

# Anlage empirischer Untersuchungen und Besonderheiten von Paneldaten

## Gliederung

1. Rückblick
2. Analyse von Sekundärdaten
3. Analysemöglichkeiten des SOEP
4. Datenstrukturen

## Rückblick

### Wohlfahrtsstaat und soziale Ungleichheit (24.4.)

- Esping-Andersen 1990 (Kap. 1): Dekommodifizierung, Wohlfahrtsregimes, Verhältnis von Markt, Staat und Familie, Deutschland als konservativ-korporatistischer Wohlfahrtsstaat
- Esping-Andersen 1990 (Kap. 3): Strukturierung von Ungleichheit durch unterschiedliche Ausgestaltung von Wohlfahrtsstaaten (Stratifizierung)
- Lewis 1992: Dekommodifizierung als primär auf Männer bezogenes Konzept, Bedeutung unterschiedlicher Ernährermodelle
- Ringen 1987: Einfluss des Wohlfahrtsstaats auf (Einkommens-)ungleichheit, Möglichkeiten und Probleme des standardmäßigen Vorgehens

## Aktuelle Perspektiven (8.5.)

- Gilbert 2002: gradueller oder grundsätzlicher Wandel, Divergenz oder Konvergenz, vom Wohlfahrtsstaat zum ‚enabling state‘
- Leitner et al. 2004: Entwicklung zum ‚adult-worker‘ Modell, (De-) Familisierung und (Re-) Kommodifizierung
- Vogel 2006: zunehmende Prekarität aufgrund des Wandels des Wohlfahrtsstaates und des Arbeitsmarktes
- Heinelt 2003: Wandel der Arbeitsmarktpolitik, von der aktiven zur aktivierenden Arbeitsmarktpolitik

## Empirische Befunde I (15.5.)

	Thema	Ungleichheit	Wohlfahrtsstaat	Daten	abh. Variable	Untersuchungsdesign	Vorgehen
Andref/Seeck 2007	Armut von Erwerbstätigen	Einkommen/ Armut	Veränderungen seit 1990	SOEP	relative Einkommensarmut	Trend	deskriptiv
Andref/Strengmann-Kuhn 1997	Arbeitsangebot im unteren Einkommensbereich	indirekt: Niedrig-einkommen	negative Arbeitsanreize durch Transfers	SOEP	Arbeitsangebot (ja/nein)	Querschnitt	multivariat
Ludwig-Mayerhofer 1992	Zusammenhang zwischen (dauerhafter) Arbeitslosigkeit und (dauerhafter) Armut	Einkommen/ Armut, Arbeitslosigkeit	indirekt: Einfluss der Verfügbarkeit von Arbeitslosenunterstützung	SOEP	relative Einkommensarmut und Sozialhilfebezug	Querschnitt unter Verwendung von Längsschnittinformationen	deskriptiv und multivariat
Becker 2002	Frauenerwerbstätigkeit und Armut von Familien	Einkommen	indirekt: Familienlastenausgleich, Kinderbetreuung	EVS	rel. Wohlstandsposition / Armut	Trend	deskriptiv

## Empirische Befunde II (22.5.)

	Thema	Ungleichheit	Wohlfahrtsstaat	Daten	abh. Variable	Untersuchungsdesign	Vorgehen
Kreyenfeld/Geisler 2006	Unterschiede in der Müttererwerbstätigkeit zwischen Ost- und Westdeutschland	Position am Arbeitsmarkt, geschlechtsspezifische Ungleichheit	familienpolitische Unterschiede zwischen Ost und West	Mikrozensus	Erwerbstätigkeit	Trend	deskriptiv und multivariat
Böhnke 2006	gesellschaftliche Desintegration als Folge sozialer Ausgrenzung	soziale Ausgrenzung	indirekt: Versagen bzw. Gefährdung des Wohlfahrtsstaats	Wohlfahrts-survey	kritische Einstellungen zum Gesellschaftssystem, resignative Lebenseinstellung, Wahlenthaltung (Parteineigung)	Querschnitt	deskriptiv und multivariat
Müller/Pollak 2004	Ungleichheit im Zugang zu tertiärer Bildung	Bildung	indirekt: Bildungsinstitutionen	diverse, u.a. SOEP	Schulabschluss (Abitur), Ausbildungsabschluss	Trend	deskriptiv und multivariat
Granato/Kalter 2001	Persistenz ethnischer Ungleichheit auf dem Arbeitsmarkt	Position am Arbeitsmarkt, ethnische Ungleichheit	indirekt: Arbeitsmarkt-institutionen, Bildungs-institutionen	Mikrozensus	berufliche Stellung	Querschnitt	deskriptiv und multivariat

## Untersuchungsdesigns

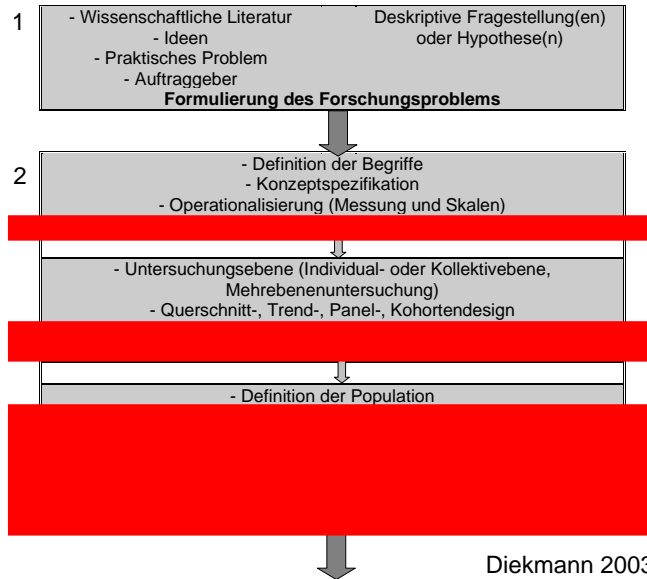
- Querschnittsdesign
  - Einmalige Befragung aller Untersuchungseinheiten (z.B. Person, Haushalt) zum gleichen Zeitpunkt
- Längsschnittsdesign
  - Mehrmalige Befragung aller Untersuchungseinheiten an verschiedenen Zeitpunkten (Paneldesign)
  - Befragung aller Untersuchungseinheiten zum gleichen Zeitpunkt, aber rückblickend über mehrere Zeitpunkte (Retrospektivdesign)
  - Befragung von unterschiedlichen Untersuchungseinheiten an mehreren Zeitpunkten (Trenddesign → wiederholte Querschnitte)

## Analyse von Sekundärdaten

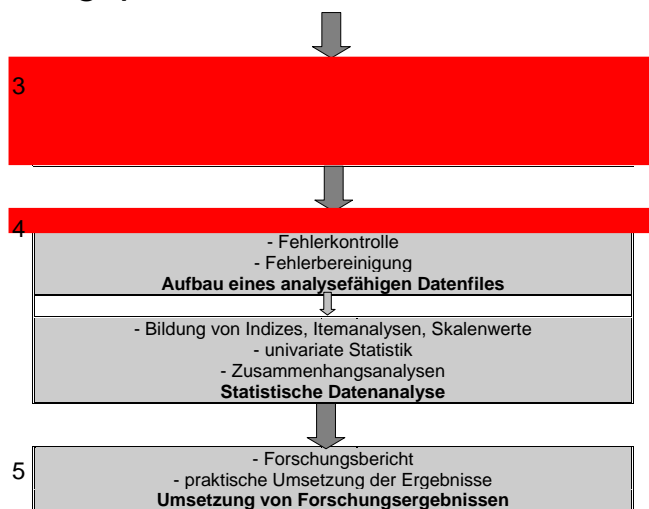
## Analyse von Sekundärdaten

„Datenerhebung ist zeitaufwendig und teuer [...] Da sich ein Datensatz häufig auf mehrere Fragestellungen bezieht, die zu bearbeiten einen einzelnen Forscher überfordern würde, ist die Auswertung vorhandener Datensätze [...] wünschenswert.“ (Bortz/Döring 2002: 374)

# Forschungsprozeß



# Forschungsprozeß



## Vorteile Sekundäranalyse

- Aufwand/fehlende Voraussetzungen (Fallzahlen, Kombination von Datenquellen: Internationale Forschung, Methodenmix, Trendanalysen)
- inhaltliche Erweiterung (nicht bearbeitete Hypothesen, nachträglich entdeckte Forschungsbereiche)
- Möglichkeit zur Exploration für eigene Erhebungen
- Kontrolle der Primärforschung durch Replikation
- positive Effekte für empirische Forschung allgemein („Demokratisierung“ der Forschung, Reduzierung der durchgeführten Erhebungen)

vgl. Diekmann 2003: 172f, Kiecolt/Nathan 1985: 11f, Dale et al. 1988: 44f

## Nachteile Sekundäranalyse

- eingeschränkte Verfügbarkeit
- Datenbeschaffungsaufwand
- evtl. unzureichende Dokumentation
- Fehler häufig nur schwer nachzuvollziehen, Vergrößerung der Fehler bei Kombination mit anderen Datensätzen
- kleine Fallzahlen bei Betrachtung rarer Populationen
- Datenqualität, Operationalisierung
- Limitierung der eigenen wissenschaftlichen Interessen

vgl. Kiecolt/Nathan 1985: 11f

## Einige Kriterien für Auswahl einer Datenquelle für Sekundäranalysen

- Inhaltliche Fragen:
  - Gegenstandsbereich
  - Verfügbarkeit benötigter Indikatoren
  - benötigte Fallzahl
  - Untersuchungsdesign: Längsschnitt/Querschnitt
  
- Pragmatische Fragen:
  - Zugangsbeschränkungen
  - Beschaffungsaufwand: Zeit, Kosten
  - Verfügbare Dokumentation

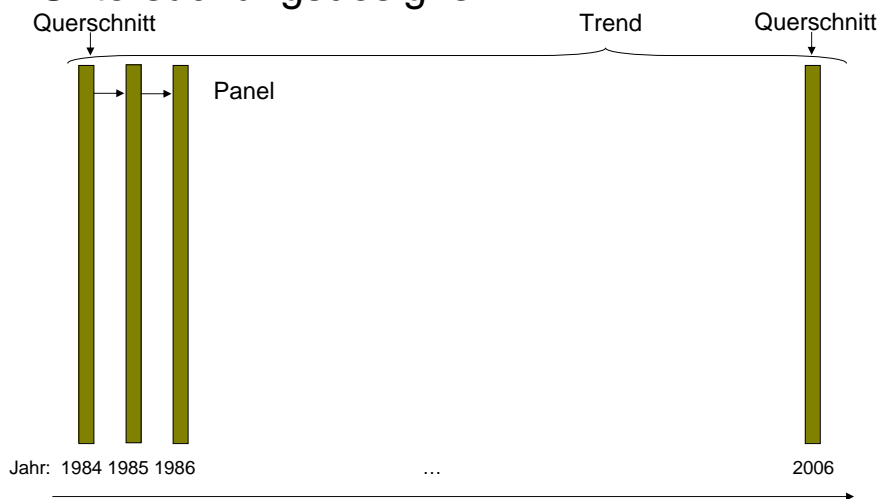
## Analysemöglichkeiten des SOEP



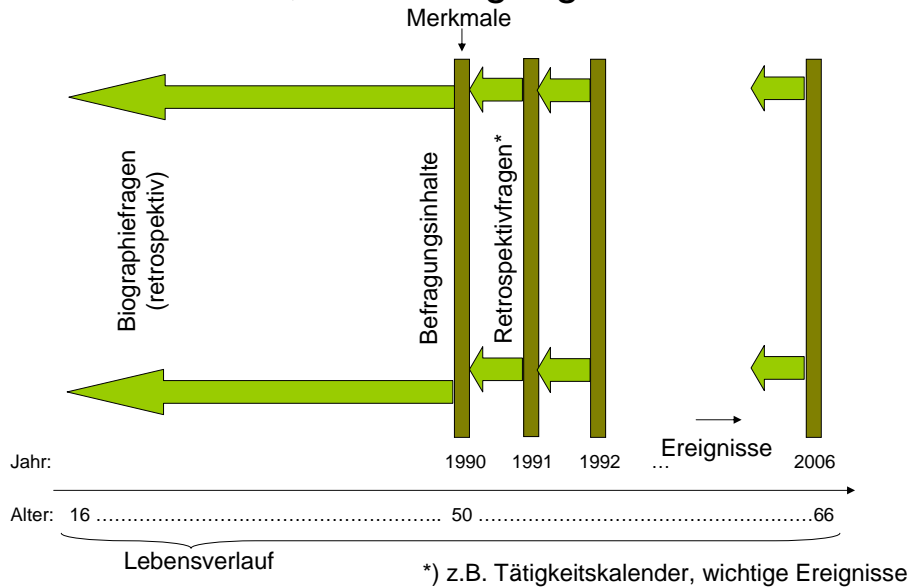
## Sozio-oekonomisches Panel

- Längsschnittbefragung privater Haushalte in Deutschland, jährliche Befragung seit 1984
- Bereiche: vor allem Sozial- und Wirtschaftsforschung
- Ziel: Repräsentative Erfassung, Analyse und Interpretation des gesellschaftlichen und wirtschaftlichen Wandels
- angesiedelt am Deutschen Institut für Wirtschaftsforschung (DIW) in Berlin → [www.diw.de](http://www.diw.de)

## Untersuchungsdesigns



## Erwachsener, Erstbefragung 1990



## Begriffe

- Ereignis: Übergang von einem Status in einen anderen Status (z.B. Beendigung Arbeitslosigkeit: arbeitslos → beschäftigt, Scheidung: verheiratet → geschieden)
- Dauer eines Status: Dauer meint den Verbleib in einem Status, also die Zeit vom Eintritt bis zum Austritt aus einem Status (Probleme: Links-/ Rechtszensierung)
- Lebensverlauf: Lebensverlauf meint die interdependenten Wege und Übergänge, die eine Person im Laufe ihres Lebens erfährt.

## Lebensverläufe und Paneldaten

- Paneldaten eignen sich zur Beschreibung von Lebensverläufen nur (mehr oder weniger) eingeschränkt

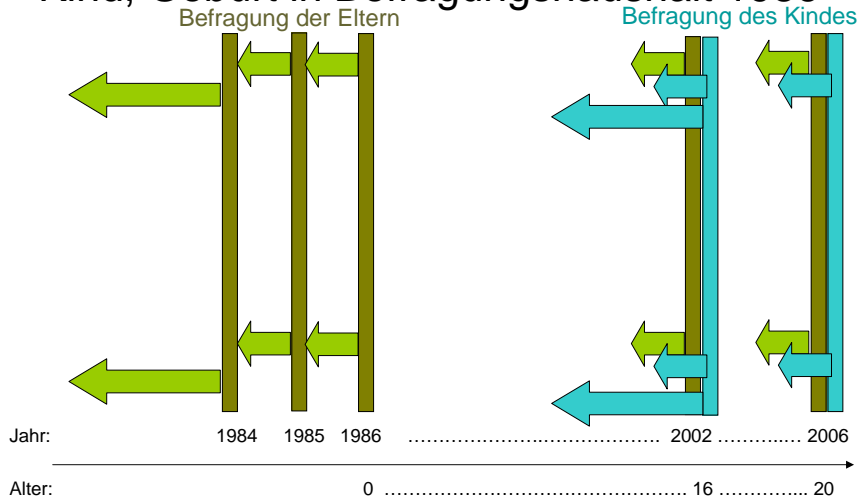
Probleme:

- Angaben über den Zeitraum vor Eintritt ins Panel
- Veränderungen zwischen zwei Befragungszeitpunkten „Behebung“:
  - retrospektive Befragungen bei Eintritt ins Panel bzw. für den Zeitraum zwischen zwei Befragungswellen
  - je länger ein Panel läuft, desto länger wird auch der Ausschnitt des Lebensverlaufs der abgebildet wird

## Geeignete multivariate Analyseverfahren

- Querschnitt: herkömmliche Regressionsmodelle (je nach abhängiger Variable)
- gepoolte Paneldaten („Trend“): Problem mehrerer Beobachtungen einer Person → Panelregression (z.B. fixed effects, random effects model)
- Ereignisse=Übergänge von einem Status in einen anderen: (diskrete) Ereignisdatenanalyse

## Kind, Geburt in Befragungshaushalt 1986



## Möglichkeiten der Analyse von Paneldaten

1. differenzierte Aussagen über sozialen Wandel möglich: kumulierte Querschnitte erlauben Aussagen über Trends, Paneldaten erlauben zusätzlich Aussagen über individuelle Veränderungen
2. bessere Möglichkeiten der Kausalanalyse: zeitliche Abfolge kann berücksichtigt werden
3. Keine Erinnerungsprobleme wie bei Retrospektivbefragungen
4. Vergrößerung der Stichprobe
5. Kontrolle unbeobachteter Heterogenität

## Beispiel: Scheidung und Einkommen

Einkommen	Fam.stand
niedrig	geschieden
hoch	verheiratet
Querschnitt	

## Beispiel: Scheidung und Einkommen

Fam.stand	Einkommen	Ereignis	Einkommen	Fam.stand
verheiratet	hoch	Scheidung	niedrig	geschieden
t-1		t	t+1	
Panel				

## Beispiel: Scheidung und Einkommen

Fam.stand	Einkommen	Ereignis	Einkommen	Fam.stand
verheiratet	hoch	Scheidung	niedrig	geschieden
verheiratet	hoch		hoch	verheiratet
t-1		t	t+1	
Panel				

Unbeobachtete Faktoren ?

## Probleme der Analyse von Paneldaten

1. keine statistische Unabhängigkeit zwischen zwei Beobachtungen einer Untersuchungseinheit
2. Probleme mit mangelnder Konstanz des Erhebungsinstrument (Änderung des Fragebogens, Bedeutungswandel von Fragen, Lerneffekte)
3. zusätzliche Schwierigkeiten zur Erlangung von Repräsentativität im Querschnitt (Panelmortalität, Veränderung der Bevölkerungsstruktur)

# Datenstrukturen

## Beantwortung von Fragen

### 1. Wie zufrieden sind Sie gegenwärtig mit den folgenden Bereichen Ihres Lebens?

Bitte kreuzen Sie für jeden Bereich auf der Skala einen Wert an:  
Wenn Sie ganz und gar **unzufrieden** sind, den Wert "0",  
wenn Sie ganz und gar **zufrieden** sind, den Wert "10".  
Wenn Sie **teils zufrieden / teils unzufrieden** sind, einen Wert dazwischen.

Wie zufrieden sind Sie ...

ganz und gar unzufrieden

ganz und gar zufrieden

– mit Ihrer Gesundheit?

0 1 2 3 4 5 6 7 8 9 10



SOEP 2006

## Datenmatrix

	1	2	...	m
1	8	$X_{12}$	...	$X_{1m}$
2	$X_{21}$	$X_{22}$	...	$X_{2m}$
3	$X_{31}$	$X_{32}$	...	$X_{3m}$
...	...	...	...	...
n	$X_{n1}$		...	$X_{nm}$

Variablen

Untersuchungs-  
einheiten bzw.  
Beobachtungen

## Querschnittsdaten

persnr	jahr	geschl	erw	eink
1	2006	1	1	1000
2	2006	2	0	1200
3	2006	1	1	7400
4	2006	1	1	3000
5	2006	2	.	.
6	.	1		
7		.		
8				
9				



## Paneldaten

persnr	jahr	geschl	erw	eink
1	2003	1	1	1700
1	2004	1	1	1750
1	2005	1	0	1050
1	2006	1	1	1500
2	2003	2	0	600
2	2004	2	0	620
2	2005	2	1	1000
2	2006	2	1	1000
3	2003	.	.	.

## Paneldaten

persnr	jahr	geschl	erw	eink
1	2003	1	1	1700
1	2004	1	1	1750
1	2005	1	0	1050
1	2006	1	1	1500
2	2003	2	0	600
2	2004	2	0	620
2	2005	2	1	1000
2	2006	2	1	1000
3	2003	.	.	.

## Trenddaten (kumulierte Querschnitte)

persnr	jahr	geschl	erw	eink
1	2004	1	1	2000
2	2004	2	1	2100
3	2004	1	1	1300
4	2005	1	1	2500
5	2005	2	1	2700
6	2005	2	0	620
7	2006	2	1	4000
8	2006	2	0	500
9	2006	1	0	550

## Trenddaten (kumulierte Querschnitte)

persnr	jahr	geschl	erw	eink
1	2004	1	1	2000
2	2004	2	1	2100
3	2004	1	1	1300
4	2005	1	1	2500
5	2005	2	1	2700
6	2005	2	0	620
7	2006	2	1	4000
8	2006	2	0	500
9	2006	1	0	550

## Literatur

- Bortz, Jürgen/ Döring, Nicola (2002): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler, Berlin/Heidelberg/New York: Springer.
- Dale, Angela/ Arber, Sara/ Procter, Michael (1988): Doing Secondary Analysis, Unwin Hyman: London.
- Diekmann, Andreas (2003): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.
- Kiecolt, K. Jill/ Nathan, Laura E. (1985): Secondary analysis of survey data. Beverly Hills u.a.: Sage.
- Ruspini, Elsibetta (2002): Introduction to Longitudinal Research. London / New York: Routledge.