

Angewandte Regressionsanalyse mit STATA

Prof. Dr. Josef Brüderl
LMU München

April 2012



Inhalt I

• I) Was ist eine Regression?	8
• II) Regression und Kausalität	15
• III) Das lineare Regressionsmodell	20
– Das multiple lineare Regressionsmodell	21
– Beispiel: Statuszuweisungsmodell	
– Was bedeutet Kontrolle?	
– Regression mit Dummy	32
– Regressionsdiagnostik	38
– Linearität, Homoskedastizität, Normalverteilung, Ausreißer	
– Interpretation von Regressionskoeffizienten	47
– Conditional-Effect-Plot, Polynomregression, semi-logarithmische Regression	
– Berücksichtigung von Interaktionseffekten	54
– Regeln für den Umgang mit Interaktionen	
– Systematik der Regressionsplots	65

Inhalt II

• IV) Maximum-Likelihood Schätzung	66
– ML-Schätzung, Tests, Modellfit	67
– ML-Output, Modellvergleich	72
• V) Regressionsmodelle für binäre Outcomes	78
– Einführung (nur bivariate Modelle)	
– Das lineare Wahrscheinlichkeitsmodell	80
– Das logistische Modell	82
– Beispiel: Bleiben am Studienort	
– STATA-Beispiel: Arbeitslosigkeit	
– Multiple logistische Regression	94
– Wahrscheinlichkeitsinterpretation	
– Interaktionseffekte	100
– Probleme mit Vergleichen	105
– Logit vs. Probit	107

Inhalt III

• VI) Regressionsmodelle für ordinale Outcomes	108
– Ordinales Logit/Probit	110
– Annahme paralleler Regressionen	114
– Interpretation	116
– Generalisiertes ordinale Logit	120
• VII) Regressionsmodelle für multinomiale Outcomes	124
– Multinomiales Logit	126
– Unabhängigkeit von irrelevanten Alternativen	131
– Interpretation	133
– Ein Modell mit Interaktionseffekten	136

Lernziele

- Kenntnis verschiedener Regressionsmodelle
 - Hier nur Querschnittsregression
 - Grundkenntnisse der linearen und der logistischen Regression werden vorausgesetzt
- 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**
- 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)
- Unabhängige Variablen
 - bild: schulische und berufliche Bildung in Jahren
 - alter: Alter in Jahren
 - exp: Berufserfahrung (alter - bild - 6)
 - prestv: Berufsprestige des Vaters (Magnitude-Skala)
 - frau: Dummy für Frau
 - ost: Dummy für Ostdeutscher
 - beruf: berufliche Stellung (Arbeiter, Angest., Beamter, Selbst.)

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



Kapitel I: Was ist eine Regression?



Regression: Pol. Einstellung auf Alter

```
. regress rechts alter if ost==0
```

Source	SS	df	MS				
Model	168.820066	1	168.820066	Number of obs =	1820		
Residual	5989.10081	1818	3.29433488	F(1, 1818) =	51.25		
				Prob > F =	0.0000		
				R-squared =	0.0274		
				Adj R-squared =	0.0269		
Total	6157.92088	1819	3.38533308	Root MSE =	1.815		

rechts	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
alter	.0177624	.0024813	7.16	0.000	.0128959	.0226288
_cons	4.364548	.1233541	35.38	0.000	4.122617	4.606479

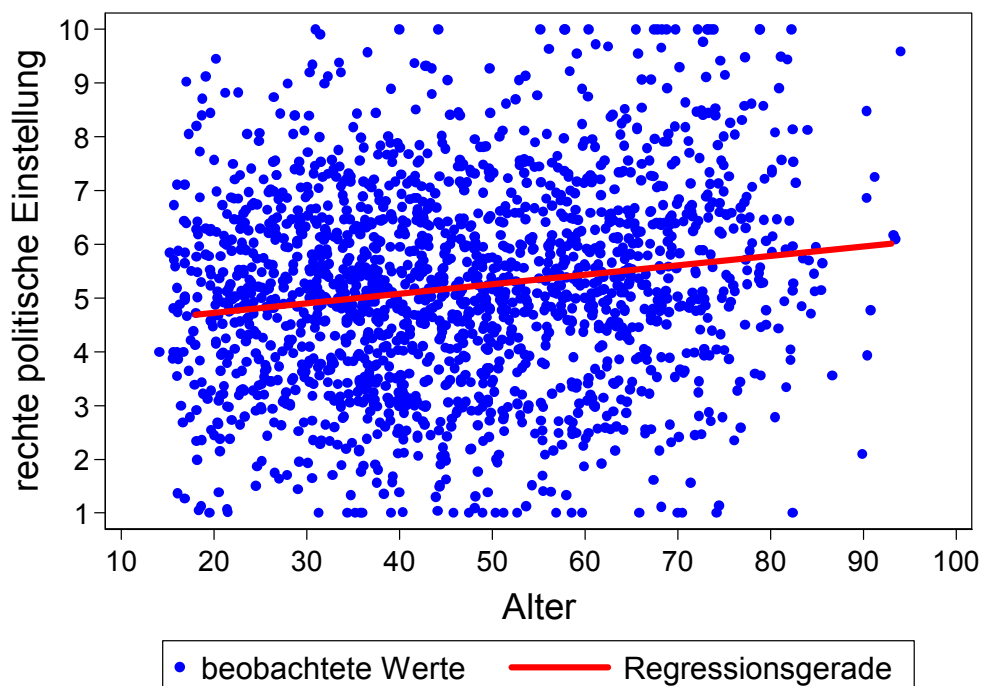
Regressionskoeffizient

t-Wert

p-Wert

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

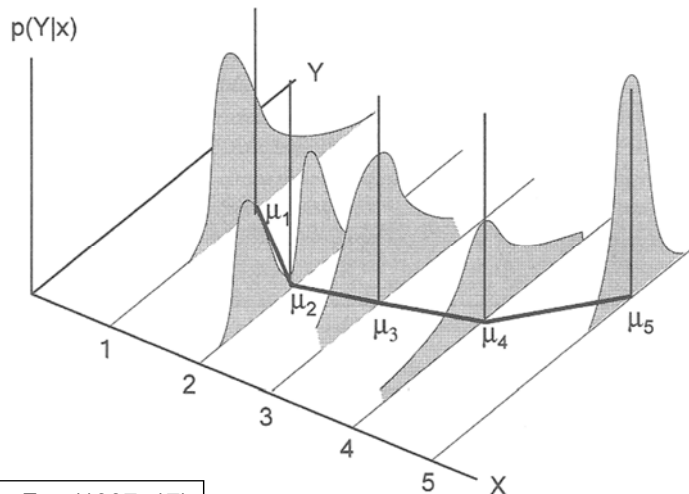
Regression: politische Einstellung auf Alter



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

Regression als bedingte Verteilung

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



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

Quelle: Fox (1997: 17)

Josef Brüderl, Regressionsanalyse, April 2012

11

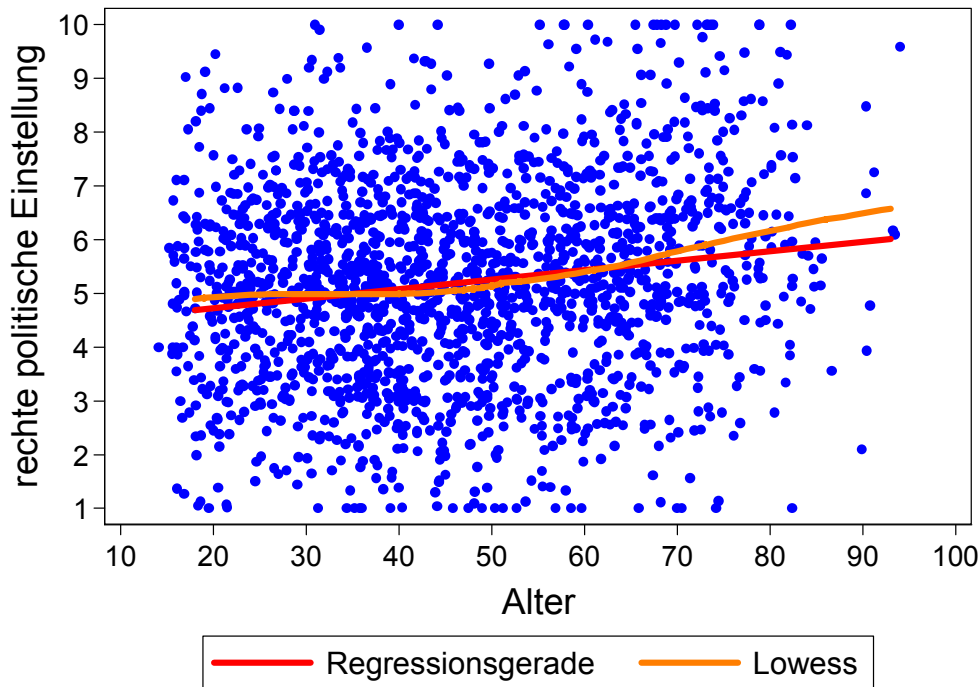
Regression als bedingte Verteilung

- 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

Josef Brüderl, Regressionsanalyse, April 2012

12

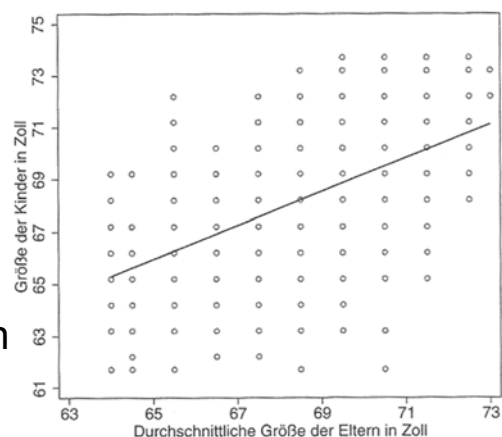
Nicht-parametrische und parametrische Regression



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

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 durch die bedingten Mittelwerte
 - 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)

Kapitel II:

Regression und Kausalität



Kausalität

- Wissenschaft sucht nach kausalen Beziehungen
 - Ursache-Wirkungs-Beziehungen
- 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 meist nicht praktikabel

Korrelation und Kausalität

- Deshalb erhebt man oft Daten über X und Y ex-post-facto und berechnet deren Korrelation
- **Korrelation ist aber nicht gleich Kausalität**
- 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)
 - Paneldaten nötig
 - Bei Querschnittsdaten hilft nur Theorie
 - Die Korrelation von X und Y bleibt erhalten, auch wenn man für dritte Variablen kontrolliert (keine unbeobachtete Heterogenität)
 - Dies macht man mit meist Regressionsverfahren
- Damit kommt der Regression in der Sozialforschung eine zentrale Rolle zu:
Sie ist der Ersatz für das Experiment

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
 - Rauchen (gucken von Privatsendern) macht dumm
 - Zähneputzen senkt das Herzinfarktrisiko

Möglichkeiten der Kausalanalyse

- „Selection on observables“
 - Die Selektionsvariablen sind gemessen
 - Man kann für sie kontrollieren
 - Regression
 - Matching
- „Selection on unobservables“
 - Die Selektionsvariablen sind nicht gemessen
 - Regressions- und Matching-Schätzer sind verzerrt
 - Scheinkorrelation, unbeobachtete Heterogenität, „omitted variable bias“
 - Unverzerrte Schätzer kann man erhalten mit
 - Instrumentalvariablen Ansatz
 - Exogene Variation identifiziert den Kausaleffekt
 - Regression Discontinuity Ansatz
 - Homogenität von Personen an einer Schwelle
 - Within-Panelanalyse
 - Homogenität einer Person über die Zeit



Kapitel III: Das lineare Regressionsmodell



Das multiple lineare Regressionsmodell

Das Modell:

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

- 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
 - β_j sagt uns, welcher Effekt verbleibt, wenn wir für die anderen uVs kontrollieren

OLS Schätzung

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

mit :

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

Annahmen :

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

$Cov(\mathbf{x}, \boldsymbol{\varepsilon}) = \mathbf{0}$ keine unbeobachtete Heterogenität

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

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

Multiples R^2

- Die vorhergesagten Werte ergeben sich aus

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

- 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

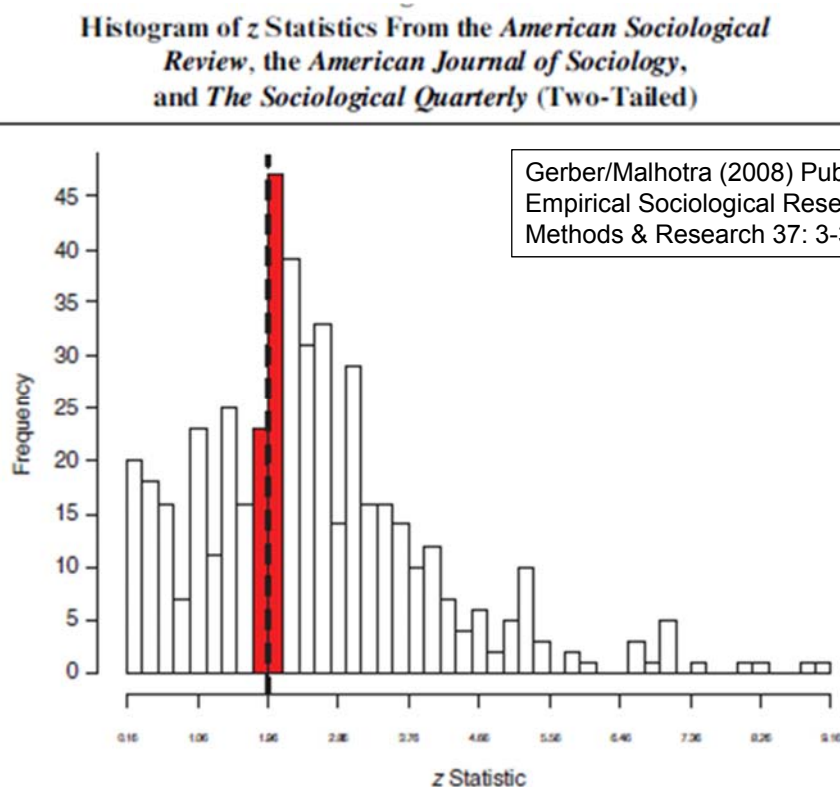
Signifikanztests

- 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 $\alpha=5\%$: $|T| > 2$)
- Test des gesamten Modells: overall F-Test
 - Nullhypothese: keine X-Variable hat einen Einfluss auf Y
 $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$
 - Die Teststatistik (F-Wert) ist
$$F = \frac{MSS/p}{RSS/(n-p-1)} \sim F(p, n-p-1)$$
 - Die H_0 wird verworfen, falls: $F > F_{1-\alpha}(p, n-p-1)$

Das Signifikanztest-Ritual

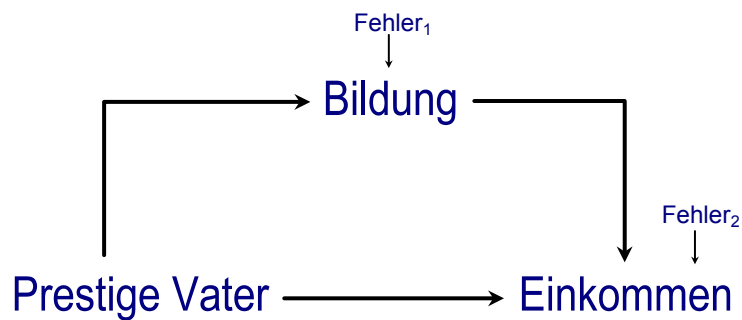
- Unsinnige Anwendung der Signifikanztests in der Praxis
 - Multiples Testen und Publikations-Bias
 - Auch wenn in Wirklichkeit kein relevanter Effekt vorliegt, so werden von 100 Forschern irrtümlicherweise 5 einen „signifikanten“ Effekt finden und genau diese 5 „signifikanten“ Effekte werden publiziert
 - Analog: Variablen-/Modellselektion anhand von t-Tests
 - Knapp 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“
 - Ziliak/McCloskey (2008): 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!
- Statt Abschaffung, Verbesserung der Praxis (s.a. Krämer 2011)
 - 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!)

Publikations-Bias



Beispiel: Statuszuweisungsmodell

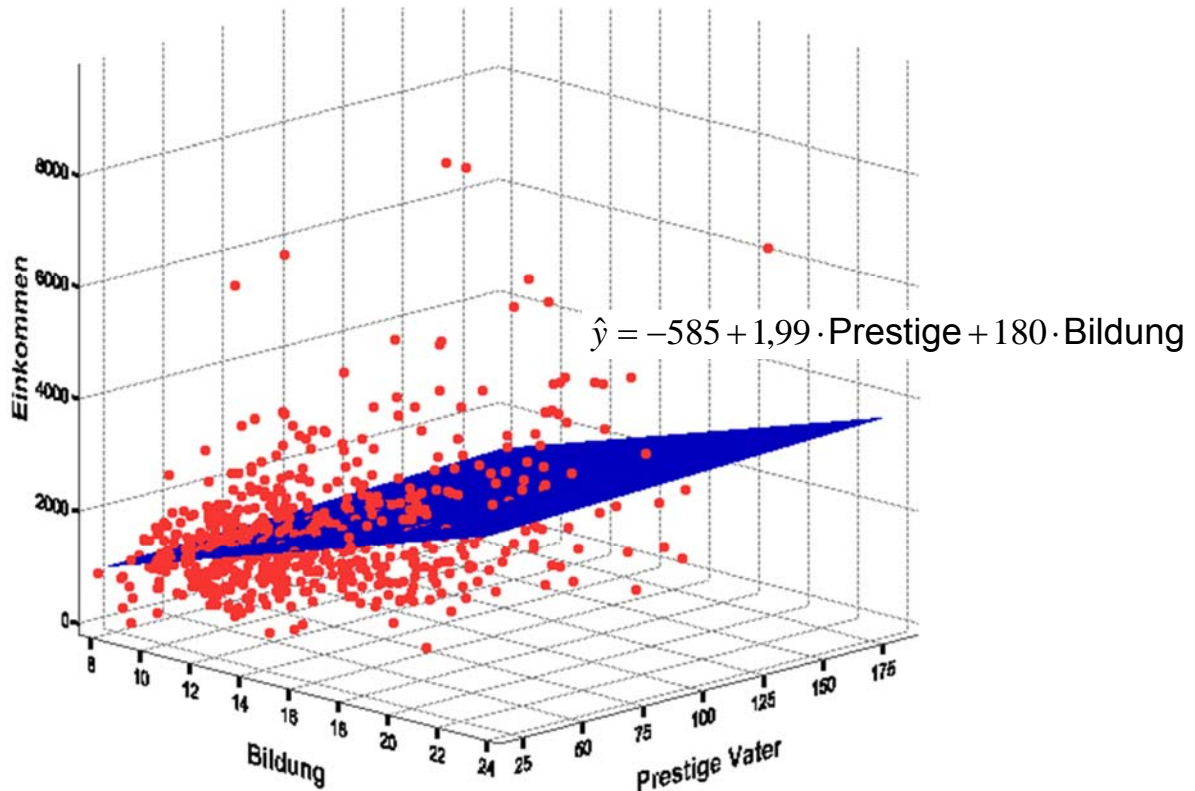
- Blau/Duncan (1967) "The American Occupational Structure"
 - Wie erlangt man seine soziale Position?
Durch „achievement“ oder Statusvererbung?
 - ALLBUS 2002:
 - Abhängige Variable: monatliches Netto-Einkommen in Euro (nur Westdeutsche, Vollzeit)
 - Status des Vaters: Magnitude-Prestigeskala (Werte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



STATA Beispiel: Statuszuweisungsmodell

<code>. regress eink prestv bild</code>						
Source	SS	df	MS	Number of obs = 670		
Model	215244302	2	107622151	F(2, 667) = 64.99		
Residual	1.1045e+09	667	1655904.67	Prob > F = 0.0000		
Total	1.3197e+09	669	1972694.65	R-squared = 0.1631		
				Adj R-squared = 0.1606		
				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

Regressionebene



Josef Brüderl, Regressionsanalyse, April 2012

29

Beispiel: Statuszuweisungsmodell

	(1)	(2)
Konstante	1343	-585
Prestige Vater	9,6	2,0
Bildung		180
R ²	0,03	0,16
N	670	670

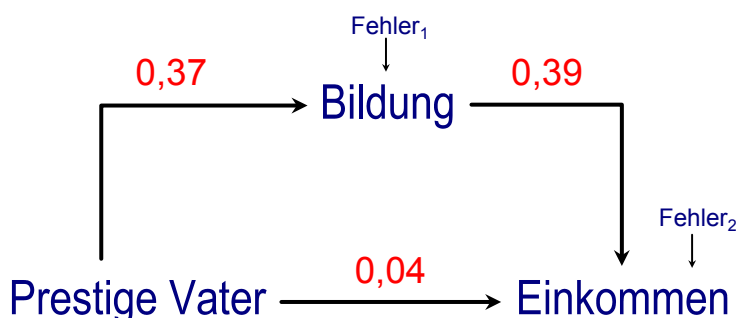
Bivariates Modell (1)

“Prestige Vater” hat einen starken Effekt

Multiples Modell (2)

Der wird unter Kontrolle der “Bildung” deutlich kleiner (Intervention)

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



Pfaddiagramm

Hier sind die standardisierten Regressionskoeff. eingetragen.

Josef Brüderl, Regressionsanalyse, April 2012

30

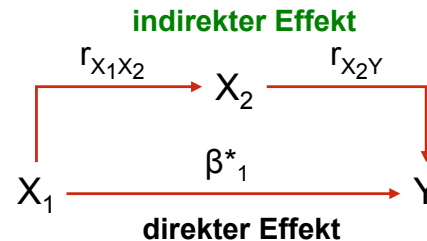
Was bedeutet „Kontrolle“?

- β_j gibt den 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 indirekten Effekten „bereinigt“

bivariater Effekt indirekter Effekt

$$\hat{\beta}_1^* = \frac{r_{X_1 Y} - r_{X_1 X_2} r_{X_2 Y}}{1 - r_{X_1 X_2}^2}$$

Standardisierung



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

$$\hat{\beta}_1^* = r_{X_1 Y}$$

- Der multiple Regressionskoeffizient ist gleich dem bivariaten

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		
Model	358146421	2	179073211	F(2, 1115) = 116.00		
Residual	1.7213e+09	1115	1543745.36	Prob > F = 0.0000		
Total	2.0794e+09	1117	1861613.69	R-squared = 0.1722		
				Adj R-squared = 0.1707		
				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: 3 LinReg Dummies.do

Kategoriale uV mit mehr als 2 Ausprägungen

- Auch solche Variablen können bei der multiplen Regression als uV benutzt werden
- 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		Total
beruf	0	
Arbeiter	0	360
Angestellter	622	0
Beamter	90	0
Selbständiger	160	0
Total	872	360

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: 3 LinReg Dummies.do

Regression mit kategorialer uV

```
. regress eink bild i.beruf
```

Source	SS	df	MS
Model	330638913	4	82659728.2
Residual	1.6674e+09	1067	1562726.39
Total	1.9981e+09	1071	1865609.69

Number of obs = 1072
F(4, 1067) = 52.89
Prob > F = 0.0000
R-squared = 0.1655
Adj R-squared = 0.1624
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: 3 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. Intervall]	
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: 3 LinReg Dummies.do

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:
 - Linearitätsannahme
 - Homoskedastizitäts-Annahme
 - Normalverteilungsannahme
 - zusätzlich: Ausreißerdiagnostik
- Dazu analysiert man die Residuen (Residuenanalyse)
 - Die Fehlerterme selbst sind nicht beobachtbar
 - Die Residuen sind allerdings Schätzer für die Fehlerterme

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

Linearität

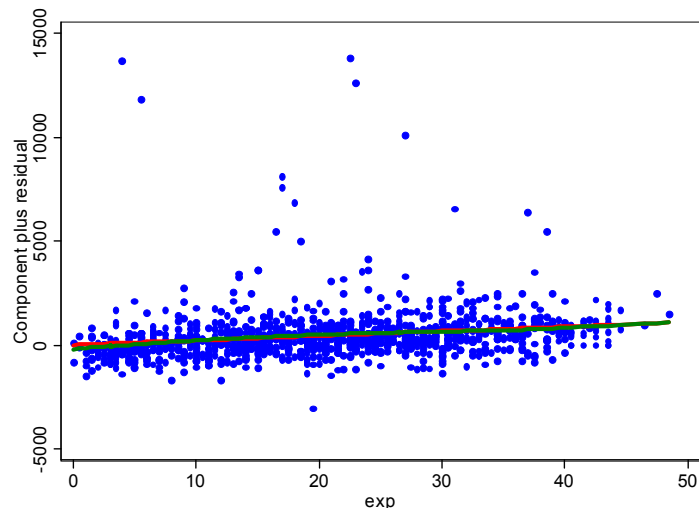
Daten: ALLBUS 2002
Do-File: 4 LinReg Diagnostik.do

- Nicht-Linearität erkennt man in einem Residuen-Plot
 - Residuen gegen die uV aufragen
 - 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.



Josef Brüderl, Regressionsanalyse, April 2012

39

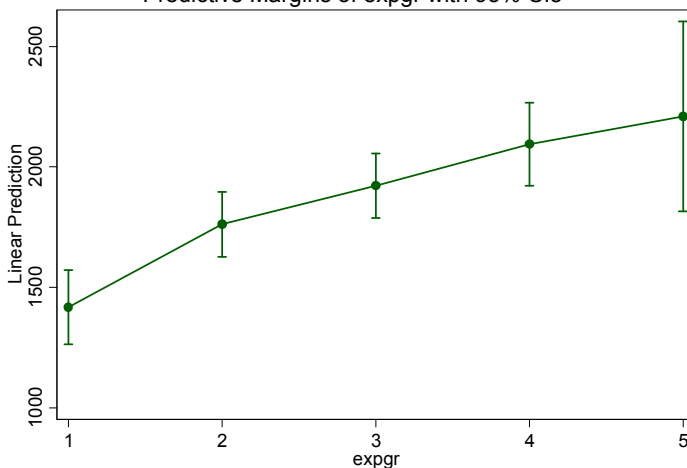
Linearität

Daten: ALLBUS 2002
Do-File: 4 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
```

Predictive Margins of expgr with 95% CIs



	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		

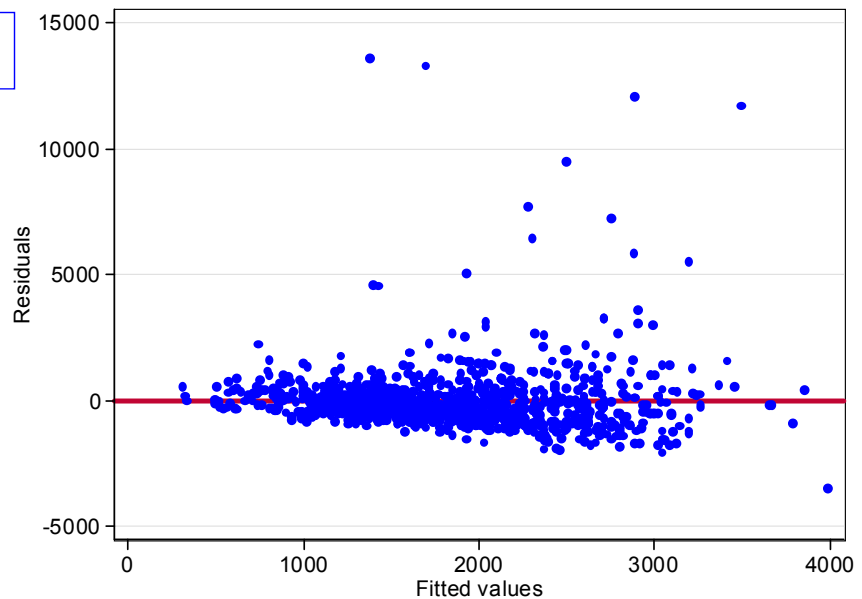
Josef Brüderl, Regressionsanalyse, April 2012

40

- 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.)

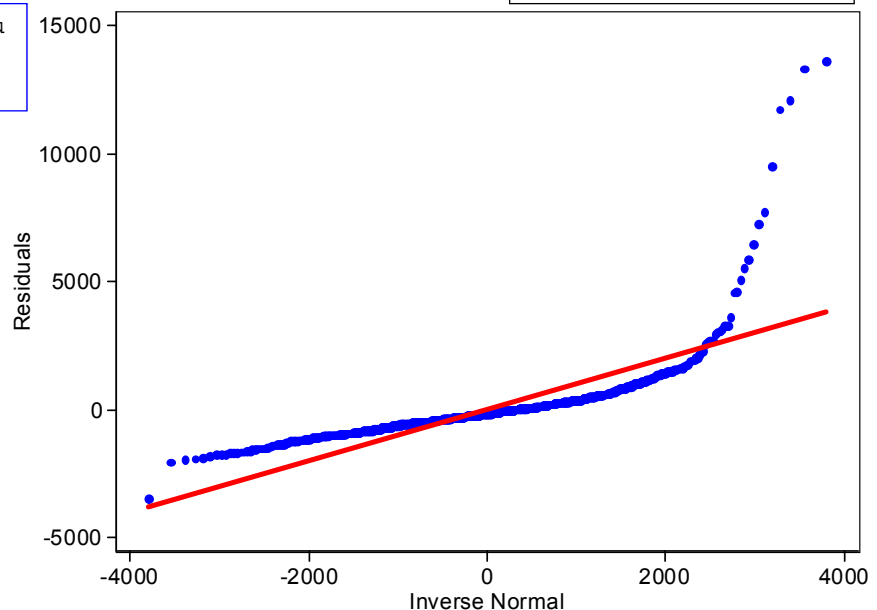


Normalverteilungsannahme

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

```
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

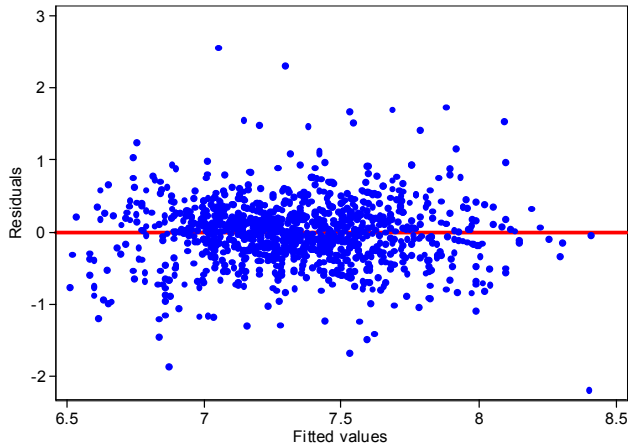


Daten: ALLBUS 2002
Do-File: 4 LinReg Diagnostik.do

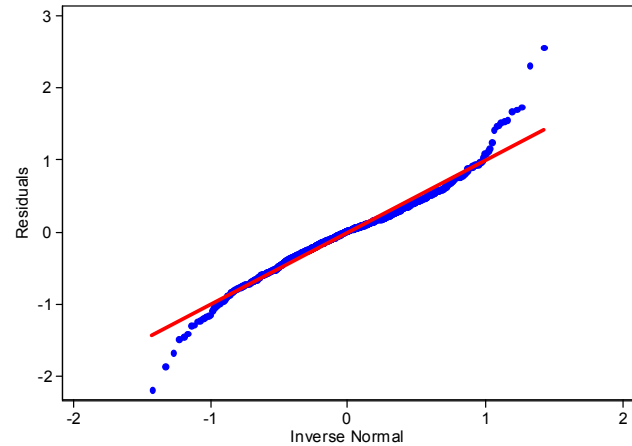
Semi-logarithmische Einkommensregression

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

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



Kein augenfälliges Muster erkennbar:
Homoskedastie

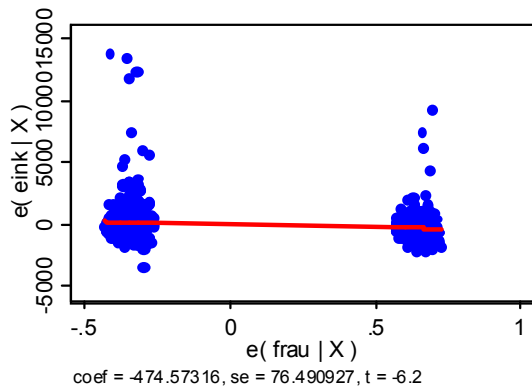
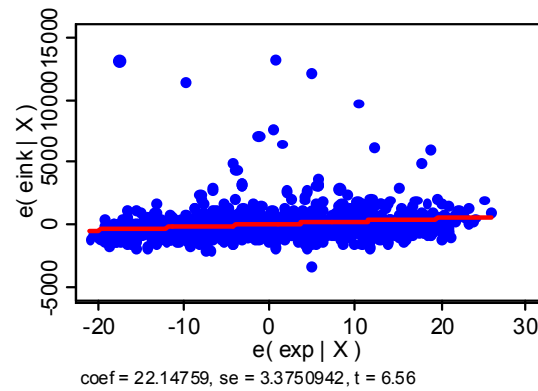
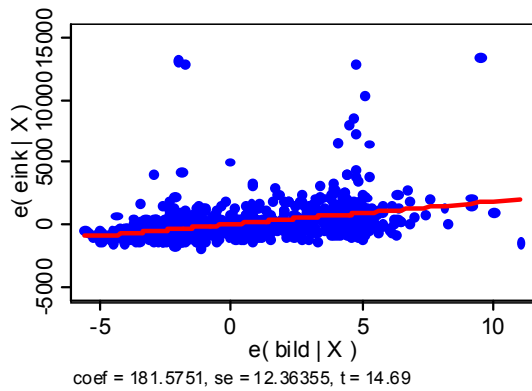


Nur mehr geringe Abweichungen an den
Ränder: Residuen fast normalverteilt

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!

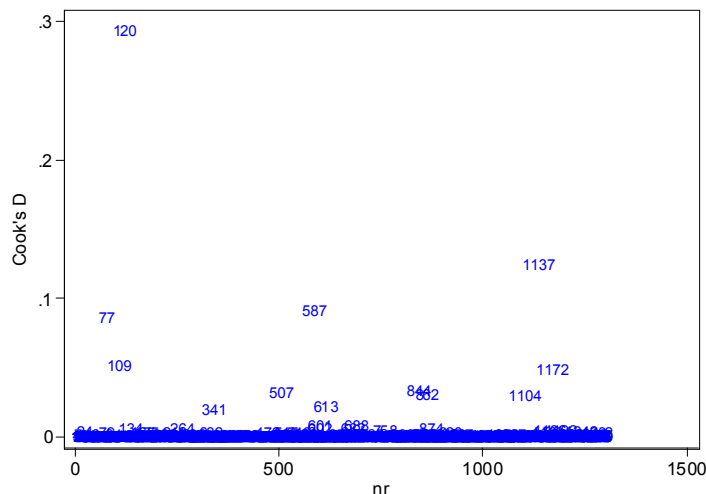
Partielle-Regressions Streudiagramme



Added-variable plots:
avplots

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

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

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

- Verschiedene Interpretationsmöglichkeiten

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

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

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

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

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

Präsentation der Regressionskoeffizienten

- Bsp. Einkommensregression
 - Monatliches Nettoeinkommen in Euro (nur Vollzeiterwerbstätige)
 - Bildungsjahre, Prestige Vater/10, Ostdeutscher, Frau, berufliche Stellung

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

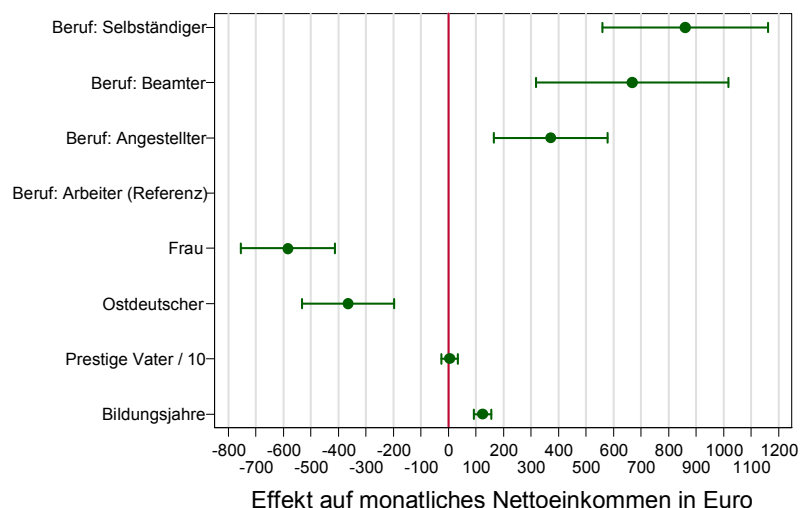
```
margins, dydx(*)
marginsplot, horizontal
```

Bildungsjahre	123.3***
	(7.67)
Prestige Vater / 10	4.7
	(0.31)
Ostdeutscher	-364.7***
	(-4.26)
Frau	-583.5***
	(-6.71)
Angestellter	371.8***
	(3.54)
Beamter	667.9***
	(3.75)
Selbständiger	860.7***
	(5.62)
Konstante	163.2
	(0.85)
N	948
R ²	0.214

t statistics in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Regressionskoeffizienten und 95%-KI

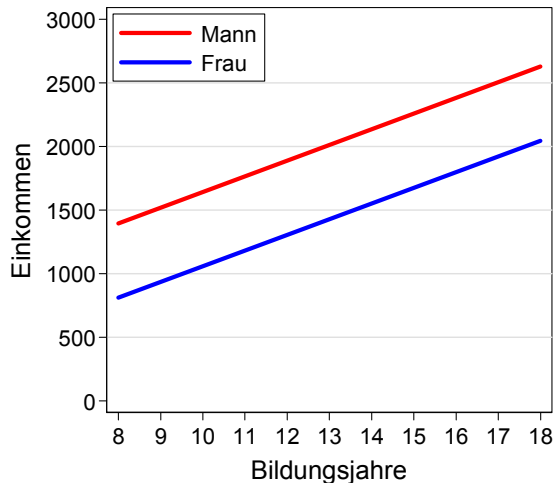


Daten: ALLBUS 2002

Do-File: 5 LinReg Interpretation.do

Conditional-Effect-Plot

- Hilfreich zur Veranschaulichung einzelner 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: 5 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: 5 LinReg Interpretation.do

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

<pre>. regress eink bild ost frau c.exp c.exp#c.exp</pre>						
Source	SS	df	MS			
Model	474909648	5	94981929.7	Number of obs =	1118	
Residual	1.6045e+09	1112	1442907.24	F(5, 1112) =	65.83	
				Prob > F =	0.0000	
				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))
```

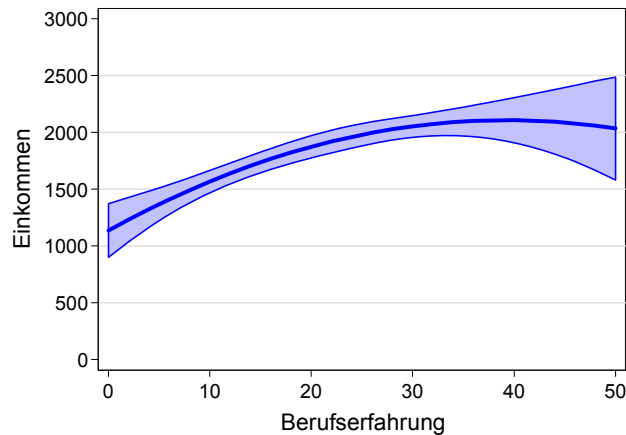
	Delta-method		z	P> z
	dy/dx	Std. Err.		
0	49.53	12.57	3.94	0.000
10	36.91	7.18	5.14	0.000
20	24.28	3.38	7.18	0.000
30	11.66	6.13	1.90	0.057
40	-0.97	11.41	-0.08	0.932
50	-13.59	17.00	-0.80	0.424

```
. test exp exp#exp
```

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

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

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



Conditional-Effect-Plot

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

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

Josef Brüderl, Regressionsanalyse, April 2012

51

Das semi-logarithmische Regressionsmodell

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

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

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

Marginaleffekt prozentuale Ver.

Gute Näherung, falls $|\beta| < 0,1$

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

Discrete Change prozentuale Veränderung

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

Josef Brüderl, Regressionsanalyse, April 2012

52

Das semi-logarithmische Regressionsmodell

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

Source	SS	df	MS	Number of obs = 1118		
Model	138.195766	5	27.6391532	F(5, 1112)	=	145.12
Residual	211.79019	1112	.190458804	Prob > F	=	0.0000
				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: 5 LinReg Interpretation.do

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

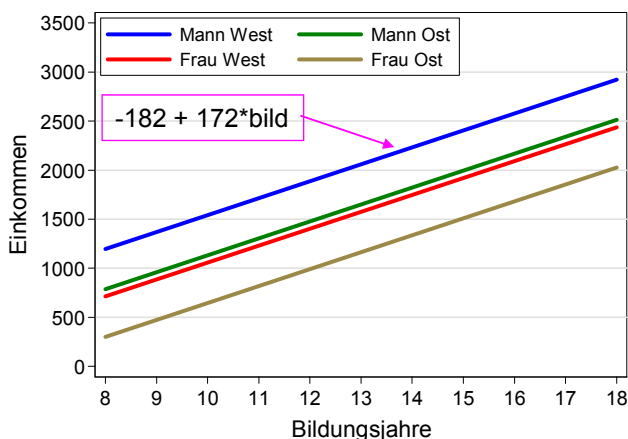
Do-File: 6 LinReg Interaktion.do

Josef Brüderl, Regressionsanalyse, April 2012

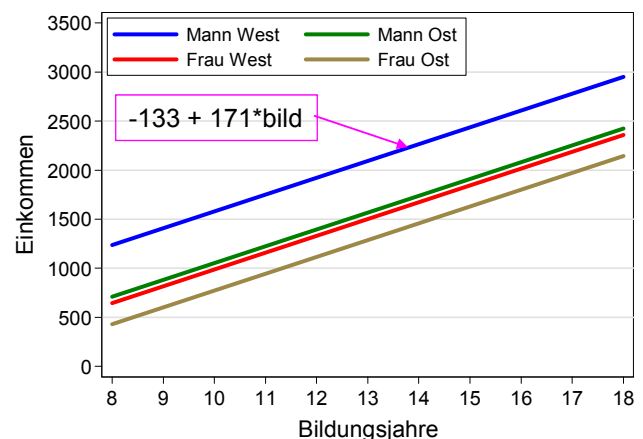
55

Dummy-Interaktion: Geschlecht/Wohnort

ohne Interaktion



mit Interaktion



Designmatrix	Frau	Ost	Ofrau	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: 6 LinReg Interaktion.do

Josef Brüderl, Regressionsanalyse, April 2012

56

Slope-Interaktion: Geschlecht/Bildung

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

eink	Coef.	Std. Err.	t	P> t	[95% Conf. Intervall]	
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

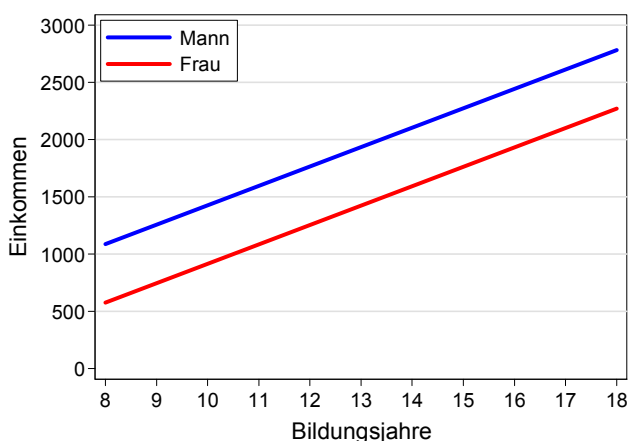
eink	Coef.	Std. Err.	t	P> t	[95% Conf. Intervall]	
bild	190.2989	15.17511	12.54	0.000	160.5239	220.0739
frau	335.2419	359.1157	0.93	0.351	-369.3776	1039.861
frau#c.bild	-63.77517	26.41776	-2.41	0.016	-115.6094	-11.94099
_cons	-546.0541	207.9355	-2.63	0.009	-954.0433	-138.0648

Frauen verdienen mehr als Männer!?

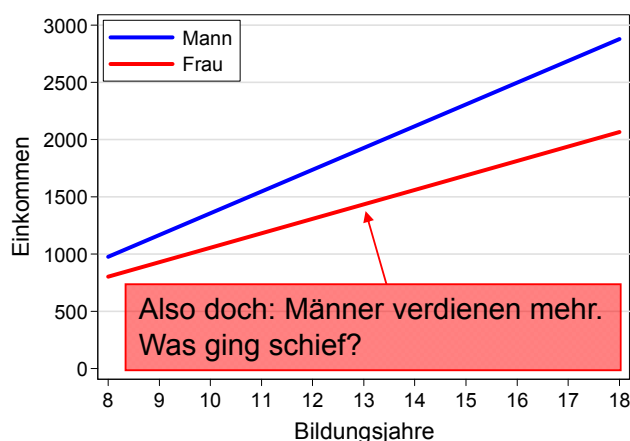
Daten: ALLBUS 2002
Do-File: 6 LinReg Interaktion.do

Slope-Interaktion: Geschlecht/Bildung

ohne Interaktion



mit Interaktion



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

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

Slope-Interaktion: Zentrierung

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

- 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		
Model	367104408	3	122368136	F(3, 1114) = 79.61		
Residual	1.7123e+09	1114	1537089.85	Prob > F = 0.0000		
Total	2.0794e+09	1117	1861613.69	R-squared = 0.1765		
				Adj R-squared = 0.1743		
				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

Josef Brüderl, Regressionsanalyse, April 2012

59

Berichte auch die konditionalen Marginal-effekte 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 Marginal-effekten
 - 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)

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

```
. regress eink cbild_m cbild_f frau
```

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

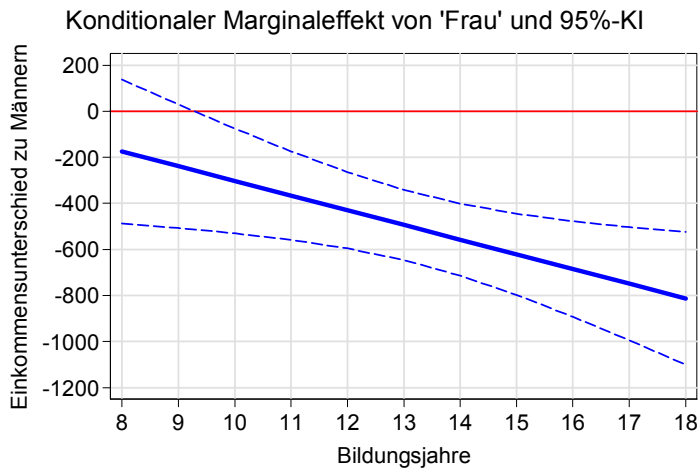
eink	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cbild_m	190.2989	15.17511	12.54	0.000	160.5239	220.0739
cbild_f	126.5237	21.62439	5.85	0.000	84.09456	168.9528
frau	-513.8864	77.55202	-6.63	0.000	-666.0509	-361.7219
_cons	1987.661	46.14574	43.07	0.000	1897.119	2078.204

Josef Brüderl, Regressionsanalyse, April 2012

60

Berichte auch die konditionalen Marginal effekte II

- Symmetrie: Bildung moderiert den Effekt von „Frau“
 - Konditionale Marginal effekte „Frau“: $335 - 64 \cdot \text{Bild}$
 - 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 Marginal-Effects-Plot

In Stata 12 leicht zu produzieren

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

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

Josef Brüderl, Regressionsanalyse, April 2012

61

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: 6 LinReg Interaktion.do

Josef Brüderl, Regressionsanalyse, April 2012

62

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 $Z = \text{mean}(Z)$
 - 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 Marginal-Effects-Plot)

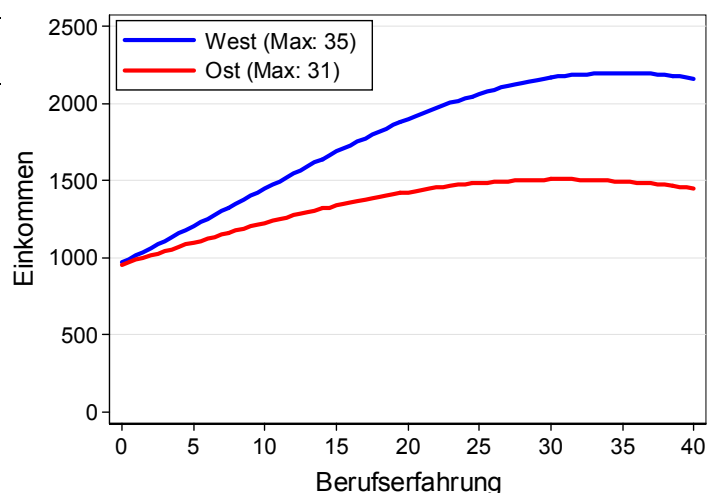
Zum Schluss: Ein Humankapitalmodell getrennt für West/Ost

AV: logarithmiertes Einkommen

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

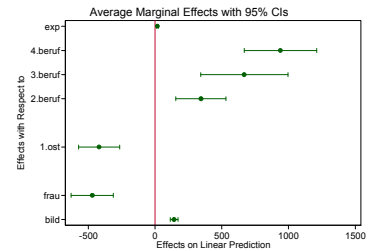
Daten: ALLBUS 2002
Do-File: 6 LinReg Interaktion.do

Systematik der Regressionsplots

```
. * Ein komplexes lineares Regressionsmodell  
. regress eink bild frau i.ost##(i.beruf c.exp c.exp#c.exp)
```

I) Plot der Regressionskoeffizienten (AMEs)

```
margins, dydx(*)  
marginsplot, horizontal xline(0) plotopts(connect(i))
```



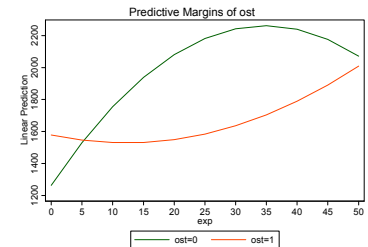
II) Profile Plot (PP) (formerly: Conditional-Effect Plot)

```
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 zweier kategorialer Variablen  
marginsplot, plotopts(connect(i))
```

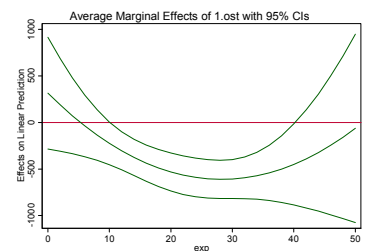
```
margins ost, at(exp=(0(5)50)) //Interaktion metrisch/kategorial  
marginsplot, noci recast(line)
```



III) Conditional-Marginal-Effects Plot (CMEP)

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

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



Josef Brüderl, Regressionsanalyse, April 2012

Daten: ALLBUS 2002
Do-File: 6a LinReg Regressionsplots.do

65



Kapitel IV:

Maximum-Likelihood Schätzung



Maximum-Likelihood Schätzung

- Schätzung meist mit Maximum-Likelihood (ML)
 - Im linearen Modell entspricht OLS dem ML-Schätzer
- Maximum-Likelihood Prinzip
 - Daten: (y_i, x_i)
 - Regressionsmodell: $f(Y=y_i|x_i, \beta)$
 - Schätzprinzip: Bestimme β so, dass die Wahrscheinlichkeit diese Daten zu beobachten, maximal wird
 - Die Wahrscheinlichkeit (Likelihood) der Daten unter dem gegebenen Modell und unabhängiger Stichprobenziehung ist

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

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

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

- Ableiten und Null-Setzen liefert die ML-Schätzer

Eigenschaften der ML-Schätzer

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

- Konsistent (unverzerrt)

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

- Normalverteilt

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

- 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

- Whs. $Y=1$: $P(Y=1)$
- Whs. $Y=0$: $P(Y=0)$
- Likelihood (McFadden 1972)

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

- 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 $Z = \frac{\hat{\beta}_j}{\hat{\sigma}_j} \stackrel{a}{\sim} N(0,1)$

- Die H_0 wird abgelehnt, falls $|Z| > z_{1-\alpha/2}$
 - Ab $n > 100$ sinnvoll (Faustregel für $\alpha=5\%$: $|Z| > 2$)

- Test des gesamten Modells: Likelihood-Ratio Test

- Nullhypothese: keine X -Variable hat einen Einfluss auf Y

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

- Die Teststatistik (LR-Wert) ist $\chi^2 = -2 \ln \left(\frac{L_0}{L_1} \right) = 2(\ln L_1 - \ln L_0)$

- 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-R²

- R² nicht sinnvoll, da keine sinnvolle Streuungszerlegung
- In Analogie: Pseudo-R² 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-R²

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

- nicht: Anteil erklärter Varianz
- Relative Log-Likelihood Verbesserung (im Vergleich zum Nullmodell)
- Fällt kleiner aus, als das R² des Linearen Wahrscheinlichkeitsmodells
- Weitere Pseudo-R² s. Long/Freeze (2006) S. 109 ff
 - Manche Autoren präferieren McKelvey/Zavoina's R²

ML-Output

- Am 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 arbloos 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

- Negative Werte der Log-Likelihood, da Likelihood fast 0
- „Iteration 0“ ist der Startwert: **ln(L₀)**
- 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(L₁)**

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)

R²-McFadden:

$$\frac{-850 + 775}{-850}$$

ML-Output

z-Wert

p-Wert

Koeffizienten Tabelle

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

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 Gruppen von VC
- **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` = $\exp(b \cdot \text{SD of X})$ = 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 Alter² gemeinsam signifikant?

```
. logit arblo bild frau ost
(output omitted)
. estimates store null
.
. logit arblo bild frau ost alter c.alter#c.alter
```

arblos	Coef.	Std. Err.	z	P> z	[95% Conf. Intervall]
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
(Assumption: null nested in full)

LR chi2(2) =    29.18
Prob > chi2 =    0.0000
```

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

Fitmaße

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

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

Measures of Fit for logit of arblo

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'

$$BIC' = -LR + p \ln N$$

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

Model:	Current logit	Saved logit	Difference
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.



Kapitel V: Regressionsmodelle für binäre Outcomes



Regression mit kategorialen Outcomes

- Bei kategorialen Outcomes kommen spezielle Regressionsverfahren zum Einsatz
 - Binäres Outcome
 - Lineares Wahrscheinlichkeitsmodell, logistische Regression (Logit), Probit
 - Ordinales Outcome
 - Ordinales Logit, ordinale Probit
 - Multinomiales Outcome
 - Multinomiales Logit, multinomiales Probit
- Binäres Outcome
 - Die dichotome aV sollte 0/1-kodiert sein
 - Beispiele:
 - Wer wählt CDU?
 - Wer wird arbeitslos?
 - Wer macht eine Uni-Ausbildung?

Das lineare Wahrscheinlichkeitsmodell

- Man kann eine dichotome aV linear modellieren

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

- Da gilt

$$E(y) = P(Y = 0) \cdot 0 + P(Y = 1) \cdot 1 = P(Y = 1)$$

- Erhält man das lineare Wahrscheinlichkeitsmodell (LWM)

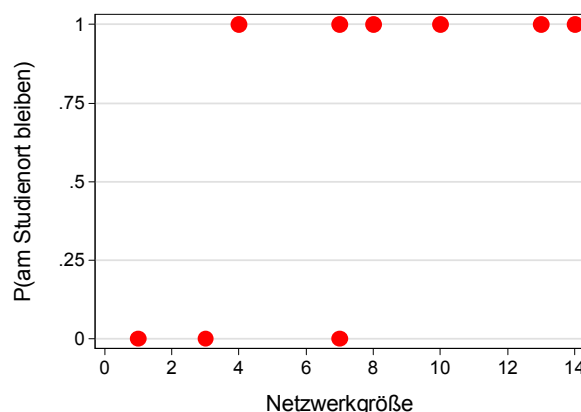
$$P(Y = 1) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = \boldsymbol{\beta}' \mathbf{x}$$

- Fiktives Bsp.: Bleiben am Studienort in Abhängigkeit von der Netzwerkgröße

Y=1: Student will am Studienort nach Beendigung des Studiums bleiben.

Y=0: Student will Studienort nach Beendigung des Studiums verlassen.

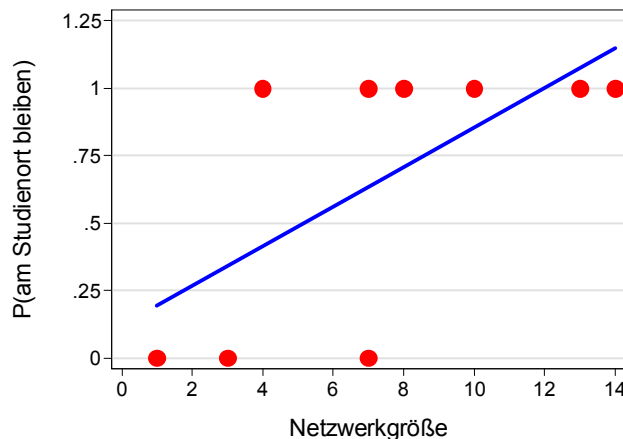
Daten: Logit Intro
Do-File: 8 Logit Intro.do



Das lineare Wahrscheinlichkeitsmodell

. regress bleiben netzgr					Number of obs = 9	
Source	SS	df	MS		F(1, 7) =	5.00
Model	.832853026	1	.832853026		Prob > F =	0.0605
Residual	1.16714697	7	.166735282		R-squared =	0.4164
					Adj R-squared =	0.3331
Total	2	8	.25		Root MSE =	.40833

bleiben	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
netzgr	.073487	.0328806	2.23	0.061	-.0042633	.1512374
_cons	.1195965	.2800758	0.43	0.682	-.5426775	.7818706



Kennt man niemanden (netzgr=0), so ist die Bleibewhs. gleich 12%. Mit jedem zusätzlichen Bekannten steigt sie um 7 Prozentpunkte.

Folge: bei 12 Bekannten liegt die Bleibewhs. über Eins!

Daten: Logit Intro
Do-File: 8 Logit Intro.do

Josef Brüderl, Regressionsanalyse, April 2012

81

Das logistische Modell

- Das LWM führt in manchen Fällen zu unsinnigen Prognosen
 - Außerdem sind die Fehler heteroskedastisch und nicht-normalverteilt
 - Die lineare Funktion unterstellt konstante Effekte. Oft macht es aber Sinn, abnehmende Effekte zu modellieren, wenn P nahe 0 oder 1 ist
- Deshalb verwendet man zur Modellierung von Wahrscheinlichkeiten oft Verteilungsfunktionen $F(\cdot)$, deren Wertebereich $[0,1]$ ist:

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

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

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

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

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

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

- Die Interpretation der Koeffizienten ist nicht einfach (s.u.)
 - Vorzeicheninterpretation geht aber auf alle Fälle: ein positiver Koeffizient besagt, dass mit steigendem X, $P(Y=1)$ zunimmt.

Josef Brüderl, Regressionsanalyse, April 2012

82

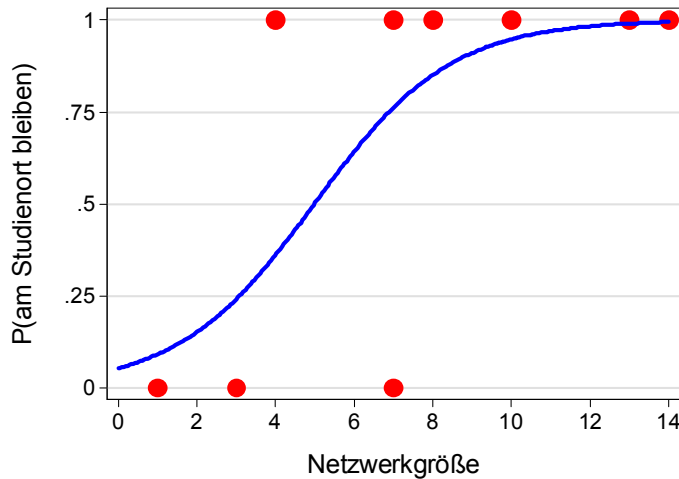
Das logistische Modell

```
. logit bleiben netzgr
```

Logistic regression

Number of obs = 9

bleiben	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
netzgr	.5769341	.3808377	1.51	0.130	-.1694942	1.323362
_cons	-2.876696	2.319757	-1.24	0.215	-7.423335	1.669943



$$\hat{\beta}'x = -2.88 + 0.58 \cdot x$$

Der Effekt der Netzwerkgröße ist positiv, d.h. je mehr Bekannte, desto höher P(Bleiben).

Daten: Logit Intro
Do-File: 8 Logit Intro.do

Interpretation

- Das Logit-Modell hat drei äquivalente Formulierungen:

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

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

Logit: $\ln\left(\frac{P(Y = 1)}{P(Y = 0)}\right) = \beta'x$ „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)

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

- Wahrscheinlichkeitseffekte sind am anschaulichsten, müssen aber mit speziellen Routinen ausgerechnet werden und sind vom Wert der X-Variablen abhängig

Beispiel: Bleiben am Studienort

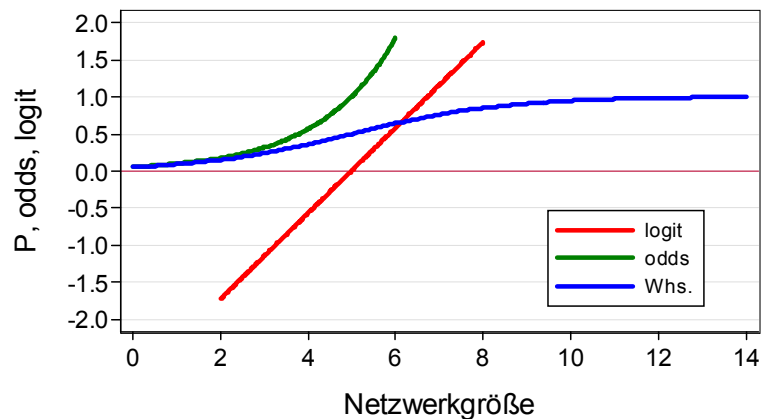
```
. drop _all
. set obs 15
. generate netzgr = _n - 1
. generate logit = -2.876696 + 0.5769341*netzgr
. generate odds = exp(logit)
. generate Whs = exp(logit) / (1+exp(logit))
. format Whs odds logit %6.2f
. list netzgr Whs odds logit, noobs
```

netzgr	Whs	odds	logit
0	0.05	0.06	-2.88
1	0.09	0.10	-2.30
2	0.15	0.18	-1.72
3	0.24	0.32	-1.15
4	0.36	0.57	-0.57
5	0.50	1.01	0.01
6	0.64	1.79	0.58
7	0.76	3.20	1.16
8	0.85	5.69	1.74
9	0.91	10.13	2.32
10	0.95	18.04	2.89
11	0.97	32.12	3.47
12	0.98	57.20	4.05
13	0.99	101.84	4.62
14	0.99	181.34	5.20

$$\text{logit} = -2.88 + 0.58 \cdot x$$

$$\text{odds} = e^{\text{logit}}$$

$$\text{Whs} = \frac{e^{\text{logit}}}{1 + e^{\text{logit}}}$$



Daten: Logit Intro
Do-File: 8 Logit Intro.do

Josef Brüderl, Regressionsanalyse, April 2012

85

Beispiel: Bleiben am Studienort

- Logit-Interpretation
 - Das Logit steigt um 0,58 mit jedem zusätzlichen Bekannten
 - Das ist sehr unanschaulich
- Odds-Interpretation
 - Das Odds (die „Chance“ zu Bleiben) nimmt mit jedem Bekannten um den Faktor $\exp(0,58) = 1,78$ zu
 - Bsp.: $\text{Odds}(6) = 1,79 = 1,01 \cdot 1,78 = \text{Odds}(5) \cdot 1,78$
 - Auch Prozentinterpretation: $(\exp(\beta_j) - 1) \cdot 100$
Das odds nimmt mit jedem Bekannten um 78% zu
 - Falsch: die Whs. erhöht sich um 78%!
 - Odds („Chancen“) sind leider auch ziemlich unanschaulich
- Whs.-Interpretation (DC)
 - Der Effekt hängt von X ab
 - Bsp.: $\text{netzgr}=5$; die Whs. des Bleibens steigt um 14 Prozentpunkte

$$P(Y = 1 | 6) - P(Y = 1 | 5) = \frac{1}{1 + e^{2.88 - 0.58 \cdot 6}} - \frac{1}{1 + e^{2.88 - 0.58 \cdot 5}} = 0.14$$

Josef Brüderl, Regressionsanalyse, April 2012

86

STATA-Beispiel: 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

```
. logit arblos bild
```

Iteration 0: log likelihood = -850.0254 [ln(L₀)]
 Iteration 1: log likelihood = -835.29614
 Iteration 2: log likelihood = -835.20727
 Iteration 3: log likelihood = -835.20726

Logistic regression

Log likelihood = -835.20726 [ln(L₁)]

Number of obs = 1329
 LR chi2(1) = 29.64
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0174

LR – Teststatistik :
 $2(-835 + 850)$

$R^2_{MF} :$
 $\frac{-850 + 835}{-850}$

arblos	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
bild	-.1083058	.0204825	-5.29	0.000	-.1484507 -.0681609
_cons	.7369764	.2693713	2.74	0.006	.2090184 1.264934

z-Wert

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

Josef Brüderl, Regressionsanalyse, April 2012

87

STATA-Beispiel: Interpretation

- Logit-Interpretation
 - Mit jedem Bildungsjahr sinkt das Logit um 0,108
- Odds-Interpretation (Odds-Ratio, OR)

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

```
. logit arblos bild, or
```

arblos	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
bild	.8973532	.01838	-5.29	0.000	.8620425 .9341102

- Mit jedem Bildungsjahr sinkt das Odds („Arbeitslosigkeitschance“) um 10,3%

- Wahrscheinlichkeits-Interpretation (DC)

$$P(Y = 1 | 14) - P(Y = 1 | 13) = \frac{1}{1 + e^{-0.737 + 0.108 \cdot 14}} - \frac{1}{1 + e^{-0.737 + 0.108 \cdot 13}} = 0.315 - 0.339 = -0.024$$

- Ausgehend vom Bildungsmittel (~13) sinkt mit einem weiteren Bildungsjahr die Arbeitslosigkeitswhs. um 2,4 Prozentpunkte

Josef Brüderl, Regressionsanalyse, April 2012

88

STATA-Beispiel: Lineares Wahrscheinlichkeitsmodell

- Das LWM kommt praktisch zum gleichen Ergebnis

```
. regress arblos bild
```

Source	SS	df	MS	Number of obs =	1329
Model	6.39192099	1	6.39192099	F(1, 1327) =	29.16
Residual	290.914324	1327	.219227072	Prob > F =	0.0000
Total	297.306245	1328	.223875185	R-squared =	0.0215
				Adj R-squared =	0.0208
				Root MSE =	.46822

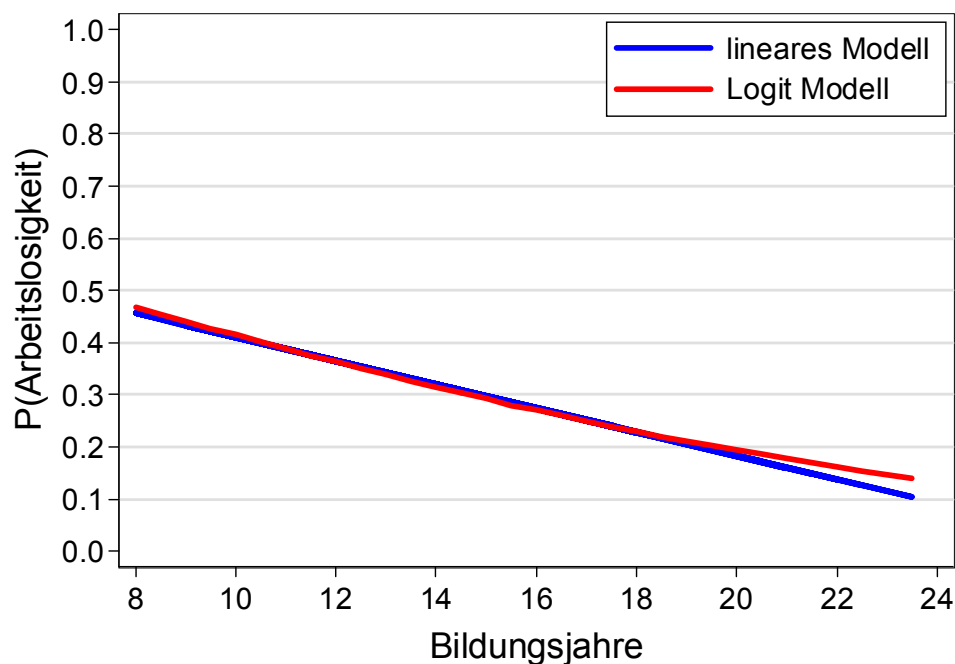
arblos	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
bild	-.0228028	.004223	-5.40	0.000	-.0310872 -.0145183
_cons	.6386421	.0571673	11.17	0.000	.526494 .7507901

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

Josef Brüderl, Regressionsanalyse, April 2012

89

STATA-Beispiel: Vergleich LWM und Logit-Modell



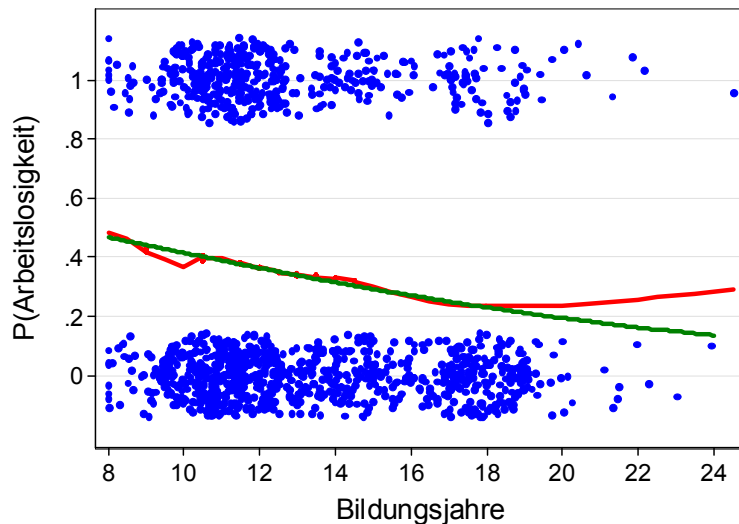
- Liegt $P(Y=1)$ im Intervall $[0,2; 0,8]$, so kommen LWM und Logit-Modell zu identischen Ergebnissen

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

Josef Brüderl, Regressionsanalyse, April 2012

90

STATA-Beispiel: Diagnostik (funktionale Form)



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

Streudiagramm nach Bildung

- Rot: Lowess
- Grün: logistisches Modell

Problem: nur bivariat

Das logistische Modell repräsentiert den Zusammenhang in den Daten ganz gut.
Der Zusammenhang ist annähernd linear, weshalb das LWM hier auch gut passen würde.
Bei hoher Bildung erkennt man eine Abweichung vom Lowess. Die ist allerdings von nur einem Ausreißer verursacht.

STATA-Beispiel: diskretes X

```
. tab arblos ost, col chi2
```

arblos	ost		Total
	0	1	
0	645 75.44	236 49.68	881 66.24
1	210 24.56	239 50.32	449 33.76
Total	855 100.00	475 100.00	1,330 100.00
Pearson chi2(1) = 90.5715 Pr = 0.000			

In diesem Fall reproduziert das logistische Modell nur 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
```

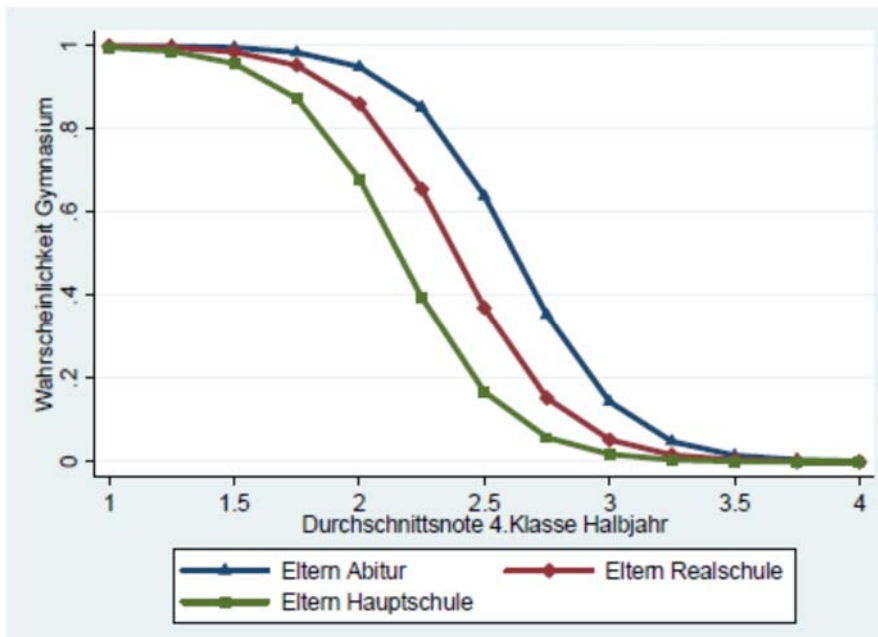
LR chi2(1) = 89.14

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

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 ja 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

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

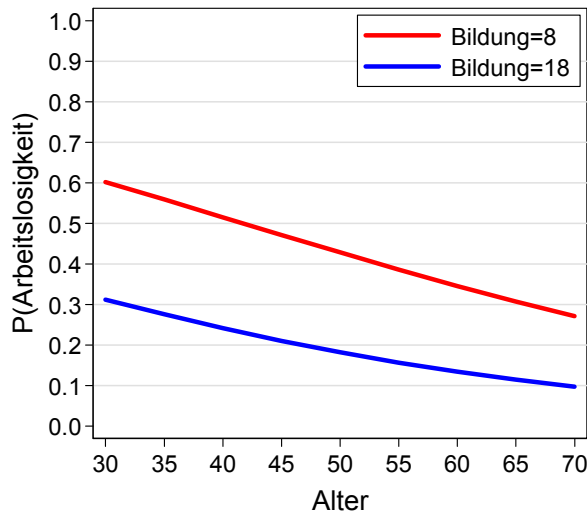
Do-File: 9 Logit Modell.do

- Dieses Modell haben wir oben schon ausführlich erläutert
- 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!)

Conditional-Effect-Plots

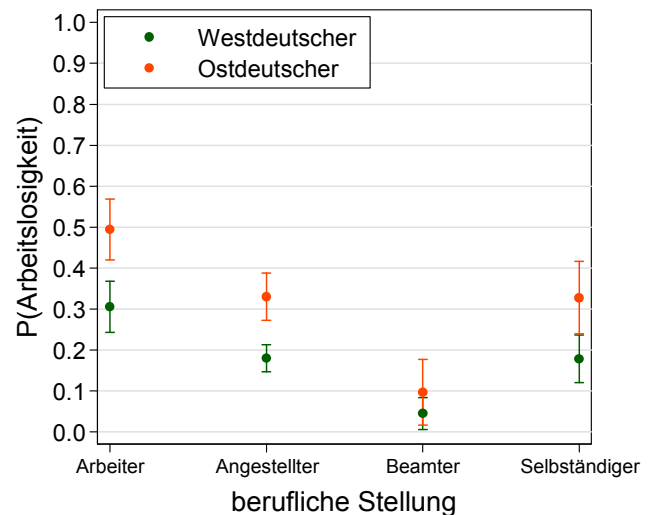
Alter und Bildung

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



Berufliche Stellung und Ost

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



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

Wahrscheinlichkeitsinterpretation

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

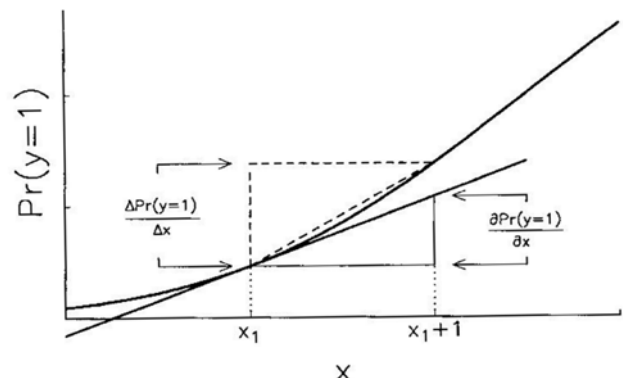
$$\frac{\Delta P(Y = 1)}{\Delta x_j} = P(Y = 1 | \bar{x}, \bar{x}_j + 1) - P(Y = 1 | \bar{x}, \bar{x}_j)$$

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

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

- DC oder ME?

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



Quelle: Long/Freese (2006) S. 169

Effekt am Mittel oder mittleren Effekt?

- Üblicherweise setzt man den Wert der X auf das Mittel und berechnet an dieser Stelle DC und ME
 - Vorgehen von Long/Freeze: `prchange`
- Alternativ kann man aber auch für jede Beobachtung ihre jeweiligen Datenwerte einsetzen und DC bzw. ME berechnen. Anschließend mittelt man.
 - Average Discrete Change (ADC)
 - Average Marginal Effect (AME)
 - Vorgehen von Stata: `margins, dydx(*)`

- Im linearen Modell ist das egal:

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

- Nicht so in nicht-linearen Modellen!

$$\frac{\partial P(Y = 1|\bar{x})}{\partial x_j} \neq \frac{\partial P(Y = 1|x_i)}{\partial x_j}$$

Wahrscheinlichkeitsinterpretation

```
. logit arblo bild alter frau ost
. prchange, help //funktioniert nur mit SPost Ados
```

logit: Changes in Probabilities for arblo

	min->max	0->1	-+1/2	-+sd/2	MargEfct
bild	-0.3833	-0.0268	-0.0284	-0.0861	-0.0284
alter	-0.2702	-0.0078	-0.0083	-0.0713	-0.0083
frau	-0.0084	-0.0084	-0.0084	-0.0041	-0.0084
ost	0.2742	0.2742	0.2604	0.1267	0.2653

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

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

Average marginal effects Number of obs = 1329

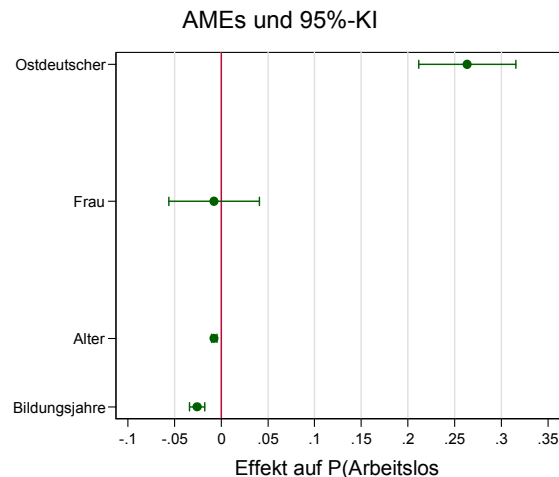
	dy/dx	Delta-method Std. Err.	z	P> z
bild	-0.0260	0.0042	-6.24	0.000
alter	-0.0076	0.0014	-5.34	0.000
1.frau	-0.0077	0.0247	-0.31	0.757
1.ost	0.2636	0.0266	9.91	0.000

Note: dy/dx for factor levels is the discrete change from the base level.

Wahrscheinlichkeitsinterpretation: Vergleich

	prchange	margins	LWM
Bildung	ME -0.0284	AME -0,0260	-0,0248
Alter	ME -0.0083	AME -0,0076	-0,0074
Frau	DC -0.0084	ADC -0,0077	-0,0092
Ost	DC 0.2742	ADC 0,2636	0,2633

- Die Unterschiede zwischen ME/AME bzw. DC/ADC sind gering
 - Das ist nicht immer so (s.u.)
 - Die (neuere) Literatur plädiert für die Verwendung der AME/ADC
- Das lineare Wahrscheinlichkeitsmodell ist nahe am AME
 - Warum nicht gleich das LWM verwenden (s.u.)?



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: Interaktionsterme in nicht-linearen Modellen erhöhen die Komplexität enorm. Um zu verstehen, was man hier eigentlich modelliert, muss man CEPs zur Hilfe nehmen.

Beispiel: Polynomregression mit Alter und Alter² und Interaktion mit „Ost“

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

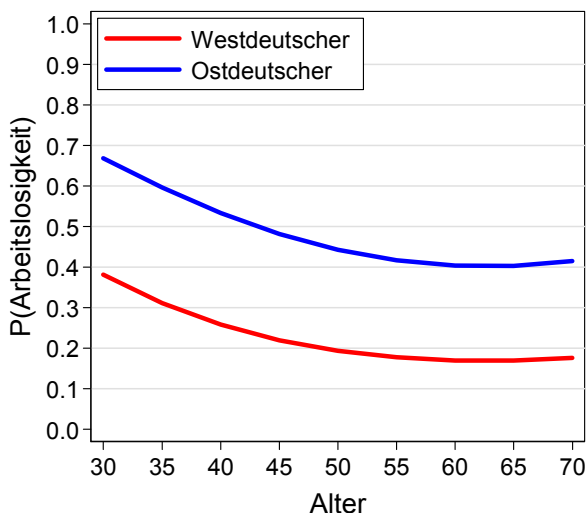
	arblos	Coef.	Std. Err.	z	P> z
bild		-0.1336	0.0220	-6.08	0.000
frau		0.0010	0.1260	0.01	0.993
1.ost		1.8372	3.3693	0.55	0.586
alter		-0.0761	0.1032	-0.74	0.461
c.alter#c.alter		0.0002	0.0012	0.15	0.880
1.ost#c.alter		-0.0786	0.1540	-0.51	0.610
1.ost#c.alter#c.alter		0.0014	0.0017	0.83	0.404
_cons		3.5295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

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

Interaktionseffekte

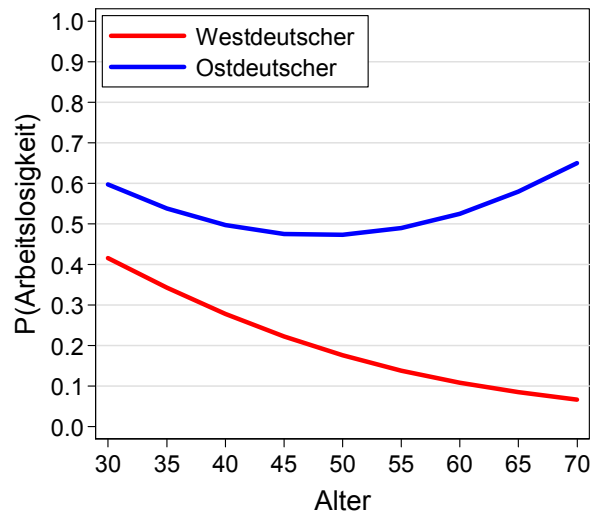
Conditional-Effects-Plot Polynomregression mit Alter²

```
logit arblo bild frau i.ost          ///  
      c.alter c.alter#c.alter  
margins ost, at(alter=(30(5)70))  
marginsplot, noci
```



Conditional-Effects-Plot Polynomreg. + Interaktion mit Ost

```
logit arblo bild frau          ///  
      i.ost##(c.alter c.alter#c.alter)  
margins ost, at(alter=(30(5)70))  
marginsplot, noci
```



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

Interaktionseffekte

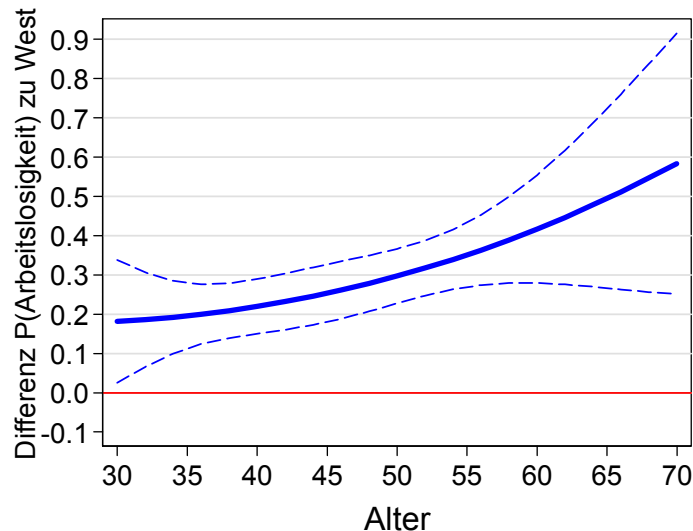
Conditional Marginal-Effects-Plot

```
logit arbloos 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 LWM ist eine gute Alternative und bekanntlich ist dort die Interpretation von Interaktionseffekten deutlich einfacher!

Konditionaler Marginaler Effekt von 'Ost' und 95%-KI



Weiterführende Literatur:

Ai/Norton (2003) Interaction terms in logit and probit models. Ec. Letters 80: 123-129

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

Alternative Herleitung 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

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

- wobei $\alpha = \frac{a}{\sigma}$ und $\beta = \frac{b}{\sigma}$
- Je nach Annahme über $F(\cdot)$ kommt man zu Probit oder Logit, wobei die Varianz von ε festgelegt wird (deshalb der Skalierungsfaktor σ)
- Die Abhängigkeit der Koeffizienten vom (unbekannten) Skalierungsfaktor ist normalerweise kein Problem, da die Wahrscheinlichkeiten trotzdem eindeutig schätzbar sind

Probleme mit Vergleichen

- Will man Koeffizienten vergleichen (über Modelle, Gruppen, Kohorten, Länder, etc.), so muss man aufpassen
 - In den Modellen/Gruppen sind normalerweise die Skalierungsfaktoren unterschiedlich (unterschiedliche unbeobachtete Heterogenität)
 - Die geschätzten Logit-/Probit-Koeffizienten werden sich allein deshalb unterscheiden
 - Vergleiche von Koeffizienten zwischen Modellen/Gruppen sind somit von unbeobachteter Heterogenität konfundiert!
 - Abhilfe
 - Man kann zeigen, dass die AME dieses Problem nicht haben
 - Ebenso kann man zeigen, dass das LWM die AME konsistent schätzt. Logit/Probit ist also nur ein komplizierter Umweg, um die AME zu schätze!
- Viel spricht also dafür, statt Logit/Probit das LWM zu verwenden!
- Dies gilt insbesondere, wenn man Vergleiche anstellen und/oder Interaktionen betrachten will

Probleme mit Vergleichen

Simulation von Mood (2010):

Logit Modell für die Wahrscheinlichkeit des Übergangs zur Universität. IQ und Geschlecht sind unkorreliert!

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

Auspurg/Hinz (2011)

zeigen bei einer Analyse zur Veränderung der Bildungsungleichheit über Geburtskohorten, dass man mit OR und AME zu unterschiedlichen Schlussfolgerungen kommt. Auch sie argumentieren, dass die OR durch unbeobachtete Kohorten-Heterogenität konfundiert sind. Sie schließen daraus, dass man für Vergleiche über Kohorten die AMEs verwenden sollte.

Weiterführende Literatur:

- Mood (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.
- Auspurg/Hinz (2011) Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen. ZfS 40: 62-73.

Logit vs. Probit

	Logit	Probit
Bildungsjahre	-0.131*** (-5.96)	-0.077*** (-6.06)
Alter	-0.038*** (-5.15)	-0.023*** (-5.32)
Frau	-0.039 (-0.31)	-0.023 (-0.30)
Ostdeutscher	1.221*** (9.68)	0.747*** (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



Kapitel VI: Regressionsmodelle für ordinale Outcomes



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 $m = 1, \dots, J$. Das latente Variablen Modell lautet:
$$y^* = \beta'x + \varepsilon$$
 - Das Schwellenwertmodell lautet nun mit $J+1$ Schwellen τ
$$y = m, \quad \text{falls } \tau_{m-1} \leq y^* < \tau_m$$
 - wobei $\tau_0 = -\infty$ und $\tau_J = \infty$. Es sind also $J-1$ Schwellen zu schätzen. Damit gilt:
$$P(y = m) = P(\tau_{m-1} \leq y^* < \tau_m)$$
$$\Rightarrow P(y = m) = F(\tau_m - \beta'x) - F(\tau_{m-1} - \beta'x)$$

Ordinales Logit/Probit

- Die Wahl der Verteilungsfunktion $F(\cdot)$ komplettiert das Modell
 - Standard-Normalverteilung: ordinales Probit
 - Standard-Logistische-Verteilung: ordinales Logit
 - z.B. ordinales Logit für $J=3$:
$$P(y = 1) = \Lambda(\tau_1 - \beta'x)$$
$$P(y = 2) = \Lambda(\tau_2 - \beta'x) - \Lambda(\tau_1 - \beta'x)$$
$$P(y = 3) = 1 - \Lambda(\tau_2 - \beta'x)$$
 - 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: 10 OLogit.do

STATA Beispiel: „rechte“ politische Einstellung

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

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

```
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

Ein Modellvergleich

	OLogit	OProbit	OLS
Alter	0.007** (2.63)	0.004** (2.69)	0.002** (2.70)
Bildungsjahre	-0.036* (-2.18)	-0.019* (-2.08)	-0.011* (-2.06)
Einkommen / 1000	0.045 (1.11)	0.026 (1.09)	0.015 (1.09)
Frau	-0.368*** (-4.15)	-0.216*** (-4.22)	-0.123*** (-4.23)
Ostdeutscher	-0.568*** (-6.24)	-0.329*** (-6.25)	-0.188*** (-6.29)
Konstante			2.094*** (25.61)
Cutpoint1	-1.663*** (-6.49)	-0.977*** (-6.75)	
Cutpoint2	1.013*** (3.98)	0.642*** (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: 10 OLogit.do

Die Annahme paralleler Regressionen

- Das ORM impliziert eine Annahme

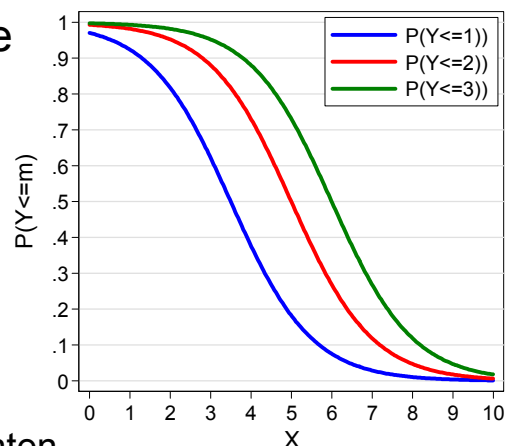
- Beispiel: ordinale Logit, J=4

$$P(y \leq 1) = \Lambda(\tau_1 - \beta'x)$$

$$P(y \leq 2) = \Lambda(\tau_2 - \beta'x)$$

$$P(y \leq 3) = \Lambda(\tau_3 - \beta'x)$$

- Daran erkennt man, dass das ORM äquivalent ist zu drei binären Logits, wobei β identisch ist (nur die Konstanten sind verschieden) [parallel regression/line assumption]
- Testprinzip: man schätzt die binären Logits und testet, ob die $\hat{\beta}$ gleich sind
- Was tun bei Verletzung der Annahme?
 - Problematische Variablen weglassen (normalerweise keine gute Idee)
 - Alternative ORM: Generalisiertes ordinale Logit (s.u.)
 - 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: 10 OLogit.do

Interpretation

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

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

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

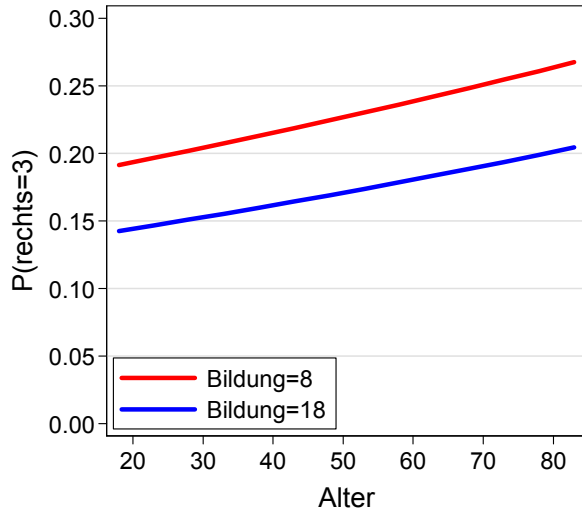
- Es gibt also $J - 1$ Odds. Aufgrund der Annahme paralleler Regressionen sind die Effekte für alle Odds identisch (deshalb auch oft: „Annahme proportionaler Odds“)
- $\exp(-\beta_j)$ ist der multiplikative Effekt auf die Odds kleiner/größer
- Die Interpretation ist einfacher, wenn man die Odds als größer/kleiner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

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

Conditional-Effect-Plots

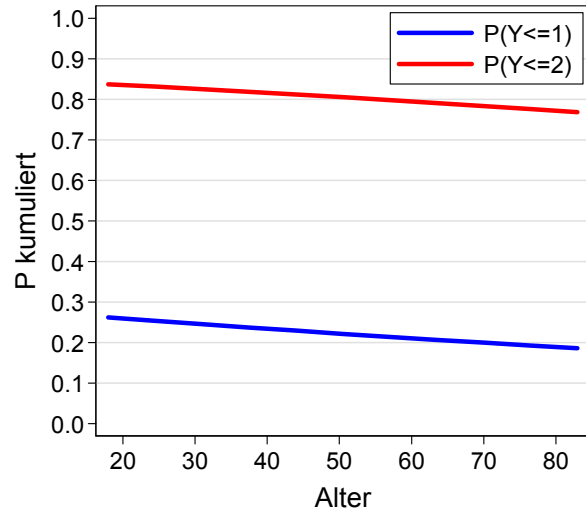
Alter und Bildung

```
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: 10 OLogit.do

Wahrscheinlichkeitsinterpretation

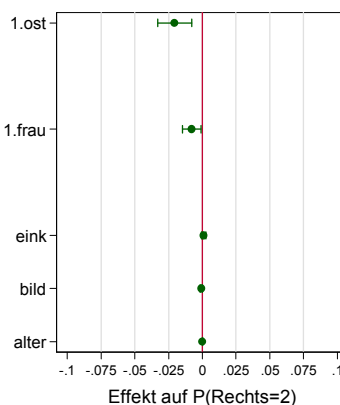
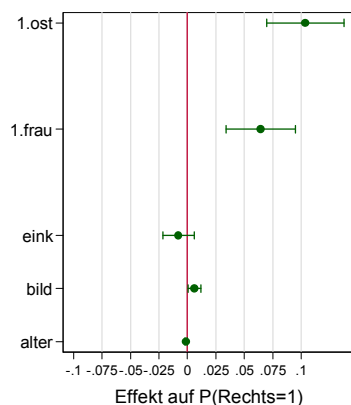
- Der Marginal Effekt (ME) im ORM ist

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

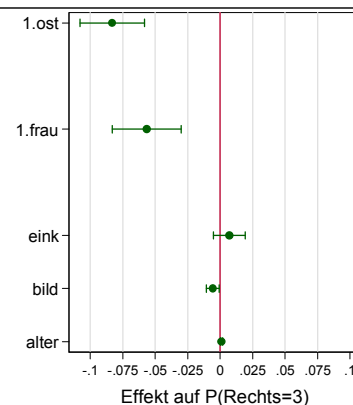
$$= \beta_j [f(\tau_{m-1} - \beta' \bar{x}) - f(\tau_m - \beta' \bar{x})]$$

- 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 CEPs (bei metrischen Variablen)
- Weiterhin gibt es natürlich auch DC, AME und ADC
 - Das oben Gesagte gilt auch hier!

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



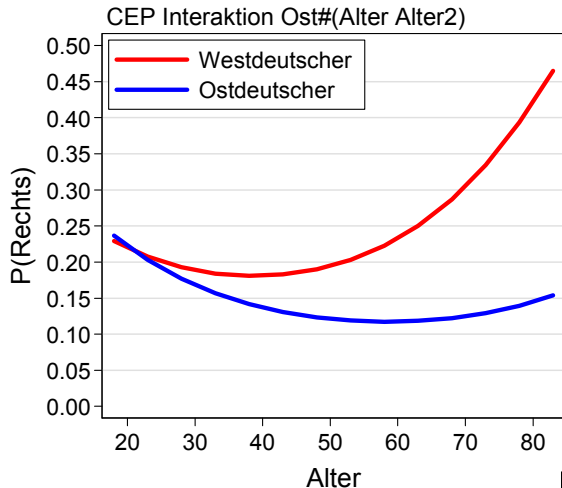
AMEs für unser Beispiel



Interaktionseffekte

Conditional-Effects-Plot Polynomreg. + Interaktion mit Ost

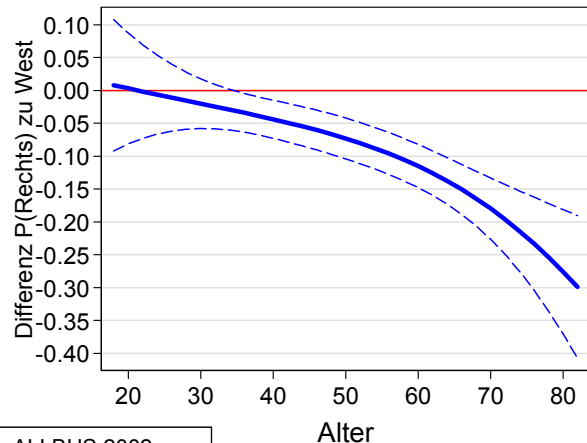
```
ologit rechts bild eink frau      ///
      i.ost##(c.alter c.alter#c.alter)
margins ost, at(alter=(18(5)83))  ///
      predict(outcome(3))
marginsplot, noci
```



Conditional Marginal-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)
```

Konditionaler Marginaleffekt von 'Ost' und 95%-KI



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

Generalisiertes ordinales Logit

- Generalisierung des ordinalen Logit (gologit)
 - Ist die Annahme proportionaler Odds verletzt, so kann man dieses Modell verwenden. Es erlaubt J-1 Koeffizientenvektoren

$$P(y = 1) = \Lambda(\tau_1 - \beta_1'x)$$

$$P(y = 2) = \Lambda(\tau_2 - \beta_2'x) - \Lambda(\tau_1 - \beta_1'x)$$

$$P(y = 3) = 1 - \Lambda(\tau_2 - \beta_2'x)$$

- Damit ist man im Prinzip beim multinomialen Modell
- Der Clou ist aber nun, dass es Mischformen gibt:
„partial proportional odds“
 - Variablen, für die proportionale Odds gelten, bleiben restringiert
 - Variablen, die keine proportionalen Odds haben, dürfen unterschiedliche Koeffizienten haben
- Beispiel
 - Wir haben oben gesehen, dass „bild“ keine proportionalen Odds hat. Also sollten wir für „bild“ zwei Koeffizienten schätzen.

Generalisiertes ordinales Logit

```
. gologit2 rechts alter bild eink frau ost, autofit //funktioniert nur wenn
                                                    „gologit2“ geladen ist
```

Testing parallel lines assumption using the .05 level of significance...

Step 1: Constraints for parallel lines imposed for eink (P Value = 0.2279)

Step 2: Constraints for parallel lines imposed for alter (P Value = 0.2021)

Step 3: Constraints for parallel lines imposed for ost (P Value = 0.1308)

Step 4: Constraints for parallel lines imposed for frau (P Value = 0.0915)

Step 5: Constraints for parallel lines are not imposed for
bild (P Value = 0.01945)

Wald test of parallel lines assumption for the final model:

```
( 1) [Links]eink - [Mitte]eink = 0
( 2) [Links]alter - [Mitte]alter = 0
( 3) [Links]ost - [Mitte]ost = 0
( 4) [Links]frau - [Mitte]frau = 0
```

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

```

      chi2( 4) =      8.24
Prob > chi2 =      0.0830
```

An insignificant test statistic indicates that the final model does not violate the proportional odds/ parallel lines assumption

If you re-estimate this exact same model with gologit2, instead of autofit you can save time by using the parameter

```
pl(eink alter ost frau)
```

Generalisiertes ordinales Logit

Generalized Ordered Logit Estimates						Daten: ALLBUS 2002	
Log likelihood = -2052.9152						Do-File: 10 OLogit.do	
(1) [Links]eink - [Mitte]eink = 0		Number of obs		=		2137	
(2) [Links]alter - [Mitte]alter = 0		Wald chi2(6)		=		78.69	
(3) [Links]ost - [Mitte]ost = 0		Prob > chi2		=		0.0000	
(4) [Links]frau - [Mitte]frau = 0		Pseudo R2		=		0.0191	

rechts		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

Links							
	alter	0.0068	0.0026	2.64	0.008	0.0018	0.0119
	bild	-0.0587	0.0190	-3.10	0.002	-0.0959	-0.0216
	eink	0.0424	0.0397	1.07	0.286	-0.0354	0.1202
	frau	-0.3637	0.0888	-4.10	0.000	-0.5377	-0.1898
	ost	-0.5684	0.0911	-6.24	0.000	-0.7470	-0.3898
	_cons	1.9558	0.2851	6.86	0.000	1.3970	2.5145

Mitte							
	alter	0.0068	0.0026	2.64	0.008	0.0018	0.0119
	bild	-0.0074	0.0202	-0.37	0.714	-0.0469	0.0321
	eink	0.0424	0.0397	1.07	0.286	-0.0354	0.1202
	frau	-0.3637	0.0888	-4.10	0.000	-0.5377	-0.1898
	ost	-0.5684	0.0911	-6.24	0.000	-0.7470	-0.3898
	_cons	-1.3706	0.2959	-4.63	0.000	-1.9507	-0.7906

Generalisiertes ordinales Logit

- Interpretation

- Die restringierten Koeffizienten sind im Vergleich zu ologit praktisch unverändert. Auch die ME/DC sind praktisch unverändert.
- Beim unrestringierten Koeffizienten von „bild“ sehen wir in der Tabelle „Links“ -0,059 in „Mitte“ -0,007 (im ologit -0,036)
 - Was das in Bezug auf Odds bedeutet, verstehen nur Experten
- Aber mit „mfx“ können wir die ME berechnen

MEs für die Variable „bild“		
	ologit	gologit
P(Links)	0,0063*	0,0102*
P(Mitte)	-0,0007	-0,0091*
P(Rchts)	-0,0055*	-0,0011

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

- Ologit sagt, dass mit einem Bildungsjahr (ausgehend vom Mittelwert aller Kovariaten) der Anteil „Links“ auf Kosten von „Rechts“ steigt
- Gologit sagt, dass mit einem Bildungsjahr der Anteil „Links“ auf Kosten von „Mitte“ steigt
 - Oder anders: der Effekt von Bildung auf „Rechts“ ist nicht so stark



Kapitel VII: Regressionsmodelle für multinomiale Outcomes



Regression für multinomiales Y

- Kategoriale aV mit mehr als zwei Ausprägungen
 - Auch für ordinal aV, wenn „parallele Regression“ verletzt
- Multinomiale logistische Regression
 - Y sei multinomial mit Ausprägungen $m = 1, \dots, J$. Ein geeignetes nicht-lineares Wahrscheinlichkeitsmodell zur Modellierung ist

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

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

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

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

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

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 eine Reparametrisierung.
 - Für $J = 2$ erhält man das binäre Logit als Spezialfall.
 - Das multinomiale Logit kann auch aus einem latenten 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
Log likelihood = -2729.6341
Number of obs = 2007
LR chi2(12) = 307.66
Prob > chi2 = 0.0000
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: 11 MLogit.do

Test der Signifikanz der Variablen

- Haben die drei Variablen einen signifikanten Einfluss auf die Wahlabsicht?
 - Die Signifikanztest der einzelnen Logit-Koeffizienten führen zu widersprüchlichen Ergebnissen.
 - 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: 11 MLogit.do

Test zum Kombinieren von Kategorien

- Häufig steht man vor dem Problem, ob man nicht einzelne Kategorien zusammenfassen könnte
 - 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: 11 MLogit.do

Unabhängigkeit von irrelevanten Alternativen

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

$$\frac{P(y = m)}{P(y = k)} = \exp\{x(\beta_m - \beta_k)\}$$

- 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: 11 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
 - $\text{Odds}(\text{Ost}) = 0,2/0,1 = 2$ $\text{Odds}(\text{West}) = 0,3/0,3 = 1$
 - Die Odds-Ratio ist also 2. Es wäre nun aber eine krasse Fehlinterpretation, wenn man schlussfolgern würde, die Ossi wählen doppelt so oft SPD!

- Odds Interpretation

- ist die OR für Variable X_j für $P(y=m) / P(y=1)$
- Die Odds Interpretation ist schwer zu verstehen

Bsp. Ost(Grüne):

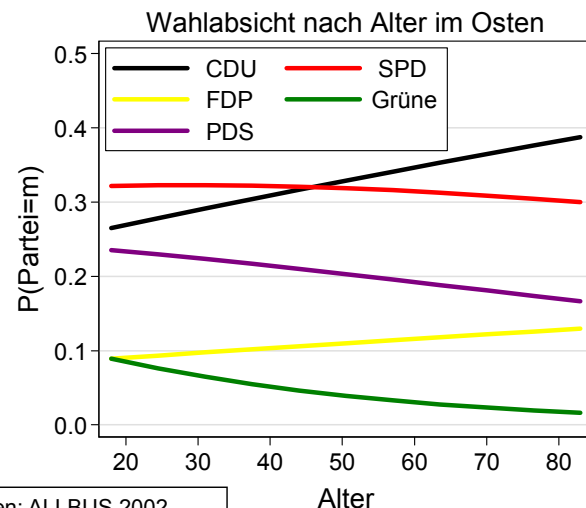
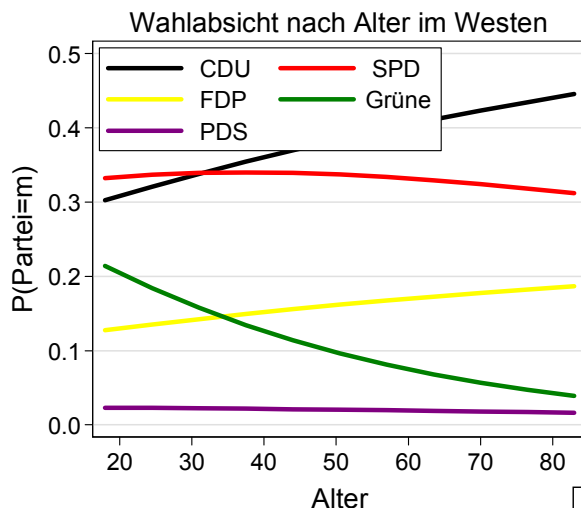
$\beta = -0,7410$, OR = 0,48
Die Odds Grüne/CDU sind im Osten um die Hälfte kleiner.

Conditional-Effect-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.



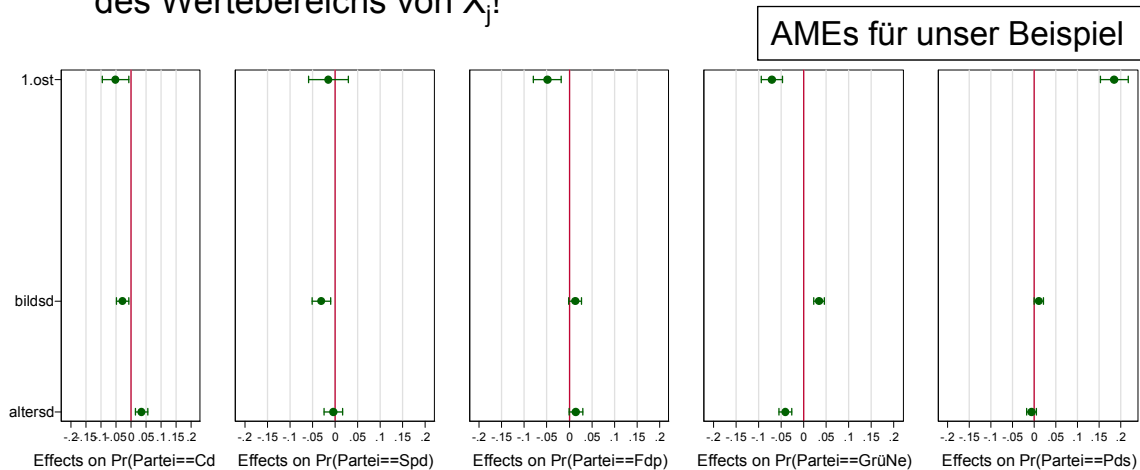
Daten: ALLBUS 2002
Do-File: 11 MLogit.do

Wahrscheinlichkeitsinterpretation

- Die Marginal Effekte (ME) im multinomialen Logit sind

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

- 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 !



Josef Brüderl, Regressionsanalyse, April 2012

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

135

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 für Ost und West getrennt abgebildet
 - Das Modell erlaubt offensichtlich gänzlich unterschiedliche Alterseffekte!

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

Ein Modell mit Interaktionseffekten!

