Einführung

Kai Arzheimer | Vorlesung Forschungsmethoden



VL Forschungsmethoden: Für wen?

- Zielgruppe
 - · Master Empirische Demokratieforschung
 - Interessierte aus anderen Studiengängen (hallo Q+!)
- Voraussetzungen
 - · Statistik I (baby steps)
 - Statistik II (Stata, Erweiterungen/Komplikationen Regressionsmodell)
- Erwartungen/Ziele
 - Aktuelle Analysen aus dem Bereich Demokratieforschung (+ ähnliche Gebiete) in groben Zügen verstehen können
 - Idee, wie eigene empirische Arbeit organisiert/strukturiert werden könnte

Warum noch eine Statistikvorlesung im Master? Warum???

- Bisher: Regression und ein paar Erweiterungen (Logit etc.)
- Forschungsorientierter Studiengang erfordert elementare Kenntnis fortgeschrittenerer Verfahren
- Anspruch:
 - Überblick und grundlegendes Verständnis
 - Vorbereitung auf spätere eigenständige Auseinandersetzung wo gewünscht und notwendig
 - Anwendungsbezug
- Wichtig für Ihr weiteres Masterstudium, deshalb Vorlesung + Übung + Seminar
 = 6 SWS
- Aber: keine Klausur, keine Angst vor Formeln, keine Panik

Organisation

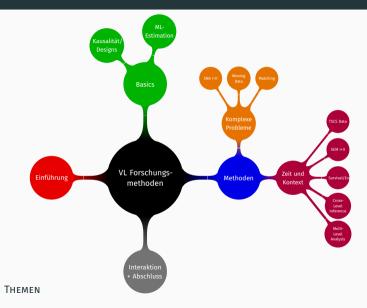
- Master-Vorlesung
 - Selbständigkeit und Eigenmotivation
 - Verbunden mit Kleingruppe (Übung) (Dr. Filip)
 - In beiden Teilen: Bei Unklarheiten sofort unterbrechen
- Empfohlene Literatur
 - Kein Buch zur Vorlesung, aber
 - King/Keohane/Verba: Designing Social Inquiry. Princeton 1994
 - Brady/Collier: Rethinking Social Inquiry. Diverse Tools, Shared Standards. Lanham 2004
 - Gelman/Hill: Data Analysis Using Regression and Multilevel/HierarchicalModels.
 Cambridge 2007
 - · Gelegentlich Hinweise zu einzelnen Themen
- Folien & videos: Moodle

Organisation (Post-Pandemie)

- Die im letzten Jahr produzierten Lehrfilme jeweils vorab (Moodle/Panopto) zur Vorbereitung (nicht an veralteten organisatorischen Hinweisen stören)
- · Intensivere Diskussion über den Stoff in Präsenz

Themen

Übersicht VL Forschungsmethoden



Kausalität und Designs

- Kausalität
 - Tendenz, sich vor Kausalaussagen zu drücken
 - Kausalität ≠ Korrelation
 - Konzepte von Kausalität
 - Wie kann man sich an Kausalaussagen annäheren?
- Designs
 - Warum funktionieren Experimente?
 - Was geht mit welchen Daten? Was funktioniert nicht?
 - Wie sehen gute Designs aus?

Maximum Likelihood Estimation

- "Frequentistische" und "Bayesianische" Statistik
- ML estimation: "Unifiying Political Methodology"?
- Wie kommt der Computer zu Schätzungen (meistens)
- Konzepte statt Details und Technik

Netzwerkanalyse I+II



- Menschen sind keine Atome (oder Inseln)
- Netzwerkanalyse: Beziehungen als Eigenschaften
- Theorie
- Graphische Darstellungen, einfache deskriptive Verfahren und Maßzahlen
- Anwendungen und komplexere Methoden

Missing Data



- Was passiert, wenn Werte fehlen?
- Einfache Imputationstechniken und ihre Grenzen
- Einführung multiple Imputation



Matching



- Kann man mit ex-post facto Daten Kausalaussagen testen ...
- ... indem man systematisch Fälle ignoriert?
- Voodoo?



Faktorenanalyse und Strukturgleichungsmodelle I+II

- Interessante Variablen sind nicht (direkt) beobachtbar
- Welche (latenten) Variablen sind für beobachtbare Kovarianzen verantwortlich
- Oder: Vom Voodoo zur Seance

Time-Series Cross-Sectional Data



- · Daten, die in Raum und Zeit strukturiert sind
- · Gewinne ich mehr Informationen, indem ich öfter hinschaue?
- Wenn ja, wie und in welchem Umfang?

THEMEN

Ereignisdaten, Selektion, Zensur etc.



- Zeitdauer, bis ein Objekt in einen anderen Zustand wechselt
 - Lebensdauer von Koalitionsregierungen
 - Wiederwahl von Abgeordneten
 - Frieden → Krieg in einer Dyade
- Zeitdauer ≥ o
- Nicht alle Sequenzen abgeschlossen, da Beobachtung irgendwann endet

Cross-Level Inference

- Idealerweise: Theorie und empirische Analyse auf selber Ebene (meistens: Mikro)
- Aber:
 - Mikro-Daten nicht verfügbar (z.B. historische Wahlforschung)
 - Datenschutz
- Schluß von Aggregatdaten (Randverteilung) auf individuelles Verhalten
 - Fehlschlußproblematik
 - Identifikation von Bandbreiten, innerhalb derer sich individuelles Verhalten bewegen muß

THEMEN

Mehrebenenanalyse



- Idealfall der sozialwissenschaftlichen Erklärung: "Badewannenmodell" (Coleman, Humell/Opp, Esser)
- Empirische Annäherung: Makrovariablen in Mikromodell einbeziehen
- Probleme: Standardfehler, Präzision der Schätzungen
- Moderne Mehrebenenanalyse

Abschlussdiskussion + Interaktionseffekte

- Interaktionen: Kein wahnsinnig fortgeschrittenes Thema
- · Passt nicht ganz in das Schema
- Macht in Master-Seminaren aber immer wieder Probleme

THEMEN 16



Datenstrukturen

Was sind Daten?

Was sind Daten?

- Systematische, quantifizierte Beobachtungen
- Ergebnis von
 - "Datengenerierendem Prozeß" (DGP)
 - Kausale Mechanismen und ihre Konsequenzen
 - Z.B. Wirtschaftsaufschwung \to Sozialisation \to individuelle Wertorientierungen \to gesellschaftlicher Wandel
 - Schwachsinniger Name
 - Unseren Beobachtungen
- Statistische Analyse
 - · Theoretische Annahmen über DGP
 - · Mathematische Modellierung des DGP
 - · Ableitung prüfbarer Hypothesen
 - · Wie gut passen Daten und Hypothesen zueinander?
 - · Probabilistische Tests

Welche Probleme gibt es (Auswahl)?

- Wirklichkeit zu komplex
- Theorien zu unspezifisch
- Messungen falsch
- Tests setzen normalerweise korrekte Modellspezifikation voraus
- Komplette Validierungsproblematik ...
- Komplexe Daten

Was bedeutet "komplexe Daten"?

- Theorien argumentieren auf unterschiedlichen Ebenen
- · Variablen haben nicht nur verschiedene Skalenniveaus
- Sondern unterscheiden sich in der Logik ihrer Konstruktion

Wie lassen sich Variablen systematisch organisieren?

Ebene	Mikro		Meso		Makro		
Тур	global relational kontextuell	$\begin{array}{c} \rightarrow \\ \rightarrow \\ \leftarrow \end{array}$	analytisch strukturell global relational kontextuell	$\begin{array}{c} \rightarrow \\ \rightarrow \\ \leftarrow \end{array}$	analytisch strukturell global relational kontextuell	$\begin{array}{c} \rightarrow \\ \rightarrow \\ \leftarrow \end{array}$	

 \rightarrow : Aggregation

 $\leftarrow\!\!: Disaggregation$

Quelle: Darstellung nach Hox 2002 und Lazarsfeld/Menzel 1961

Was bedeutet das? i

- Globale Variablen
 - · Eine Ebene, eine Einheit
 - Bsp: individuelle Arbeitslosigkeit, Alter, Geschlecht, Einstellungen, Wahlsystem, Dezentralisierung ...
- · Analytische Variablen
 - · Zwei Ebenen, mehrere Einheiten
 - Basieren auf Aggregation (Mittelwerte, Anteile ...) untergeordneter Einheiten
 - Z.B. Arbeitslosenquote

Was bedeutet das? ii

- Relationale Variablen
 - · Eine Ebene, mehrere Einheiten
 - · Beziehungen zwischen Objekten auf einer Ebene
 - Z.B. Kommunikation zwischen Wählern ("ties": ja/nein; Häufigkeit)
- Strukturelle Variablen
 - · Zwei Ebenen, mehrere Einheiten
 - Aggregation relationaler Variablen, z.B. Dichte der ties vgl. mit anderen Einheiten

Datenstrukturen 22

Was bedeutet das III: kontextuell

- Bezieht sich auf (vermutete) Wirkung übergeordneter auf untergeordnete Ebene
- Kontextuelle Variablen
 - · Zwei Ebenen, mehrere Einheiten
 - · Eigenschaften der übergeordneten Ebene
 - z.B. alle Wähler demselben Wahlsystem ausgesetzt (global, analytisch, strukturell)

Fazit

Zusammenfassung

- Stand der Forschung erfordert, daß man über das Programm von SM I+II hinausgeht
- · Weites Feld
- Wichtiger als Detailwissen
 - · Verständnis von Konzepten
 - · Vorstellung davon, was warum (nicht) geht

FAZIT 24

Literaturempfehlungen für die nächsten Wochen

- Kausalität und Designs: King, Keohane & Verba 1994
- Maximum-Likelihood: King 1998 (Unifying Political Methodology), Eliason 1993
 (Maximum Likelihood Estimation: Logic and Practice)
- Netzwerkanalyse: Ebook von Hanneman/Riddle: http://www.faculty.ucr.edu/~hanneman/nettext/

FAZIT 25