

Analysen politikwissenschaftlicher Datensätze mit Stata

JOHANNES
GUTENBERG
UNIVERSITÄT
MAINZ

Sitzung 8: „The Public’s Conditional
Response to Supreme Court decisions“
(Johnson/Martin 1998)

Vorbereitung

- Bitte laden Sie den Datensatz
z: \daten\jm\webster.dta

Kontext

- Supreme Court (SC, Oberstes Gericht der USA) kann (u.a.) Gesetze des Bundes oder der Staaten für verfassungswidrig erklären und ist oberstes Appellationsgericht
- Häufig letzte Instanz in gesellschaftlichen Streitfragen (Rassentrennung, Abtreibung, Todesstrafe etc.), Legislative und Exekutive an Urteile gebunden
- Besetzt mit neun Richtern auf Lebenszeit

Kontext

- Diese werden vom Präsidenten vorgeschlagen und vom Senat gewählt
- Wichtige Fälle unter dem Namen der Konfliktparteien bekannt
- Richter nicht an frühere Entscheidungen gebunden

Kontext: Abtreibungsrecht

- Roe v. Wade (1973):
 - Abtreibung aus nicht-medizinischen Gründen im zweiten Trimester legal
 - Gesetze der Bundesstaaten, die dem widersprechen, verfassungswidrig

Kontext: Abtreibungsrecht

- Webster v. Reproductive Health Service (1989)
 - Ein Gesetz des Staates Missouri, nach dem das Leben mit der Konzeption beginnt
 - und die Finanzierung von Abtreibungen aus öffentlichen Mitteln sowie die Durchführung von Abtreibungen in öffentlichen Kliniken verbietet,
 - ist *nicht* verfassungswidrig

Kontext: Todesstrafe

- Furman v. Georgia (1972)
 - Anwendung und Durchführung der Todesstrafe in Georgia verstoßen gegen das verfassungsmäßige Verbot „grausamer und ungewöhnlicher Bestrafung“
 - Todesstrafe in Georgia und einer Reihe anderer Staaten vorläufig ausgesetzt.

Kontext: Todesstrafe

- Gregg v. Georgia (1976)
 - Todesstrafe nicht in grundsätzlichem Widerspruch zur Verfassung
 - überarbeitete Gesetze verfassungsmäßig
- McCleskey v. Kemp (1987)
 - Statistische Belege dafür, daß Schwarze bei gleicher Sachlage häufiger zum Tode verurteilt werden, stehen einer Anwendung der Todesstrafe nicht entgegen
 - wenn im konkreten Fall keine Diskriminierung nachweisbar ist

Theoretische Grundannahmen

- Leitfrage: Wie wirken sich Entscheidungen des Supreme Court auf die Einstellungen zu politischen Streitfragen (issues) aus?

Theoretische Grundannahmen

- Bisherige Ansätze:
 - „Positive Response Theory“ (Dahl 1957):
 - Bevölkerung akzeptiert Legitimität des SC als letzte Instanz
 - Einstellungen passen sich deshalb den Urteilen an
 - „Structural Response Theory“ (Franklin/Kosaki 1989)
 - Einstellungen zu *wichtigen Fragen* vor allem vom sozialen Kontext beeinflusst
 - Nach wichtigen Urteilen öffentliche Diskussion → größere Homogenität in sozialen Gruppen → Polarisierung bzw. Kristallisierung
 - Nicht notwendigerweise Veränderung der aggregierten Einstellungen

Theoretische Grundannahmen

- „Conditional Response Hypothesis“
 - Erweiterung der „Structural Response Theory“ (SRT) um „Elaboration Likelihood Model“
 - Bei einer ersten wichtigen Entscheidung kommt es wie von SRT angenommen zu einer Kristallisierung
 - weil der SC eine Quelle mit hoher Glaubwürdigkeit und Legitimität darstellt
 - Wenn Menschen zu einer wichtigen Frage eine Einstellung herausgebildet haben, ändern neue Informationen (auch von glaubwürdigen Quellen) daran nicht mehr viel
 - Deshalb haben Folgeentscheidungen des SC selbst dann, wenn sie von früheren Entscheidungen abweichen, keinen großen Effekt mehr

Hypothesen

- Nach einer ersten wichtigen Entscheidung
 - verstärken sich die Einstellungsunterschiede zwischen sozialen Gruppen
 - kommt es generell zu einer Verschiebung in der öffentlichen Meinung
- Bei Folgeentscheidungen
 - bleiben Einstellungsunterschiede zwischen Gruppen weitgehend konstant
 - ändert sich kaum etwas an der Gesamtverteilung

Design/Daten

- Webster als Beispiel für eine Folgeentscheidung (Rode bereits von Franklin/Kosaki analysiert) im Bereich Abtreibung
- Furman als Beispiel für eine erste wichtige Entscheidung im Bereich Todesstrafe
- Gregg und McCleskey als Beispiele für Folgeentscheidungen
- Quasi-Experimentelles Trend-Design: Verglichen werden jeweils Querschnittsbefragungen vor und nach den Urteilen

Modellierung

- Ordinale bzw. dichotome Probit-Modelle
- Sozialstrukturelle Variablen als Indikatoren für Gruppenmitgliedschaften
- Für jeden Fall werden Befragungen vor/nach der Entscheidung in einem Datensatz zusammengefaßt

Modellierung

- Dummy codiert Zeit nach dem Urteil
- Interaktionen zwischen Dummy und Sozialstrukturvariablen erfassen Verstärkung/Abschwächung der Effekte (=Polarisierung/Depolarisierung)
- Modelle werden einmal mit und einmal ohne Interaktionen geschätzt → Likelihood Ratio-Test für die *gemeinsame* Signifikanz aller Interaktionen

Mögliche Kritikpunkte

- Annahme, daß SC als besonders legitim/glaubwürdig gilt und deshalb Einstellungen beeinflussen kann, nicht direkt modelliert (vgl. FN 1)
- D.h. Vorher-Nachher-Effekte nicht unbedingt auf Urteil selbst zurückzuführen (Drittvariablen)
- Hypothese 2 im Grunde einfach Nullhypothese zu Hypothese 1 (und als solche leicht beizubehalten)
- Kein Paneldesign, d.h. keine individuellen Veränderungen sichtbar (aber Untersuchung auf Individualabene!)

Ergebnisse

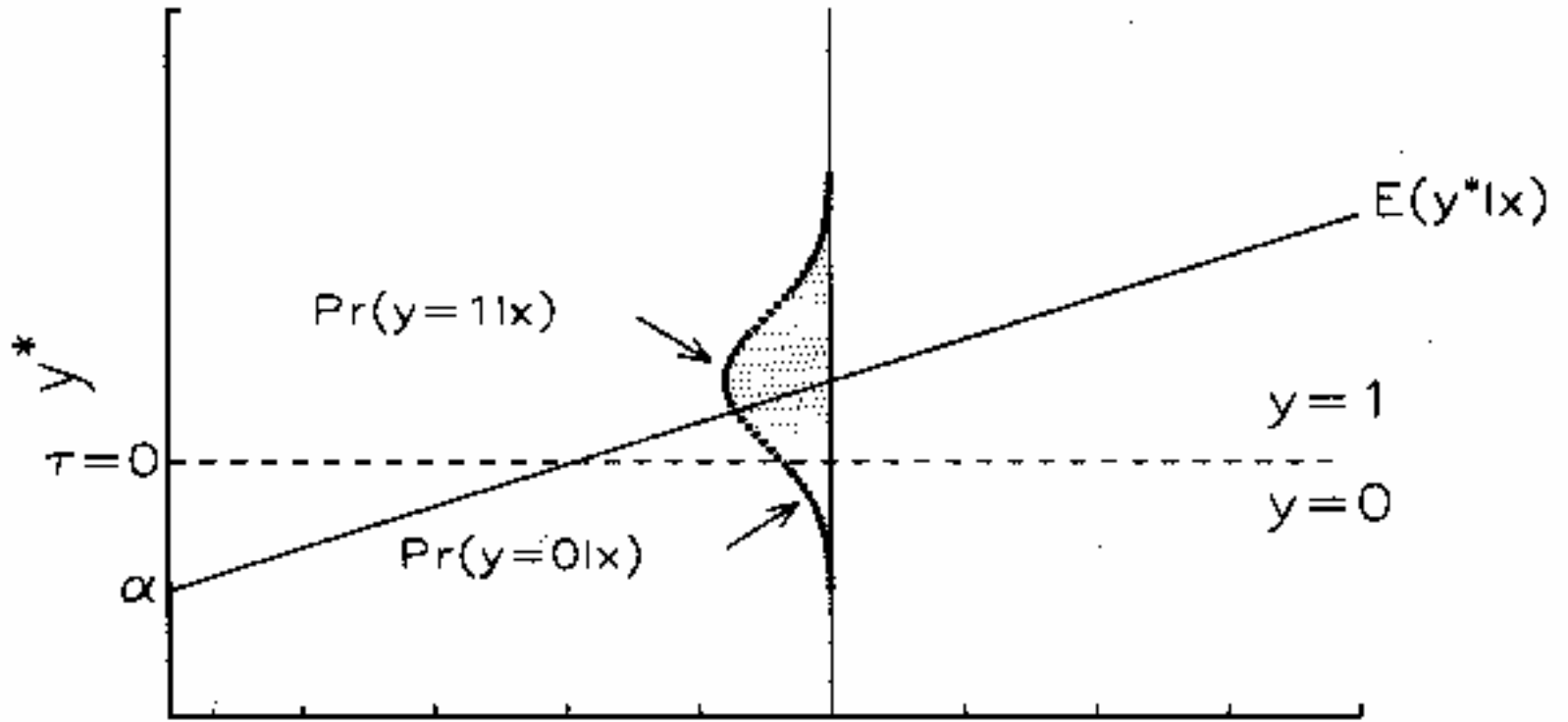
- „Conditional Response Hypothesis“ im wesentlichen bestätigt
 - SC dann besonders einflußreich, wenn er erstmals in einer Sache urteilt
 - Keine signifikanten Effekte in den anderen Fällen

Replikation: Tabelle 1

- Der Datensatz enthält zwei ordinale abhängige Variablen (discret, health)
 - Diese entsprechen der Zahl der Umstände, unter denen die Respondenten Abtreibungen für akzeptabel halten (0, 1, 2 oder mehr)
 - und zwar mit Blick auf medizinische Gründe (health) und andere Gründe (discret)
- Diese analysieren Johnson/Martin mit ordinalen Probit-Modellen
- Die (gemeinsame) Signifikanz der Interaktionsterme prüfen sie mit einem LR-Test

Probit-Modelle

- Logit-Modelle können als Modelle mit einer latenten, unbeschränkten intervallskalierten Variablen aufgefaßt werden, die nur kategorial beobachtet werden kann (vgl. letzte Stunde)
- $y^* = a + bx + e$
 - mit $y=0$ wenn $y^* \leq 0$
 - mit $y=1$ wenn $y^* > 0$



Quelle: Long/Freese 2001

Probit-Modelle

- Die Varianz der Störgröße e
 - ist nicht bekannt
 - kann nicht aus den Daten geschätzt werden, da die latente Variable nicht direkt beobachtet werden kann
 - muß deshalb fixiert werden
- Im Prinzip kann hier von einer beliebigen Verteilung mit einem Mittelwert von 0 ausgegangen werden
- Eine logistische Verteilung der Störgröße mit einer Varianz von $\sim 3,29$ (entspricht fast einer t -Verteilung mit df 7) führt zum Logit-Modell
- Eine Standardnormalverteilung führt zum Probit-Modell
- Das Probit-Modell ist das historisch ältere
- Die Form des Logit-Modells (auf der Ebene der Wahrscheinlichkeiten) ist wesentlich einfacher

Probit-Modelle

- Die Parameter beider Modelle werden mit der ML-Methode bestimmt
- Die Form der S-Kurven unterscheidet sich etwas, weil die logistische Verteilung von einer etwas höheren Wahrscheinlichkeit größerer Stör-Werte ausgeht
- In den allermeisten Fällen liefern Probit- und Logit-Modelle sehr ähnliche Ergebnisse, wobei sich die Koeffizienten in der Regel um einen Faktor von ~ 1.7 unterscheiden
- Wenn man Zugriff auf moderne Computerprogramme hat, ist die Entscheidung zwischen Logit und Probit eine Frage von Geschmack und Tradition

ML-Schätzung / LR-Test

- Parameter für lineare Regression werden mit OLS geschätzt – algebraische Lösung
- Parameter von Logit-, Probit- und ähnliche Modellen werden nach der Maximum Likelihood-Methode bestimmt
- Likelihood ~ Plausibilität / Wahrscheinlichkeit, aber nicht auf 0-1 beschränkt (Wahrscheinlichkeit * datenspezifische Konstante)
- Likelihood ist eine Funktion von Parameterschätzung und empirischen Daten

ML-Schätzung / LR-Test

- Gesucht werden solche Parameterschätzungen, die die Likelihood-Funktion maximieren, d.h. in Anbetracht der Daten am plausibelsten sind
- Der Stichprobenmittelwert ist die ML-Schätzung für den wahren Mittelwert in der GG; OLS-Schätzung sind die ML-Schätzung für die wahren Parameter des linearen Regressionsmodells in der GG
- Für die meisten Modelle existiert keine algebraische Lösung bzw. diese ist zu kompliziert

ML-Schätzung / LR-Test

- → iterative Lösung, d.h. Maximierung der Likelihood-Funktion durch (systematisches) „Probieren“, bis ein Konvergenzkriterium erreicht ist
- In der Praxis wird meist der Logarithmus der Likelihood (Log-Likelihood) maximiert
- Für Logit-/Probit-Modelle liegt der Wert der Likelihood-Funktion immer zwischen 0 und 1 → Log-Likelihood immer negativ

ML-Schätzung / LR-Test

- Differenzen in der Log-Likelihood zwischen Modellen können für Signifikanztests genutzt werden
 - Wenn Modelle „geschachtelt“ (nested) sind, d.h. alle Parameter des Modells A auch in Modell B enthalten sind
 - Ist die Differenz der Log-Likelihoods $\times -2$ Chi-Quadrat verteilt
 - Mit der Zahl der zusätzlichen Parameter als Freiheitsgrade
 - Gemeinsamer Signifikanztest: Wie wahrscheinlich ist die Veränderung in der Log-Likelihood, wenn *alle* zusätzlichen Parameter in der GG den Wert null haben?
 - Vergleich eines Modells mit dem Nullmodell, daß nur die Konstante enthält (von Stata als LR chi2 ausgegeben, entspricht F-Test für lineare Regression)
 - Vergleich konkurrierender Modelle, wenn diese geschachtelt sind (im Text: Modelle mit/ohne Interaktionsterme)

Replikation Tabelle 1

- `tab disc`
- Schätzung von Probit-Modellen mit dem Kommando `probit / oprobit`
- `oprobit disc` → „Nullmodell“
 - Reproduziert empirische Verteilung
 - Keine Iteration notwendig
- `oprobit disc postdum educ posteduc gender postgen race postrace cath postcath attchrch`
 - $LR\ chi2(10) = 258 = -2 * \text{Differenz der Loglikelihoods bei zehn zusätzlichen Parameterschätzungen}$
 - Extrem unwahrscheinlich, wenn alle zehn Parameter tatsächlich null (hätte aber auch niemand vermutet)
 - Iteration konvergiert sehr schnell

Replikation Tabelle 1

- Stimmen alle Werte überein?
 - Für cutpoint/Konstante wurde eine andere Parametrisierung gewählt
 - Constant = $_{b[_cut1]} * -1$
 - Cut Point = $(_{b[_cut1]} - _{b[_cut2]}) * -1$
 - Beim Wert für den Dummy haben sich Johnson/Martin um den Faktor 10 vertan
 - Warum gibt es keine Interaktion für Kirchgang?
- outreg using modell1.txt
 - Öffnen Sie diese Datei (u:\stataseminar\modell1.txt) in Word
 - Markieren Sie alles und wählen Sie Tabelle → Umwandeln → Text in Tabelle
 - outreg ist ein Zusatzprogramm und kennt sehr viele nützliche Optionen. Außerdem werden Variablenlabel berücksichtigt (hier nicht relevant)
 - Schließen Sie Word wieder (wichtig)

Replikation Tabelle 1

- Was machen die Interaktionseffekte (gibt es eine Polarisierung?)
 - jeder einzelne Effekt ist nicht-signifikant
 - trotzdem ist es theoretisch möglich, daß alle *gemeinsam* signifikant von null verschieden sind
- LR-Test
 - mit `est store modell1full` die Ergebnisse des Modells speichern
 - (mit `fitstat,saving(m1f)` außerdem die Anpassungsmaße speichern)
 - Modell schätzen, in dem Interaktionen auf null gesetzt sind (am einfachsten: Interaktionen weglassen!):
`oprobit disc postdum educ gender race cath attchrch`
 - `lrtest modell1full,stats` vergleicht gespeichertes Modell mit aktuellem → Unterschiede zwischen beiden nicht signifikant, d.h. Interaktionen nicht von null verschieden ($p=0,16$)

Replikation Tabelle 1

- (fitstat,using(m1f) zeigt dementsprechend, daß Modell ohne Interaktionen "besser" ist, d.h. mit weniger Parametern eine vergleichbare Anpassung an die Daten erreicht)
- Mit einem Wald-Test erhalten Sie fast das gleiche Ergebnis wie beim LR-Test
 - nochmals das vollständige Modell schätzen oder einfach mit est restore modell1full die Ergebnisse zurückholen
 - test posteduc postgen postrace postcath
 - In beiden Fällen entsprechen die vier df den Parametern, die im Modell ohne Interaktion nicht geschätzt werden

Replikation Tabelle 1

- Signifikanz von Interaktionstermen immer trickreich, aber hier nur eine Möglichkeit der Codierung
- Interessante Frage: Werden Koeffizienten nach der Entscheidung signifikant, die es vorher nicht sind?
 - z.B. lincom gender + postgender
 - lincom cath+postcath
- Warum ist der Effekt von Katholizismus weder vor noch nach der Entscheidung signifikant?

Replikation Tabelle 1

- Johnson/Martin gehen kaum darauf ein, daß sich für das (zutreffendere) Modell ohne Interaktion ein signifikanter Periodeneffekt ergibt
- Dieser ist positiv, d.h. die Unterstützung für Abtreibungen *nimmt in welchen Gruppen zu?*
- *Referenzgruppe(n): niedriggebildet, männlich, weiß, nicht-katholisch*
- *No!*
- Inhaltliche Bedeutung → geschätzte Wahrscheinlichkeiten
 - Modell ohne Interaktionen noch mal schätzen
 - predict p0 p1 p2
 - tabstat p0 p1 p2,by(postdum)
 - Vorsicht: Das muß nicht am Urteil liegen!

Replikation Tabelle 1

- Ist ordinales Probit-Modell angemessen?
- `outreg using modell1b.txt`
- `brant`
- `ologit disc postdum educ gender race cath attchrch`
- `outreg using modell1b.txt,append`
- `brant`
- Öffnen Sie `u:\stataseminar\modell1b.txt` in Word, markieren Sie alles und wandeln Sie den Text in eine Tabelle um
- Vgl. Sie Koeffizienten und Std.fehler

Replikation Tabelle 1

- predict log0 log1 log2
- tabstat log* ,by(postdum)
- tabstat p0 p1 p2* ,by(postdum)
- Ergeben sich aus einem (angemesseneren) multinomialen Modell substantiell andere Ergebnisse?
 - mlogit disc postdum educ gender race cath attchrch ,base(0)
 - predict m0 m1 m2
 - tabstat m0 m1 m2* ,by(postdum)
- Zwischenfazit: Keine Polarisierung (soweit ok), aber generell substantielle Gegenreaktion auf die Gerichtsentscheidung

Replikation Tabelle 2

- `use z:\daten\jm\Furman.dta,replace`
- `tab favcap`
- `probit favcap postdum educ posteduc
sex postsex race postrace pid postpid
jewish postjewi attend`
- Stimmt alles?
- Zwei Interaktionen sind signifikant
von null verschieden – gehen diese
in die richtige Richtung?

Replikation Tabelle 2

- Sind die Interaktionen gemeinsam von null verschieden?
 - Wald-Test: `test posteduc postsex postrace postpid postjewi`
 - LR-Test: `est store modell2full`
 - `probit favcap postdum educ sex race pid jewish attend` (Dezimalfehler bei PI)
 - `lrtest modell2full`
- Was bedeutet dies inhaltlich?

Replikation Tabelle 2

- Polarisierung?
 - Wie unterscheiden sich weiße, männliche, nicht-jüdische (starke) Anhänger der Demokraten (0) /Republikaner (6) mit High School Abschluß (1), die mehrmals im Jahr zur Kirche gehen (3) vor und nach der Entscheidung?
 - est restore modell2full
 - prvalue,x(postdum 0 educ 1 posteduc 0 sex 0 postsex 0 race 0 postrace 0 pid 0 postpid 0 jewish 0 postjewi 0 attend 3)
 - prvalue,x(postdum 0 educ 1 posteduc 0 sex 0 postsex 0 race 0 postrace 0 pid 6 postpid 0 jewish 0 postjewi 0 attend 3)
 - prvalue,x(postdum 1 educ 1 posteduc 1 sex 0 postsex 0 race 0 postrace 0 pid 0 postpid 0 jewish 0 postjewi 0 attend 3)
 - prvalue,x(postdum 1 educ 1 posteduc 1 sex 0 postsex 0 race 0 postrace 0 pid 6 postpid 6 jewish 0 postjewi 0 attend 3)

Replikation Tabelle 2

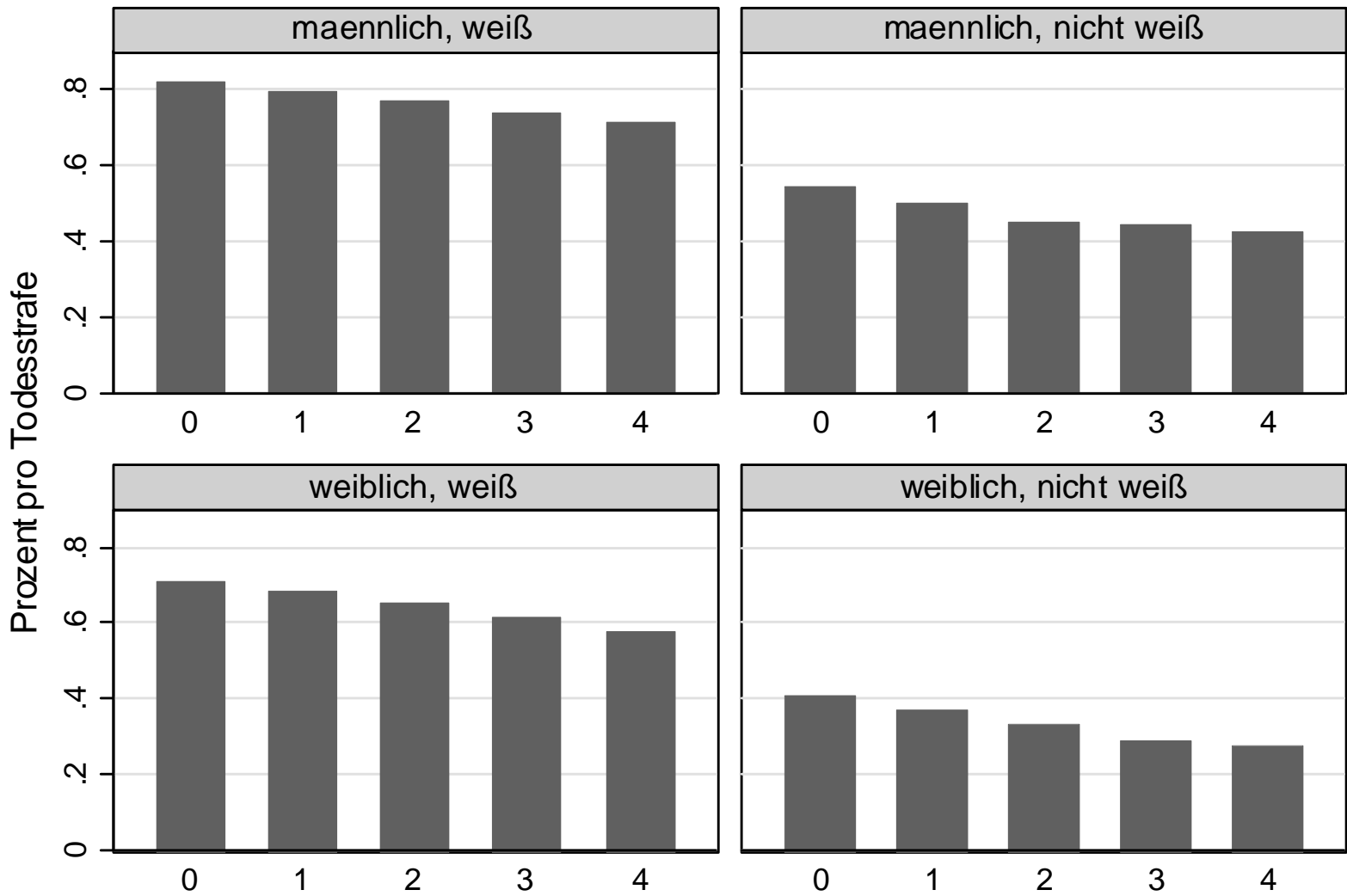
- Wie groß fällt der Unterschied bei Frauen mit ansonsten gleichen Eigenschaften aus?
 - `prvalue,x(postdum 0 educ 1 posteduc 0 sex 1 postsex 0 race 0 postrace 0 pid 0 postpid 0 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 0 educ 1 posteduc 0 sex 1 postsex 0 race 0 postrace 0 pid 6 postpid 0 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 1 educ 1 posteduc 1 sex 1 postsex 1 race 0 postrace 0 pid 0 postpid 0 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 1 educ 1 posteduc 1 sex 1 postsex 1 race 0 postrace 0 pid 6 postpid 6 jewish 0 postjewi 0 attend 3)`

Replikation Tabelle 2

- Wie stark vergrößert sich die Kluft zwischen niedrig- (0) und hochgebildeten (4) nicht-jüdischen weißen Frauen, die keiner Partei zuneigen (3)?
 - `prvalue,x(postdum 0 educ 0 posteduc 0 sex 1 postsex 0 race 0 postrace 0 pid 3 postpid 0 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 0 educ 4 posteduc 0 sex 1 postsex 0 race 0 postrace 0 pid 3 postpid 0 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 1 educ 0 posteduc 0 sex 1 postsex 1 race 0 postrace 0 pid 3 postpid 3 jewish 0 postjewi 0 attend 3)`
 - `prvalue,x(postdum 1 educ 4 posteduc 4 sex 1 postsex 1 race 0 postrace 0 pid 3 postpid 3 jewish 0 postjewi 0 attend 3)`

Hausaufgabe

- Schreiben Sie eine Datei `jm.do`, die
 - den Datensatz `gregg` öffnet
 - Tabelle 3 aus dem Text repliziert
 - die gemeinsame Signifikanz der Interaktionsterme per Wald- und per LR-Test prüft
 - Auf der Grundlage des Modells ohne Interaktionsterme *eine* Balkengrafik erzeugt, die den Einfluß der Bildung auf die Befürwortung der Todesstrafe zeigt
 - und zwar getrennt nach Geschlecht und Ethnizität der Befragten (über alle anderen Variablen wird gemittelt)
 - Am einfachsten geht das mit `predict`
 - Bis zum nächsten Mittwoch an dofiles@politik.uni-mainz.de
 - Das ganze soll so aussehen:



Graphs by Geschlecht and Ethnizitaet