehmen (mehr dazu findet man in Ai/Norton 2003).

Beispiel: Polynomregression mit Alter und Alter^2 und Interaktion mit „Ost“

```
. logit arblos bild frau i.ost##(c.alter c.alter#c.alter)

------------------------------------------------------------------------------
             arblos |      Coef.   Std. Err.      z    P>|z|
--------------------+---------------------------------------------------------
               bild |    -0.1336     0.0220    -6.08   0.000
               frau |     0.0010     0.1260     0.01   0.993
             1.ost  |     1.8372     3.3693     0.55   0.586
              alter |    -0.0761     0.1032    -0.74   0.461
    c.alter#c.alter |     0.0002     0.0012     0.15   0.880
      1.ost#c.alter |    -0.0786     0.1540    -0.51   0.610
1.ost#c.alter#c.alter |     0.0014     0.0017     0.83   0.404
              _cons |     3.5295     2.2561     1.56   0.118
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

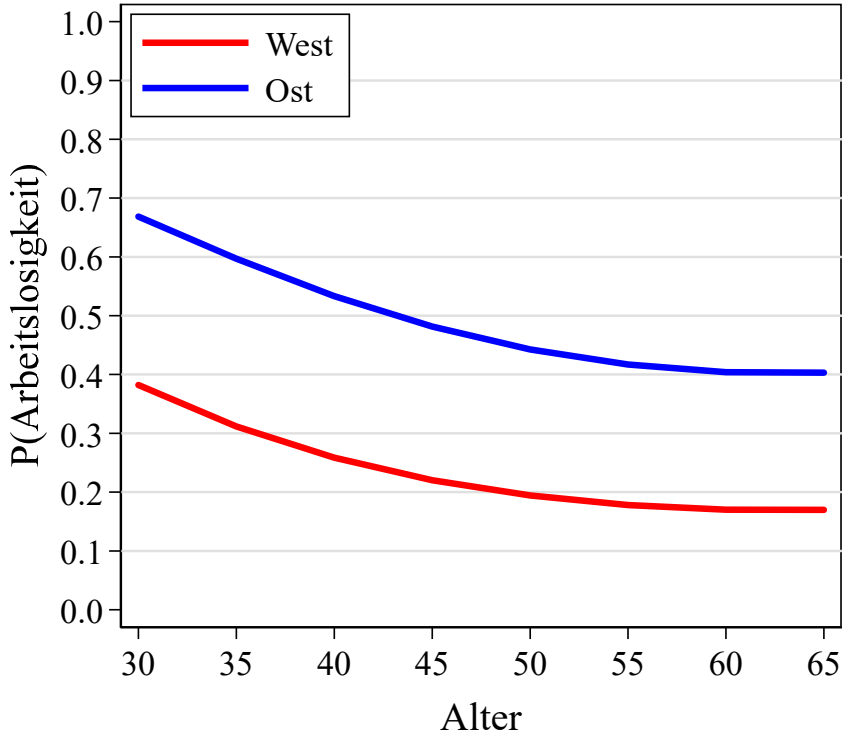
STATA-Beispiel: Interaktionseffekte

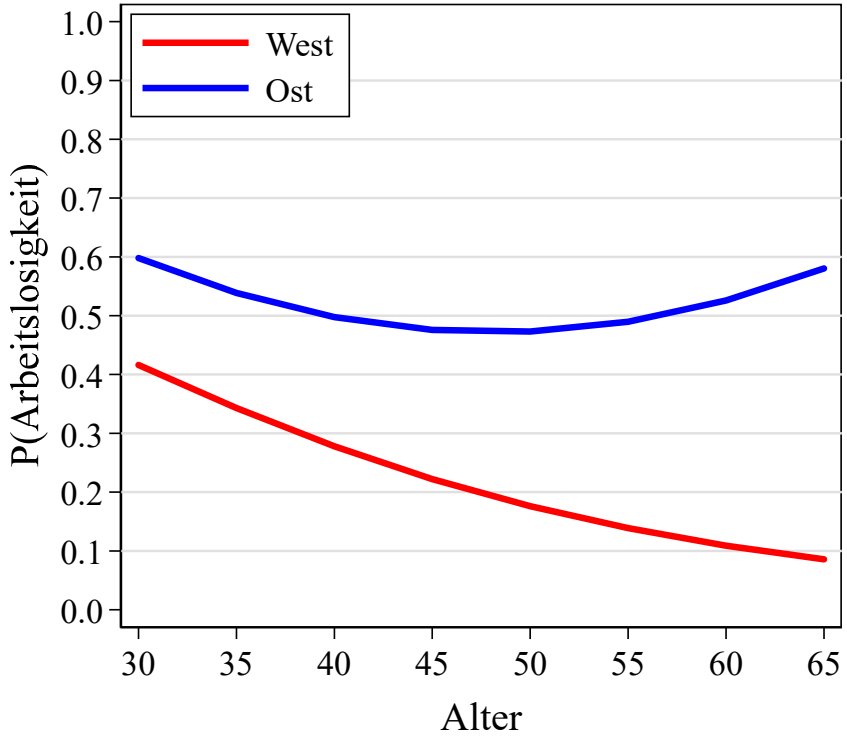
Profile-Plot
Polynomregression mit Alter²

```
logit arblos bild frau i.ost       ///
        c.alter c.alter#c.alter
margins ost, at(alter=(30(5)65))
marginsplot,   noci
```

Profile-Plot
Polynomreg. + Interaktion mit Ost

```
logit arblos bild frau             ///
    i.ost##(c.alter c.alter#c.alter)
margins ost, at(alter=(30(5)65))
marginsplot,   noci
```





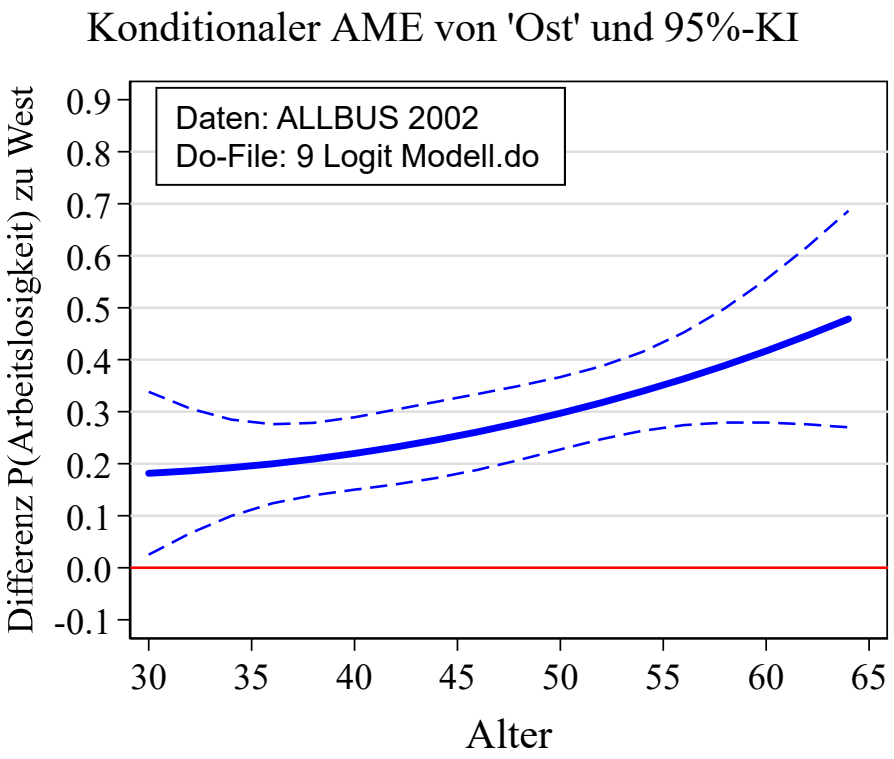
Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

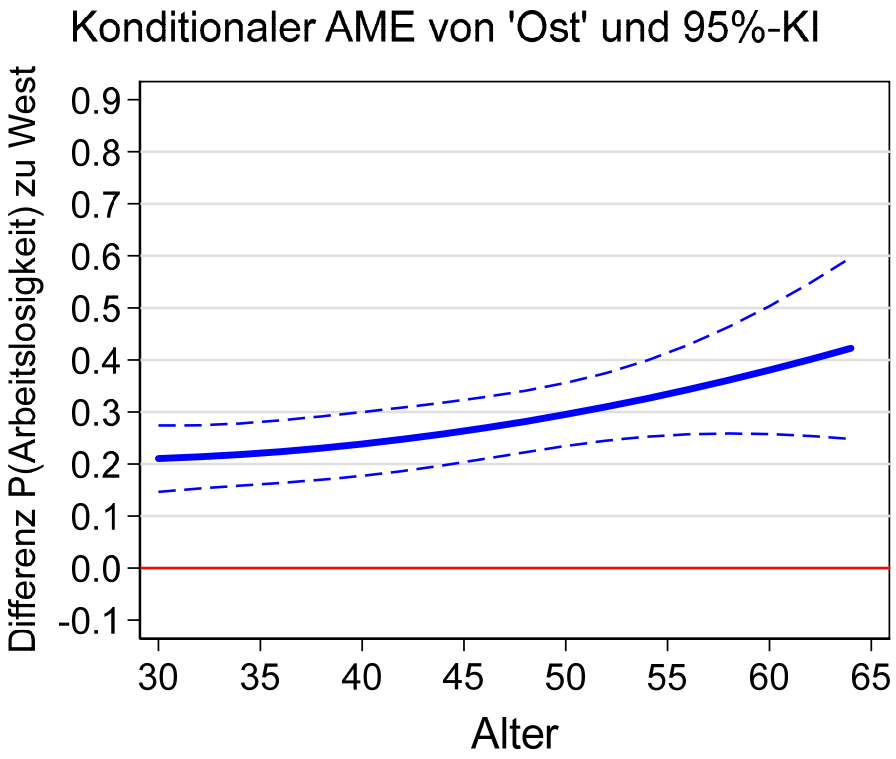
Interaktionseffekte

Conditional AME-Plot

```
logit arblos bild frau                    ///
     i.ost##(c.alter c.alter#c.alter)
margins, at(alter=(30(2)70)) dydx(ost)
marginsplot, recast(line) recastci(rline)
```

Wem das alles zu kompliziert ist:
Das LPM ist eine gute Alternative und bekanntlich ist dort die Interpretation von Interaktionseffekten deutlich einfacher!





NLPMs als latentes Variablen Modell

Y^* sei eine latente Variable, die wir linear modellieren

$$y^* = a + bx + \sigma\varepsilon$$

- wobei ε ein Fehlerterm ist und σ ein Skalierungsfaktor

– Das binäre, beobachtete Y ergibt sich aus folgendem Schwellenwertmodell

$$y = \begin{cases} 1, & y^* > 0 \\ 0, & y^* \leq 0 \end{cases}$$

– Daraus ergibt sich $P(Y = 1)$ als

$$P(Y = 1) = P(y^* > 0) = P(a + bx + \sigma\varepsilon > 0) =$$
$$P(\sigma\varepsilon > -(a + bx)) = P(\varepsilon < \alpha + \beta x) = F(\alpha + \beta x)$$

- wobei $\alpha = \frac{a}{\sigma}$ und $\beta = \frac{b}{\sigma}$

– Je nach Annahme über $F(.)$ kommt man zu Probit oder Logit

- Die Varianz von ϵ ist dabei fest (s.o.)
- Deshalb benötigt das Modell den Skalierungsfaktor σ

Das Skalierungsproblem

- Die Abhängigkeit der Koeffizienten vom (unbekannten) Skalierungsfaktor ist normalerweise kein Problem, da die Wahrscheinlichkeiten trotzdem eindeutig schätzbar sind
- Der Skalierungsfaktor wird allerdings zum Problem, wenn man vergleichen will (Überblick bei Breen et al. 2018)
  1. Genestete Modelle (Modellvergleich)
  2. Modelle über Gruppen, Kohorten, Länder, etc. (Gruppenvergleich)
- Die Skalierungsfaktoren unterscheiden sich zwischen Modellen/ Gruppen (unterschiedliche unbeobachtete Heterogenität)
  - Die geschätzten Logit-Koeffizienten unterscheiden sich allein deshalb (gilt dann natürlich analog für Odds-Ratio Effekte)
  - Vergleiche von Koeffizienten zwischen Modellen/Gruppen sind somit von unbeobachteter Heterogenität konfundiert!

Modellvergleich

Erweiterung eines Logit-Modells um Z. Dann:

1. Die Erklärungskraft des linearen Prädiktors steigt. Da die Varianz von ε festgelegt ist, muss σ sinken. Deshalb werden die Koeffizienten größer (Koeffizienten-Inflation, KIF)
   - $\sigma_2 < \sigma_1$, weshalb die Logit-Koeffizienten alle (!) größer werden
2. Der Effekt von X verändert sich (falls Confounding oder Mediation)

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \frac{a}{\sigma_1} + \frac{b}{\sigma_1}x$$

Generell gilt: die Veränderung des Effekts von X wird in Richtung 0 verzerrt

Deshalb ist die soz. Literatur voll von Artefakt-Befunden, dass Drittvariablen keine Confounder/Mediatoren sind

$$\ln\left(\frac{P(Y=1)}{P(Y=0)}\right) = \frac{a}{\sigma_2} + \frac{b}{\sigma_2}x + \frac{c}{\sigma_2}z$$

	Modell (1)	Modell (2)
IQ	0,80	0,99
Mädchen		2,00
Konstante	-0,01	-1,01

Simulation von Mood (2010):
Logit Modell für die Wahrscheinlichkeit des Übergangs zur Universität.
IQ und Geschlecht sind unkorreliert!

Modellvergleich

- Es existieren diverse Lösungsvorschläge
- Bei kleinem R^2 ist die KIF gering (Best/Wolf 2012)
  - Allerdings ist es kein gutes Argument, mit der schlechten Erklärungskraft des Modells zu argumentieren
- Karlson/Holm/Breen (2012): KHB-Verfahren
  - Erweitere die Regression um das Residuum der Regression von Z auf X. Damit hat das reduzierte Modell die gleiche Erklärungskraft und es kommt zu keiner KIF

$$Logit = \alpha + \beta x + \gamma \hat{\varepsilon}_{z|x}$$

  - Danach noch AME, wg. besserer Interpretierbarkeit
  - Stata-Ado: khb
- AMEs
  - Simulationen (z.B. Best/Wolf 2012) zeigen, dass die AMEs nicht von der KIF betroffen sind
- Das LPM ist vom Skalierungsproblem nicht betroffen

Gruppenvergleich

- Gruppenvergleich
  - Die Skalierungsfaktoren (σ) sind im Normalfall in den Gruppen unterschiedlich und man kennt sie nicht
  - Logit-Effekte werden allein deshalb unterschiedlich ausfallen
  - Gruppenvergleiche von NLPM-Koeffizienten beruhen auf nicht überprüfbaren Annahmen (Breen et al. 2018)
- Lösungsvorschläge
  - Heute werden oft die AMEs auch zum Gruppenvergleich verwendet
    - Auspurg/Hinz (2011) zeigen bei einer Analyse zur Veränderung der Bildungsungleichheit über Geburtskohorten, dass man mit OR und AME zu unterschiedlichen Schlussfolgerungen kommt. Sie empfehlen die Verwendung der AMEs.
    - Aber auch die AMEs scheinen nur bedingt vergleichbar zu sein (?)
  - Eine gangbare Lösung scheint das LPM zu sein (?) (Breen et al. 2018)

Logit vs. Probit

	Logit	Probit
Bildungsjahre	-0.131***	-0.077***
	(-5.96)	(-6.06)
Alter	-0.038***	-0.023***
	(-5.15)	(-5.32)
Frau	-0.039	-0.023
	(-0.31)	(-0.30)
Ostdeutscher	1.221***	0.747***
	(9.68)	(9.83)
Konstante	2.222***	1.328***
N	1329	1329
R^2	0.088	0.089

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Logit und Probit liefern fast immer praktisch identische Ergebnisse. Die Logit-Koeffizienten sind in etwa um den Faktor 1,7 größer (s. Long, 1997, S. 48):

bild:	-0,131 / -0,077 =	1,70
alter:	-0,038 / -0,023 =	1,65
frau:	-0,039 / -0,023 =	1,70
ost:	1,221 / 0,747 =	1,63

Daten: ALLBUS 2002
Do-File: 9 Logit Modell.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 11:
Multinomiales Logit

Regression für multinomiales Y

- Kategoriale aV mit mehr als zwei Ausprägungen
- Multinomiale logistische Regression
  - Y sei multinomial mit Ausprägungen m = 1, …, J. Ein geeignetes nicht-lineares Wahrscheinlichkeitsmodell zur Modellierung ist

$$P(y = m) = \frac{\exp(\boldsymbol{\beta'_m x})}{\sum_{k=1}^{J} \exp(\boldsymbol{\beta'_k x})}$$

  - Alle J Koeffizientenvektoren sind nicht identifiziert, weshalb man einen auf 0 setzen muss. Für J = 3 und 1 als Referenzkategorie:

$$P(y = 1) = \frac{1}{1 + \exp(\boldsymbol{\beta'_2 x}) + \exp(\boldsymbol{\beta'_3 x})}$$

$$P(y = 2) = \frac{\exp(\boldsymbol{\beta'_2 x})}{1 + \exp(\boldsymbol{\beta'_2 x}) + \exp(\boldsymbol{\beta'_3 x})}$$

$$P(y = 3) = \frac{\exp(\boldsymbol{\beta'_3 x})}{1 + \exp(\boldsymbol{\beta'_2 x}) + \exp(\boldsymbol{\beta'_3 x})}$$

Multinomiales Logit

Anmerkungen:

- Es gilt: $\sum_m P(y = m) = 1$
- Die Wahl der Referenzkategorie ist beliebig. In Stata wird automatisch die Modal-Kategorie gewählt.
- Die Koeffizienten sind der Logit-Effekt jeweils im Vergleich zur Referenzkategorie. Je nach gewählter Referenzkategorie fallen also die Koeffizienten anders aus. Das Modell ändert sich aber nicht, es handelt sich nur um ein Reparametrisierung.
- Für J = 2 erhält man das binäre Logit als Spezialfall.
- Das multinomiale Logit kann auch aus einem latente Variablen Modell hergeleitet werden. Das Problem mit dem Skalierungsfaktor σ existiert hier also auch!
- Schätzung mit ML: $L = \prod P_i$

STATA Beispiel: Wahlabsicht (Sonntagsfrage)

```
. generate partei = v521
. recode   partei 1=1 2=2 3=3 4=4 6=5  5 7 8=.
. label define partlbl 1 "CDU" 2 "SPD" 3 "FDP" 4 "Grüne" 5 "PDS"
. label value partei partlbl

. tab partei,m

     partei |      Freq.     Percent        Cum.
------------+-----------------------------------
        CDU |        722       25.60       25.60
        SPD |        660       23.40       49.01
        FDP |        284       10.07       59.08
      Grüne |        201        7.13       66.21
        PDS |        161        5.71       71.91
          . |        792       28.09      100.00
------------+-----------------------------------
      Total |      2,820      100.00
```

„Missing“: Republikaner, andere Partei, Nichtwähler, Unentschlossene, Verweigerer, Ausländer

Daten: ALLBUS 2002
Do-File: 0 Datenaufbereitung.do

STATA Beispiel: Wahlabsicht (Sonntagsfrage)

```
. mlogit partei alter bild ost, base(1)        //CDU als Referenzkategorie

Multinomial logistic regression                 Number of obs   =       2007
                                                LR chi2(12)     =     307.66
                                                Prob > chi2     =     0.0000
Log likelihood = -2729.6341                     Pseudo R2       =     0.0533
------------------------------------------------------------------------------
      partei |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
CDU          |  (base outcome)
-------------+----------------------------------------------------------------
SPD          |
       alter |    -0.0069     0.0032    -2.14   0.032      -0.0133     -0.0006
        bild |    -0.0051     0.0198    -0.26   0.797      -0.0439      0.0338
         ost |     0.0995     0.1191     0.84   0.403      -0.1339      0.3329
-------------+----------------------------------------------------------------
FDP          |
       alter |    -0.0001     0.0042    -0.02   0.982      -0.0084      0.0082
        bild |     0.0604     0.0241     2.50   0.012       0.0131      0.1077
         ost |    -0.2284     0.1624    -1.41   0.160      -0.5467      0.0899
-------------+----------------------------------------------------------------
Grüne        |
       alter |    -0.0320     0.0054    -5.96   0.000      -0.0425     -0.0215
        bild |     0.1586     0.0270     5.88   0.000       0.1057      0.2115
         ost |    -0.7410     0.2138    -3.47   0.001      -1.1600     -0.3221
-------------+----------------------------------------------------------------
PDS          |
       alter |    -0.0112     0.0057    -1.97   0.049      -0.0223     -0.0000
        bild |     0.0763     0.0310     2.46   0.014       0.0156      0.1370
         ost |     2.4515     0.2247    10.91   0.000       2.0111      2.8918
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Test der Signifikanz der Variablen

- Haben die drei Variablen einen signifikanten Einfluss auf die Wahlabsicht?
  - Die Signifikanztests der einzelnen Logit-Koeffizienten können zu widersprüchlichen Ergebnissen führen.
  - Wir brauchen einen LR-Test

```
. mlogtest, lr                  //funktioniert nur mit "SPost"

**** Likelihood-ratio tests for independent variables (N=2007)

 Ho: All coefficients associated with given variable(s) are 0.

             |       chi2   df   P>chi2
-------------+--------------------------
       alter |     42.040    4    0.000
        bild |     45.950    4    0.000
         ost |    209.727    4    0.000
-----------------------------------------
```

Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Test zum Kombinieren von Kategorien

- Kann man Kategorien zusammenfassen?
  - Wenn die Logit-Effekte in einer Kategorie alle Null sind, so ist die Kategorie von der Referenzkategorie nicht unterscheidbar und die beiden Kategorien können zusammengefasst werden
  - Hierfür gibt es einen LR-Test

```
. mlogtest, lrcomb                    //funktioniert nur mit "SPost"

**** LR tests for combining alternatives (N=2007)

 Ho: All coefficients except intercepts associated with a given pair
     of alternatives are 0 (i.e., alternatives can be collapsed).

Alternatives tested|     chi2   df   P>chi2
-------------------+------------------------
     SPD-     FDP  |   12.383    3    0.006
     SPD-   Grüne  |   80.599    3    0.000
     SPD-     PDS  |  154.159    3    0.000
     SPD-     CDU  |    5.429    3    0.143
     FDP-   Grüne  |   47.655    3    0.000
     FDP-     PDS  |  151.772    3    0.000
     FDP-     CDU  |    8.196    3    0.042
   Grüne-     PDS  |  180.703    3    0.000
   Grüne-     CDU  |   93.209    3    0.000
     PDS-     CDU  |  173.940    3    0.000
-----------------------------------------
```

Interessanterweise scheint es, als ob wir CDU und SPD zusammenfassen könnten. Das werden wir aber aus inhaltlichen Gründen nicht tun!

Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Unabhängigkeit von irrelevanten Alternativen

- Das multinomiale Logit impliziert eine spezielle Annahme
  - Unabhängigkeit von irrelevanten Alternativen (IIA)

$$\frac{P(y=m)}{P(y=k)} = \exp\{\boldsymbol{x}(\boldsymbol{\beta}_m - \boldsymbol{\beta}_k)\}$$

  - Das Odds ist also unabhängig von den anderen Alternativen
  - In unserem Beispiel dürfte sich das Odds SPD/CDU nicht verändern, wenn eine neue Partei aufträte
    - Wenn z.B. die „Piraten“ hinzukämen, müssten proportional identische Teile der SPD- und CDU-Wähler zu den Piraten wechseln, damit die Odds SPD/CDU gleich bleiben und die IIA gilt
- Es gibt viele Tests der IIA
  - Grundidee: Man vergleicht die Schätzer des vollen Modells mit einem restringierten Modell, in dem eine Kategorie fehlt. Unterscheiden sich die Schätzer, so ist die IIA verletzt.

Unabhängigkeit von irrelevanten Alternativen

- Probleme der IIA-Tests
  - Long/Freese zeigen, dass diese Tests in finiten Stichproben nicht funktionieren
    - Hausman-McFadden Test liefert oft negative Werte
    - Small-Hsiao hängt ab von der Zufallsaufteilung der Stichprobe
  - Also bleiben eigentlich nur Plausibilitätsüberlegungen:
    „Sind die Kategorien klar unterscheidbare Alternativen?“
- Abhilfe wenn IIA verletzt ist
  - Nested Logit (`nlogit`)

```
. mlogtest, suest base      //funktioniert nur mit "SPost"

**** suest-based Hausman tests of IIA assumption (N=2007)

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other
alternatives.

 Omitted |      chi2   df   P>chi2   evidence
---------+-----------------------------------
     SPD |    11.430   12    0.493   for Ho
     FDP |    13.196   12    0.355   for Ho
   Grüne |     5.381   12    0.944   for Ho
     PDS |    13.095   12    0.362   for Ho
     CDU |     5.041   12    0.957   for Ho
-------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Interpretation

- Vorzeicheninterpretation
  - Das Vorzeichen von β gibt Richtung des Logit- und Odds-Effektes
  - **Dieses Vorzeichen ist aber nicht unbedingt identisch mit dem Vorzeichen des Wahrscheinlichkeitseffektes!**
  - Das wird gern übersehen und kann zu falschen Schlussfolgerungen führen. Ein Beispiel:
    - Im Westen wählen jeweils 30% SPD und CDU
    - Im Osten wählen 20% SPD und 10% CDU
    - Odds(Ost) = 0,2/0,1 = 2 Odds(West) = 0,3/0,3 = 1
    - Die Odds-Ratio ist also 2. Es wäre nun aber eine krasse Fehlinterpretation, wenn man schlussfolgern würde, die Ossis wählen doppelt so oft SPD!
- Odds Interpretation
  - e^{β} ist die OR für Variable X_j für P(y=m) / P(y=1)
  - Die Odds Interpretation ist schwer zu verstehen

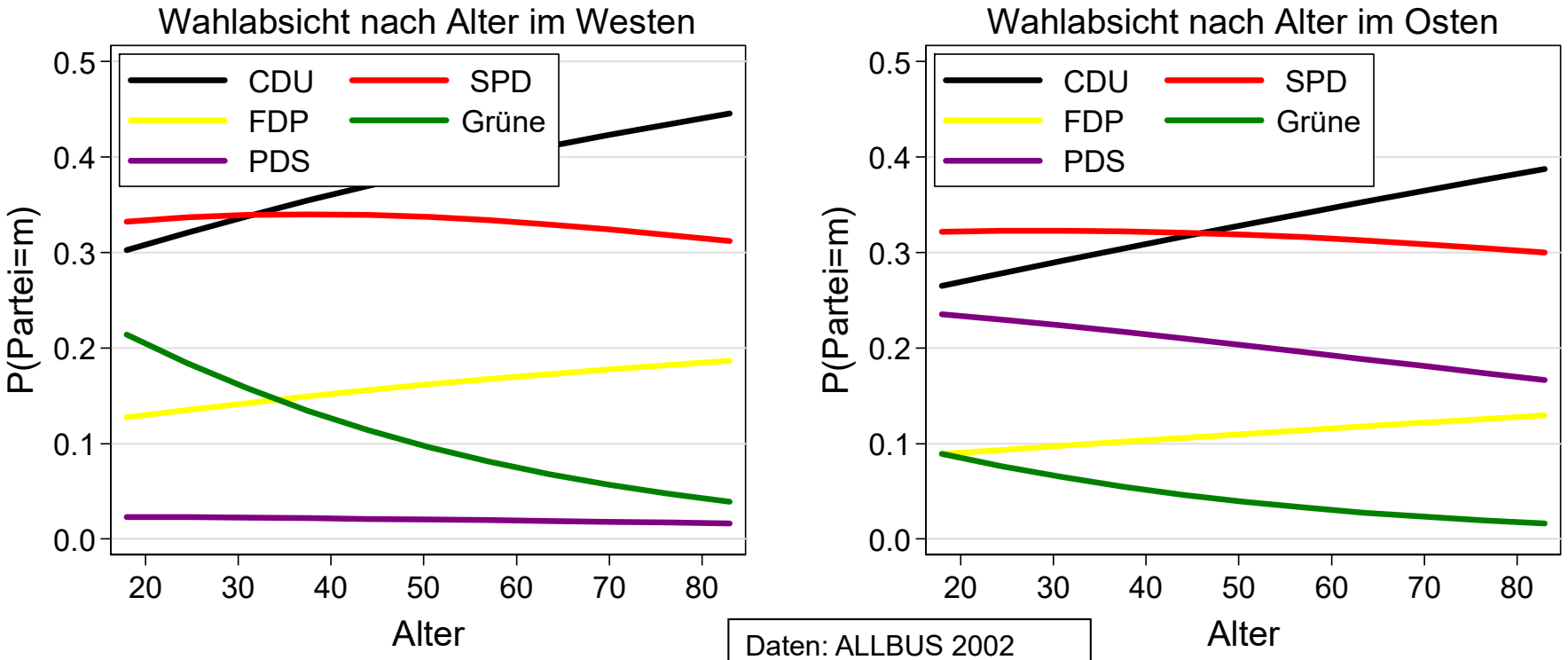
Bsp. Ost(Grüne):
ß = -0,7410, OR = 0,48
Die Odds Grüne/CDU sind im Osten um die Hälfte kleiner.

Vorhergesagte Wahrscheinlichkeiten (Profile-Plots)

Alterseffekt nach Ost/West

```
mlogit partei alter bild ost, base(1)
prgen alter, from(18) to(83) x(ost 1) generate(o)
twoway connected op1 op2 op3 op4 op5 ox
```

Man beachte: Obwohl dieses Modell keine Produktterme enthält, sehen wir deutliche Interaktionen. Z.B. ist der Alterseffekt auf die Grünen im Westen stärker als im Osten.



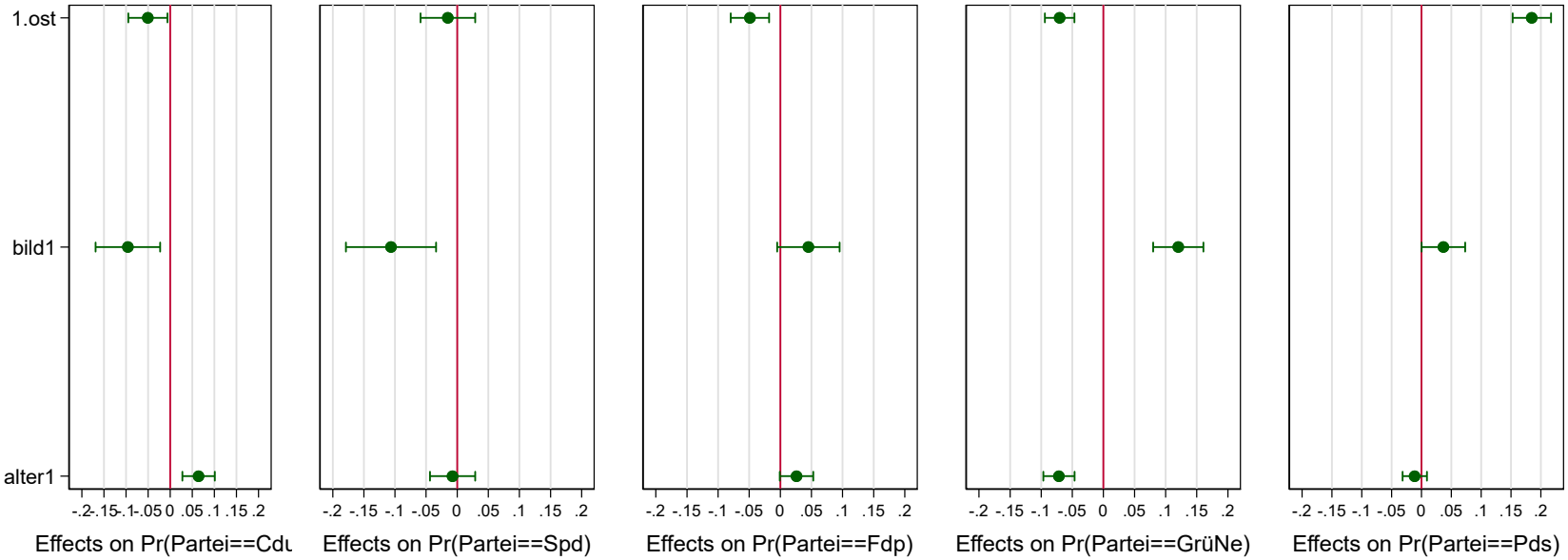
Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Wahrscheinlichkeitsinterpretation

- Die Marginal Effekte (ME) im multinomialen Logit sind

$$\frac{\partial P_m}{\partial x_j} = P_m\left(\beta_{m,j} - \sum_{k=1}^{J}\beta_{k,j}P_k\right), \qquad \text{wobei } P_m = P(y=m)$$

  - Der ME kombiniert alle Koeffizienten der Variable X_j. Er hängt von allen anderen Kovariatenwerten ab.
  - Es ist sogar möglich, dass der ME sein Vorzeichen wechselt innerhalb des Wertebereichs von X_j!
  - Deshalb verwendet man sinnvollerweise auch hier AMEs



AMEs für unser Beispiel

Daten: ALLBUS 2002
Do-File: 10 MLogit.do

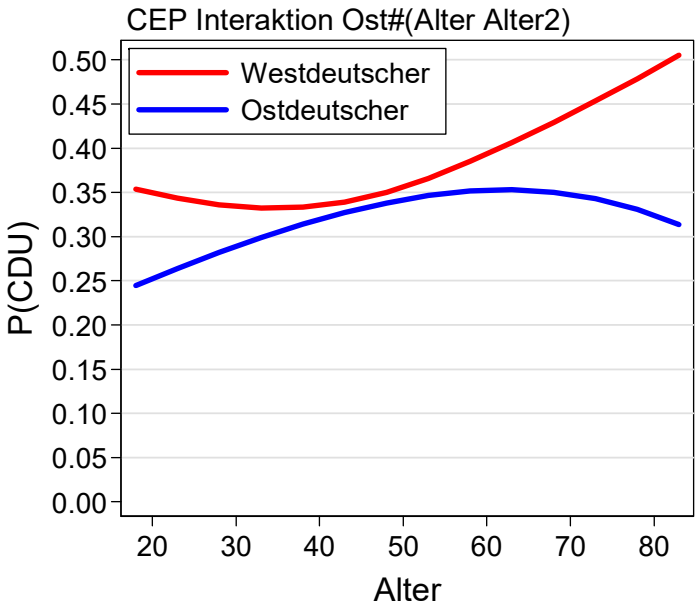
Ein Modell mit Interaktionseffekten!

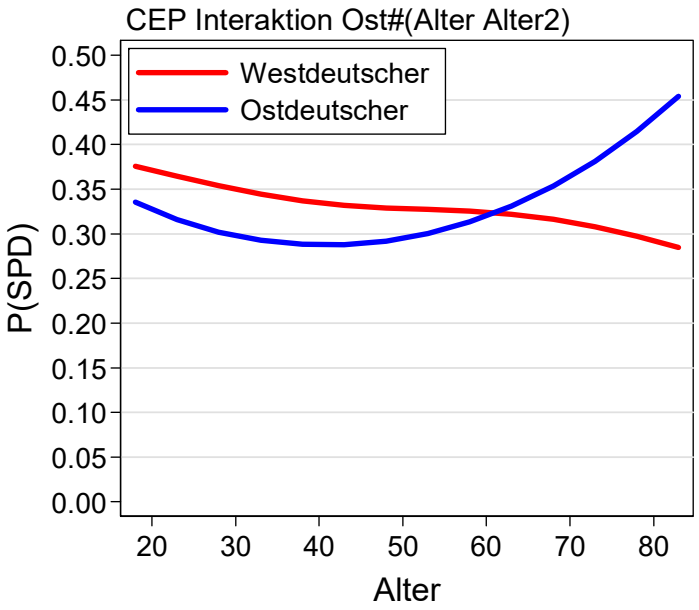
- Wir schätzen ein komplexes Modell
  - Quadratisches Polynom in Alter
  - Interaktion mit Ost
  - `mlogit partei bild i.ost##(c.alter c.alter#c.alter)`

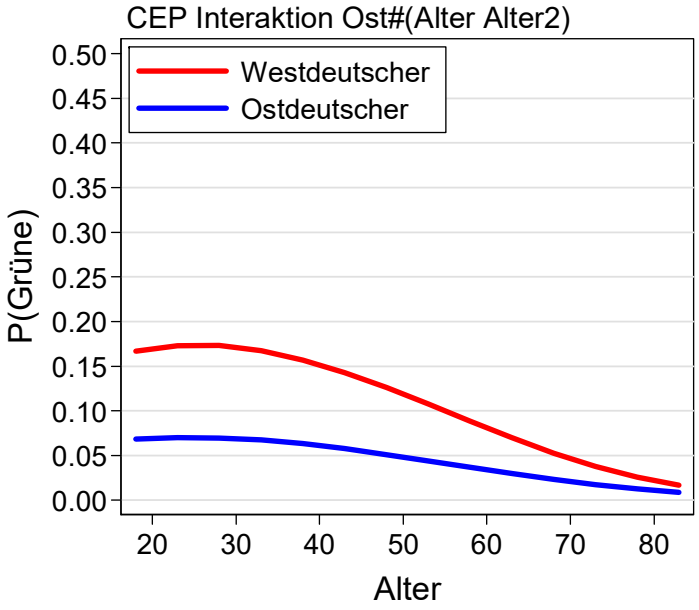
- Interpretation
  - Eine Regressionstabelle hilft hier nicht mehr weiter
    - Regressionstabellen komplexer nicht-linearer Modelle sind Verschwendung von Papier und Zeit
  - Hier muss man Regressionsplots einsetzen
  - Auf der folgenden Folie sind illustrativ die Alterseffekte auf die Parteien (nicht: FDP) für Ost und West getrennt abgebildet
    - Das Modell erlaubt offensichtlich gänzlich unterschiedliche Alterseffekte!

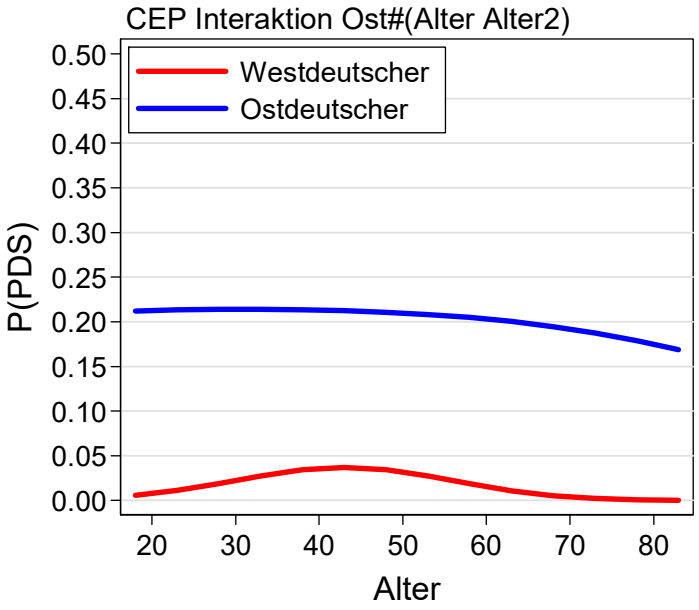
Daten: ALLBUS 2002
Do-File: 10 MLogit.do

Ein Modell mit Interaktionseffekten!









LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 12:
Ordinales Logit

Regression für ordinales Y

- Ordinale aV werden häufig mit OLS analysiert
  - Annahme: Distanzen zwischen den Kategorien sind gleich
  - Diese Annahme ist häufig nicht plausibel, weshalb man ein Regressionsverfahren für ordinale aV verwenden sollte
  - Dennoch liefert das lineare Modell meist ähnliche Ergebnisse (s.u.)
- Ordinal Response Model (ORM)
  - Y sei ordinal mit Ausprägungen m = 1, …, J. Das latente Variablen Modell lautet:

$$y^* = \boldsymbol{\beta}'\boldsymbol{x} + \varepsilon$$

  - Das Schwellenwertmodell lautet nun mit J+1 Schwellen τ

$$y = m, \quad \text{falls} \quad \tau_{m-1} \leq y^* < \tau_m$$

  - wobei $\tau_0 = -\infty$ und $\tau_J = \infty$. Es sind also J-1 Schwellen zu schätzen. Damit gilt:

$$P(y = m) = P(\tau_{m-1} \leq y^* < \tau_m)$$

$$\Longrightarrow \quad P(y = m) = F(\tau_m - \boldsymbol{\beta}'\boldsymbol{x}) - F(\tau_{m-1} - \boldsymbol{\beta}'\boldsymbol{x})$$

Ordinales Logit/Probit

- Die Wahl der Verteilungsfunktion F(.) komplettiert das Modell
  - Standard-Normalverteilung: ordinales Probit
  - Standard-Logistische-Verteilung: ordinales Logit
  - z.B. ordinales Logit für J=3:

$$P(y = 1) = \Lambda(\tau_1 - \boldsymbol{\beta}'\boldsymbol{x})$$

$$P(y = 2) = \Lambda(\tau_2 - \boldsymbol{\beta}'\boldsymbol{x}) - \Lambda(\tau_1 - \boldsymbol{\beta}'\boldsymbol{x})$$

$$P(y = 3) = 1 - \Lambda(\tau_2 - \boldsymbol{\beta}'\boldsymbol{x})$$

  - Identifikation: es ist nicht möglich J - 1 Schwellen und eine Konstante zu schätzen. Stata setzt deshalb die Konstante auf 0.
  - Für J = 2 erhält man die binären Modelle als Spezialfall.
  - In der Ableitung haben wir den Skalierungsfaktor σ ignoriert. Das Problem existiert hier aber natürlich auch!
  - Schätzung mit ML: $L = \prod P_i$

STATA Beispiel: „rechte“ politische Einstellung

```
. recode rechts 1/3=1  4/6=2 7/10=3
. label define relbl 1 "Links" 2 "Mitte" 3 "Rechts"
. label value rechts relbl
. tab rechts, m

     rechts |      Freq.     Percent        Cum.
------------+-----------------------------------
      Links |        584       20.71       20.71
      Mitte |      1,560       55.32       76.03
     Rechts |        517       18.33       94.36
          . |        159        5.64      100.00
------------+-----------------------------------
      Total |      2,820      100.00
```

Die aV wird hier gruppiert, um die Outputs übersichtlicher zu machen. Das sollte man bei ernsthaften Analysen nicht tun!

Daten: ALLBUS 2002
Do-File: 11 OLogit.do

STATA Beispiel: „rechte“ politische Einstellung

```
. replace   eink = eink/1000      //bessere Skalierung
. ologit   rechts alter bild eink frau ost

Iteration 0:   log likelihood = -2092.9781
Iteration 1:   log likelihood = -2055.7797
Iteration 2:   log likelihood = -2055.5577
Iteration 3:   log likelihood = -2055.5576

Ordered logistic regression                     Number of obs   =       2137
                                                LR chi2(5)      =      74.84
                                                Prob > chi2     =     0.0000
Log likelihood = -2055.5576                     Pseudo R2       =     0.0179

------------------------------------------------------------------------------
      rechts |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       alter |      0.0068     0.0026     2.63   0.009       0.0017      0.0118
        bild |     -0.0358     0.0164    -2.18   0.029      -0.0680     -0.0036
        eink |      0.0448     0.0403     1.11   0.266      -0.0342      0.1239
        frau |     -0.3677     0.0886    -4.15   0.000      -0.5415     -0.1940
         ost |     -0.5678     0.0910    -6.24   0.000      -0.7461     -0.3894
-------------+----------------------------------------------------------------
       /cut1 |     -1.6625     0.2562                      -2.1647     -1.1604
       /cut2 |      1.0128     0.2542                       0.5145      1.5111
------------------------------------------------------------------------------
```

Daten: ALLBUS 2002
Do-File: 11 OLogit.do

Ein Modellvergleich

	OLogit	OProbit	OLS
Alter	0.007**	0.004**	0.002**
	(2.63)	(2.69)	(2.70)
Bildungsjahre	-0.036*	-0.019*	-0.011*
	(-2.18)	(-2.08)	(-2.06)
Einkommen / 1000	0.045	0.026	0.015
	(1.11)	(1.09)	(1.09)
Frau	-0.368***	-0.216***	-0.123***
	(-4.15)	(-4.22)	(-4.23)
Ostdeutscher	-0.568***	-0.329***	-0.188***
	(-6.24)	(-6.25)	(-6.29)
Konstante			2.094***
			(25.61)
Cutpoint1	-1.663***	-0.977***	
	(-6.49)	(-6.75)	
Cutpoint2	1.013***	0.642***	
	(3.98)	(4.44)	
N	2137	2137	2137
R^2			0.035
pseudo R^2	0.018	0.018	

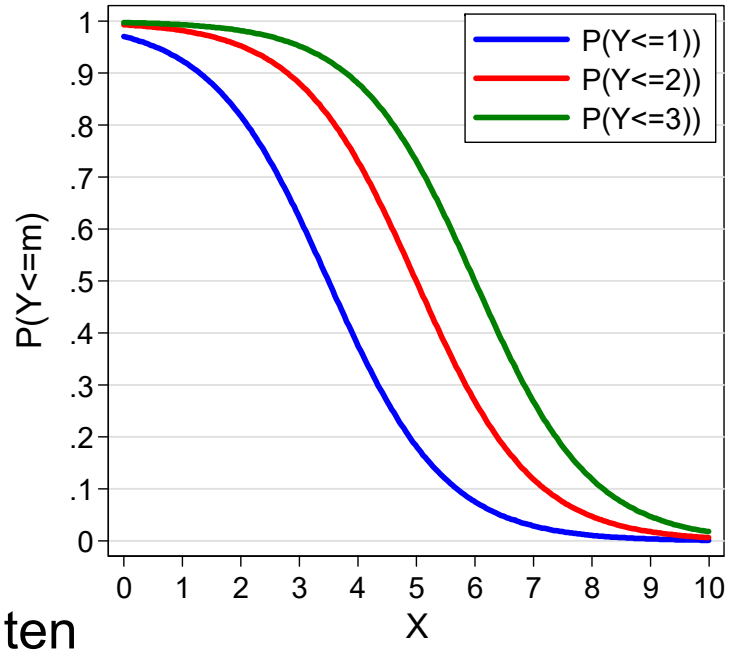
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Daten: ALLBUS 2002
Do-File: 11 OLogit.do

Die Annahme paralleler Regressionen

- Das ORM impliziert eine Annahme
  - Beispiel: ordinales Logit, J=4

$$P(y \leq 1) = \Lambda(\tau_1 - \boldsymbol{\beta}'\boldsymbol{x})$$
$$P(y \leq 2) = \Lambda(\tau_2 - \boldsymbol{\beta}'\boldsymbol{x})$$
$$P(y \leq 3) = \Lambda(\tau_3 - \boldsymbol{\beta}'\boldsymbol{x})$$



  - Daran erkennt man, das das ORM äquivalent ist zu drei binären Logits, wobei $\boldsymbol{\beta}$ identisch ist (nur die Konstanten sind verschieden) [parallel regression/line assumption]
  - Testprinzip: man schätzt die binären Logits und testet, ob die $\widehat{\boldsymbol{\beta}}$ gleich sind
  - Was tun bei Verletzung der Annahme?
    - Problematische Variablen weglassen (normalerweise keine gute Idee)
    - Alternative ORM: z.B. generalisiertes ordinales Logit
    - Multinomiale Modelle

Die Annahme paralleler Regressionen

```
. brant, detail                        //funktioniert nur mit "SPost"

Estimated coefficients from j-1 binary regressions

               y>1         y>2
alter    .00398831    .0102355
 bild   -.06185737   .00013163
 eink    .07913162   .00707855
 frau   -.25742583  -.51034535
  ost   -.46197847  -.72218376
_cons    1.9772367    -1.48633

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+---------------------------
         All |     13.84    0.017     5
-------------+---------------------------
       alter |      2.49    0.115     1
        bild |      6.53    0.011     1
        eink |      1.20    0.273     1
        frau |      3.33    0.068     1
         ost |      3.14    0.076     1
-----------------------------------------
A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

Daten: ALLBUS 2002
Do-File: 11 OLogit.do

Interpretation

- Vorzeicheninterpretation
  - Vorzeichen von $\widehat{\boldsymbol{\beta}}$ gibt Richtung des Effektes auf Y* bzw. auf P(Y=J)
- Odds Interpretation (nur Logit)

$$\frac{P(y \leq 1)}{P(y > 1)} = \exp(\tau_1 - \boldsymbol{\beta}'\boldsymbol{x})$$

$$\frac{P(y \leq 2)}{P(y > 2)} = \exp(\tau_2 - \boldsymbol{\beta}'\boldsymbol{x})$$

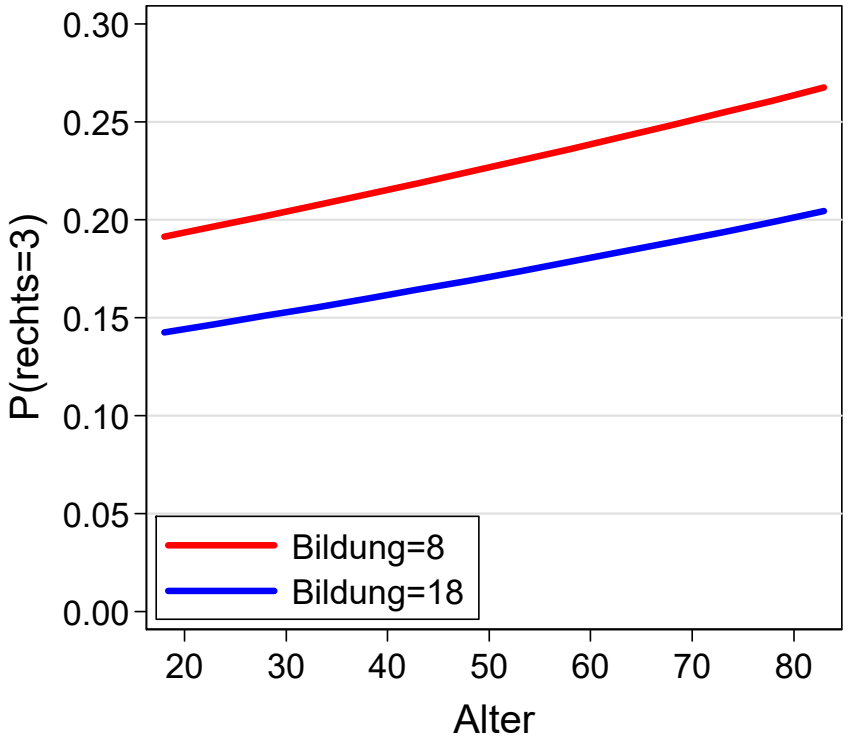
  - Es gibt also J - 1 Odds. Aufgrund der Annahme paralleler Regressionen sind die Effekte für alle Odds identisch (deshalb auch oft: „Annahme proportionaler Odds“)
  - $\exp(-\beta_j)$ ist der multiplikative Effekt auf die Odds kleiner/größer
  - Die Interpretation ist einfacher, wenn man die Odds als größer/ kleiner formuliert: $\exp(\beta_j)$
  - Die Odds Interpretation ist schwer zu verstehen

Bsp. Frau: ß=-0,3677, OR=0,69
Frauen haben eine um 31% geringere Chance eher rechts als links zu sein.

Profile-Plots

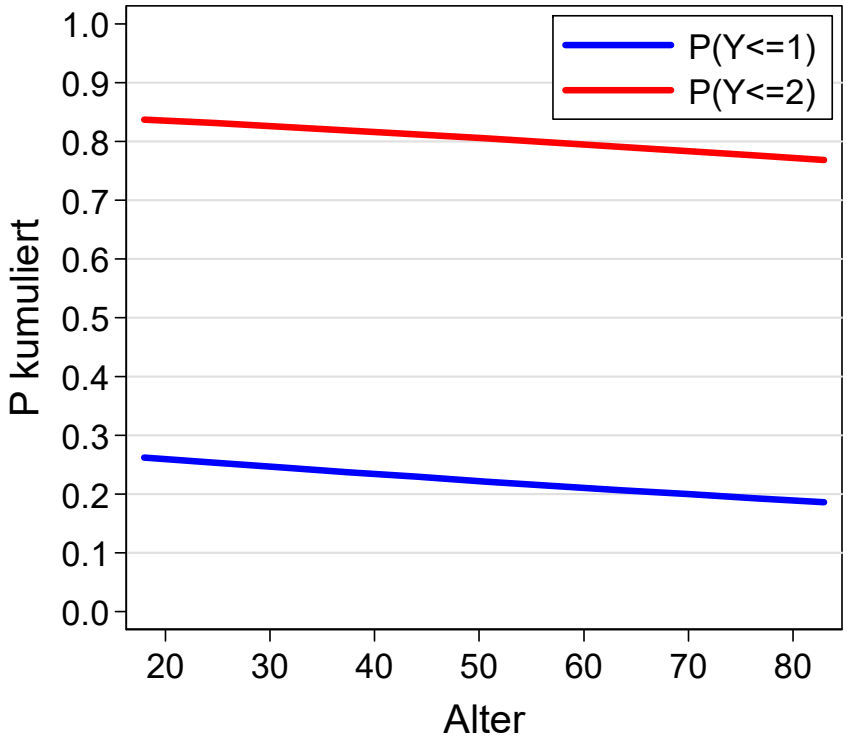
Alter und Bildung

```
ologit rechts alter bild eink frau ost
margins, at(alter=(18(5)83) ///
    bild=(8 18)) predict(outcome(3))
marginsplot, noci
```



Alter (P kumuliert)

```
ologit   rechts alter bild eink frau ost
prgen alter, from(18) to(83) generate(pr)
       //funktioniert nur mit "SPost"
twoway connected prs1 prs2 prx
```



Daten: ALLBUS 2002
Do-File: 11 OLogit.do

Wahrscheinlichkeitsinterpretation

Daten: ALLBUS 2002
Do-File: 11 OLogit.do

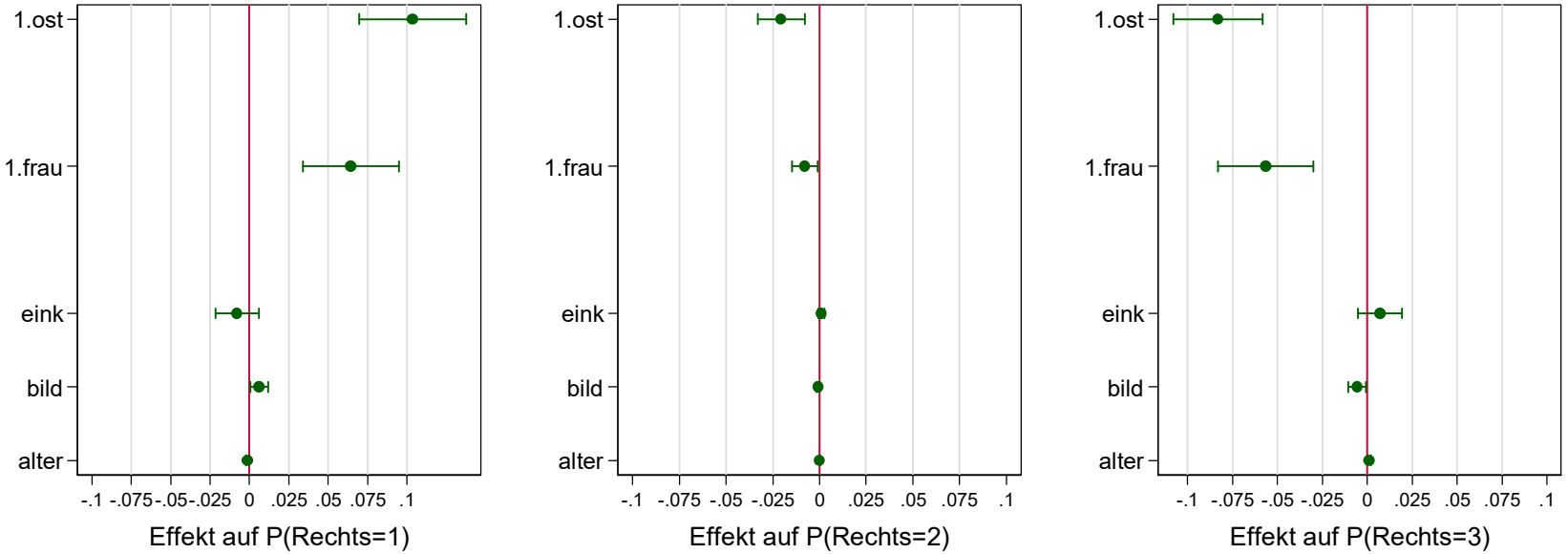
- Der Marginal Effekt (ME) im ORM ist

$$\frac{\partial P(Y=m)}{\partial x_j} = F'(\tau_m - \boldsymbol{\beta}'\overline{\boldsymbol{x}}) - F'(\tau_{m-1} - \boldsymbol{\beta}'\overline{\boldsymbol{x}}) =$$

$$= \beta_j \left[f(\tau_{m-1} - \boldsymbol{\beta}'\overline{\boldsymbol{x}}) - f(\tau_m - \boldsymbol{\beta}'\overline{\boldsymbol{x}}) \right]$$

  - Man beachte, dass es hier mehrere ME gibt (nämlich J Stück)
    - Die summieren sich über alle Kategorien zu 0!
  - Die ME sind die Steigungen in den PPs (bei metrischen Variablen)
- Weiterhin gibt es natürlich auch DC, AME und ADC
  - Das oben Gesagte gilt auch hier!

AMEs für unser Beispiel



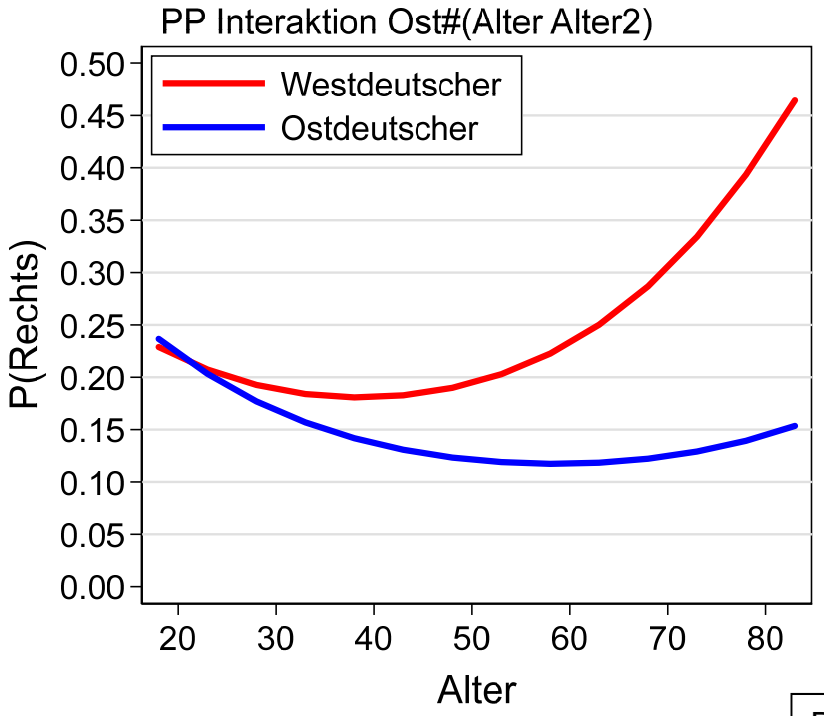
Interaktionseffekte

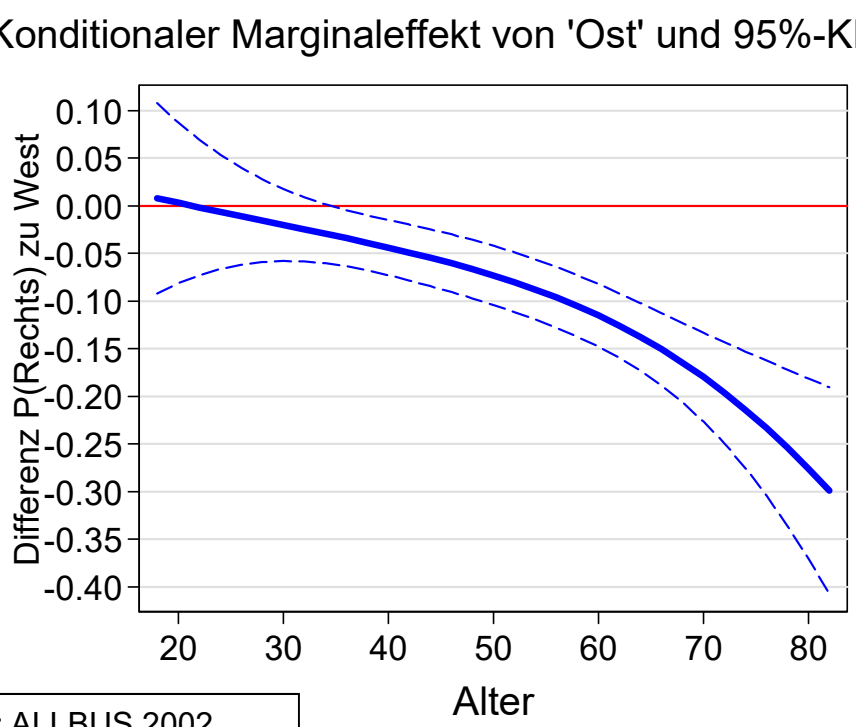
Profile-Plot
Polynomreg. + Interaktion mit Ost

```
ologit rechts bild eink frau      ///
    i.ost##(c.alter c.alter#c.alter)
margins ost, at(alter=(18(5)83))  ///
    predict(outcome(3))
marginsplot,   noci
```

Conditional-Effects-Plot

```
ologit rechts bild eink frau      ///
     i.ost##(c.alter c.alter#c.alter)
margins, at(alter=(18(2)82)) dydx(ost)
     predict(outcome(3))
marginsplot, recast(line)
             recastci(rline)
```





Daten: ALLBUS 2002
Do-File: 11 OLogit.do

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Kapitel 13:
Ausblick/Literatur

Ausblick

- Multiple Imputation
  - `mi` Prefix
- Bayesianische Statistik
  - `bayes` Prefix
- Komplexe Surveydaten
  - `svy` Prefix
- Zähldaten
  - `poisson, nbreg`
- Zensierte/trunkierte Daten
  - `tobit, intreg, truncreg, heckman`
- Strukturgleichungsmodelle
  - `sem`
- Mehrebenenmodelle
  - `mixed, me...`
- Ereignisdatenmodelle
  - `stcox, streg, st...`
- Paneldatenmodelle
  - `xtreg, xt...`

Literatur

Ai, C. und E. Norton (2003) Interaction terms in logit and probit models. Economics Letters 80: 123-129

Auspurg, K. und Th. Hinz (2011) Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen. ZfS 40: 62-73.

Best, H. und C. Wolf (2012) Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. KZfSS 64: 377-395

Breen, R., K. Karlson und A. Holm (2018) Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. Ann. Rev. Soc. 44: 39-54.

Karlson, K., A. Holm und R. Breen (2012) Comparing regression coefficients between same-sample nested models using logit and probit: a new method. Sociol. Methodol. 42:274–301.

Mood, C. (2010) Logistic Regression: Why we cannot do what we think we can do, and what we can do about it. Eur. Soc. Rev. 26: 67-82.