

## Analysepotenziale sozialwissenschaftlicher Forschungsdaten

Workshop 23. bis 27. September 2013  
an der Universität Bielefeld



Datenservicezentrum

Betriebs- und OrganisationsDATEN



Schon länger besteht in den Sozial- und Wirtschaftswissenschaften Einigkeit darüber, dass viele Forschungsfragen nur im Längsschnitt adäquat untersucht werden können. Längsschnittliche Forschungsdaten werden deshalb immer häufiger genutzt, sie setzen aufgrund ihrer komplexen Datenstruktur bestimmte Kenntnisse und Erfahrungen im Datenhandling voraus, die in der universitären Lehre nur begrenzt vermittelt werden (können). In den letzten Jahren werden jedoch mehr und mehr auch andere, komplexe Datenbestände in den Sozial- und Wirtschaftswissenschaften aufgebaut, die auch die sozialen Kontexte, in denen Individuen und Haushalte eingebunden sind, abbilden. Dazu zählen Daten über Partnerschaften und Beziehungsstrukturen, Betriebe und Unternehmen. Letztere sind nicht nur für die Organisationsforschung im engeren Sinne, sondern zunehmend auch für andere Forschungsfelder wichtig z.B. der Arbeitsmarktforschung, der Konjunkturforschung oder der familien- oder ungleichheitssoziologischen Forschung. Diese Datenbestände finden in den bestehenden Ausbildungsprogrammen nur eingeschränkte Beachtung.

Beim diesjährigen *data lab*-Workshop vom 23. bis 27. September 2013 an der Universität Bielefeld werden die Nutzungsmöglichkeiten des Sozio-ökonomischen Panels (SOEP), des Beziehungs- und Familienpanels (pairfam) sowie die Betriebs- und Linked-Employer-Employee-Datensätze (Mannheimer Innovationspanel, Datensätze des Ifo Instituts und SOEP-Linked-Employer-Employee Datensatz, SOEP-LEE) vorgestellt.

Der Workshop bietet eine praxisorientierte Einführung in die Nutzung dieser Datensätze. Neben der Vermittlung von Basisinforma-

tionen zum Design, Aufbau und den Analysepotentialen werden erste Erfahrungen in der Nutzung der Datensätze durch intensive Übungen am Computer vermittelt. Die Vorstellung der jeweiligen Datensätze erfolgt durch Mitarbeiter der Forschungsdatenzentren der jeweiligen Datenproduzenten in zwei Modulen mit drei parallel angebotenen Veranstaltungslinien (streams):

### Modul 1: Nutzung komplexer Datensätze

#### Stream 1: Sozio-oekonomisches Panel (SOEP)

Das SOEP ist eine seit 1984 laufende jährliche Wiederholungsbefragung in Deutschland, mit den Themenschwerpunkten Haushaltszusammensetzung, Erwerbs- und Familienbiographie, Einkommensverläufe, Gesundheit und Lebenszufriedenheit, Persönlichkeit und Kompetenzen. Der Workshop bietet einen fundierten Einblick in das Arbeiten mit dem SOEP. So wird ein umfassender Überblick über Aufbau und Analysepotentiale des SOEP gegeben. Außerdem werden verschiedene methodische Aspekte (wie Stichprobenziehung, Gewichtung und Imputation) vorgestellt. Neben den Kernthemen des SOEP, die in Vorlesungen und praktischen Übungen behandelt werden, bieten wir in diesem Jahr zusätzlich eine Überblicksveranstaltung zum SOEP in der Migrations- und Integrationsforschung.

#### Stream 2: pairfam

Das von der Deutschen Forschungsgemeinschaft als Langfristvorhaben geförderte Projekt ist die bislang umfangreichste Studie zur Erforschung partnerschaftlicher und familialer Lebensformen in Deutschland. Mehr als 12400 per Zufall ausgewählte Personen im Alter von 15 bis 17, 25 bis 27 und 35 bis 37 Jahren nahmen an der ersten Befragungswelle im Jahr 2008 teil. Über einen Zeitraum von 14 Jahren werden diese Personen sowie deren Partner, Eltern und Kinder im jährlichen Abstand wiederholt befragt. Themenschwerpunkte sind die Ausgestaltung von Partnerschaften, Fertilität und Elternschaftsentscheidungen, kindliche Entwicklung und Erziehung sowie intergenerationale Beziehungen. Im Workshop werden Aufbau, Datenstruktur und Analysepotentiale vorgestellt.

#### Stream 3: Betriebs- und Linked-Employer-Employee-Daten

*Mannheimer Innovationspanel (MIP)* des Zentrums für Europäische Wirtschaftsforschung (ZEW): Das MIP enthält Informationen zur Einführung neuer Produkte, Dienstleistungen und Verfahren in Unternehmen, den Aufwendungen für Innovationen und den Erfolg, den Unternehmen mit neuen Produkten, neuen Dienstleistungen und verbesserten Verfahren erzielen. Das MIP wird seit 1993 jährlich als Panelbefragung erhoben und deckt die Bereiche Bergbau, verarbeitendes Gewerbe, Energie, Baugewerbe, unternehmensnahe Dienstleistungen und distributive Dienstleistungen ab. Die Erhebung ist für Deutschland repräsentativ und ermöglicht

Hochrechnungen für die deutsche Wirtschaft insgesamt sowie für einzelne Branchengruppen. Im Workshop werden die Daten und ihr Analysepotenzial vorgestellt.

*Unternehmensdaten des ifo Instituts:* Seit 1949 führt das ifo Institut Unternehmensbefragungen in Deutschland durch. Im Rahmen des ifo Konjunkturtests werden monatlich Betriebe u.a. zur Entwicklung der allgemeinen Geschäftslage, den Verkaufspreisen, Warenbeständen und der Beschäftigtenzahl befragt. Die Umfrage ist zugleich die Basis für den *ifo Geschäftsklimaindex*. Daneben existieren noch weitere Befragungen, z.B. der zweimal jährlich durchgeführte *ifo Investitionstest*, welcher die realisierten und die geplanten Unternehmensinvestitionen beinhaltet, sowie der einmal jährlich erhobene *ifo Innovationstest*, der die Evaluation der unternehmerischen Innovationstätigkeit zum Ziel hat. Die Mikrodaten des ifo Konjunkturtests werden innerhalb des Business Expectations Panels (BEP) mit Bilanzdaten der Unternehmensdatenbanken Amadeus und Hoppenstedt verknüpft, ebenso wie die Mikrodaten des ifo Investitionstests im Business Investment Panel (BIP) und die Mikrodaten des ifo Innovationstests im Business Innovation Panel (BINP). Sie stehen auch für Analysen zur Verfügung. Im Rahmen der Übungen wird auf die Datensätze und ihre Besonderheiten, insbesondere die Datenstruktur, eingegangen. Anhand von ausgewählten Fragestellungen werden Analysepotentiale vorgestellt.

*SOEP-LEE:* Dieser Datensatz stellt Informationen zum Arbeitskontext und zu den Arbeitsbedingungen von insgesamt 1700 abhängig beschäftigten Befragten des SOEP bereit, die im Rahmen einer gesonderten standardisierten Betriebsstättenbefragung im Jahr 2012 erhoben wurden. Die Verknüpfung von Personen-, Haushalts- und Betriebsdaten eröffnet insbesondere für die Ungleichheitsforschung neue Analysepotentiale, die in diesem Workshop – neben den technischen und methodischen Besonderheiten dieses Datensatzes – vorgestellt werden.

### Modul 2: Grundlegende und fortgeschrittene Analyseverfahren

Durch Lehrvorträge und darauf bezogene praktische Übungen wird ein vertiefender Einblick in (1) grundlegende und (2) fortgeschrittene Verfahren der Längsschnittanalyse sowie ihre Nutzungspotentiale gegeben (Einführung in die Panelregression, Panelregression mit dichotomen abhängigen Variablen, Analyse von Dyaden und Strukturgleichungsmodelle). Neben gemeinsamen Veranstaltungen für die Workshopteilnehmer gibt es jeweils parallel ablaufende Übungen, in denen die entsprechenden Verfahren auf Basis der Datensätze SOEP und pairfam angewendet werden.

## Ablauf Modul1

Montag, 23. September 2013

10.00-12.00 Propädeutikum in Stata für Teilnehmer/innen ohne Stata-Vorkenntnisse

13.00-14.00 Begrüßung, Einführung und Kurzvorstellung der Streams

Stream 1: SOEP	Stream 2: pairfam	Stream 3: Firmen-/ LEE-Daten
-------------------	----------------------	---------------------------------

14.00-15.30 <u>Einführung SOEP</u> Datenstruktur, Inhalte, Dokumentation	<u>Einführung pairfam</u> Datenstruktur, Dokumentation	<u>Einführung MIP</u> Datenstruktur, Inhalte, Design
---	---	---

16.00-17.30 <u>Übung I</u> Dokumentation und Datenorganisation	<u>Übung I</u> Dokumentation und Datenorganisation (Teil1)	<u>Übung I</u> MIP
---	---	-----------------------

18.00-19.30 <u>Übung II</u> Datenmanagement Querschnitt	<u>Übung II</u> Datenmanagement Querschnitt	<u>Übung II</u> MIP
--	--	------------------------

Dienstag, 24. September 2013

SOEP	pairfam	Firmen- / LEE-Daten
------	---------	---------------------

9.30-11.00 Das SOEP in der Migrations- und Integrationsforschung	Multi-Actor-Design: Datenstrukturen und Anwendungsbeispiele	Unternehmensdaten des ifo-Instituts
--	---	-------------------------------------

11.30-13.00 <u>Übung III</u> Datenmanagement: Längsschnitt	<u>Übung III</u> Datenmanagement Längsschnitt	<u>Übung I</u> ifo Mikrodaten
---	--	----------------------------------

14.00-16.00 <u>Übung IV</u> Intergenerationale Verknüpfungen	<u>Übung IV</u> Multi-Actor-Verknüpfungen	<u>Übung II</u> EBDC Panel
---	--	-------------------------------

16.30-18.30 Individuelle Beratung (nach Bedarf)	Individuelle Beratung (nach Bedarf)	<u>Vortrag</u> Datensätze für die Betriebs- und Organisationsforschung: Das Angebot des DSZ-BO
---	-------------------------------------	---

Mittwoch, 25. September 2013

SOEP	pairfam	Firmen- / LEE-Daten
------	---------	---------------------

9.00-10.30 Stichproben, Gewichtung und Hochrechnung im SOEP	Analysebeispiele und Analysepotenziale	Verknüpfte Personen- und Betriebsdaten SOEP-LEE
---	--	---

11.00-12.15 Geodaten in den Sozialwissenschaften - die Nutzung im Rahmen der Konstruktion von Sozialindizes und für methodische Analysen  
*Prof. Jörg-Peter Schräpler*

13.00-15.00 <u>Übung V</u> Gewichtung	<u>Übung V</u> Deskriptive Panelanalysen	<u>Übung I</u> SOEP-LEE
--	---	----------------------------

15.15-17.00 <u>Übung VI</u> Deskriptive Panelanalysen	Netzwerkbildung	<u>Übung II</u> SOEP-LEE
--	-----------------	-----------------------------

## Ablauf Modul2

Donnerstag, 26. September 2013

9.30-11.00 Grundlegende Verfahren (GV) I:  
Verfahren der Längsschnittanalyse Ia: Panelregression  
*Dr. Marco Giesselmann*

9.00-11.00 Fortgeschrittene Verfahren (FV) I:  
Einführung in lineare Strukturgleichungsmodelle  
*PD Dr. Oliver Arránz Becker*

Stream 1: SOEP	Stream 2: pairfam
----------------	-------------------

11:30-13:30 <u>GV I:</u> Übung Panelregression	<u>GV I:</u> Übung Panelregression
--	------------------------------------

FV I: Übung Strukturgleichungsmodelle (SOEP & pairfam)

14:30-18:00 <u>GV I:</u> Übung Panelregression	<u>GV I:</u> Übung Panelregression
--	------------------------------------

FV I: Übung Strukturgleichungsmodelle (SOEP & pairfam)

Freitag, 27. September 2013

9.30-11.00 Grundlegende Verfahren (GV) II:  
Verfahren der Längsschnittanalyse Ib: Panelregression mit dichotomen abhängigen Variablen (Logit)  
*Dr. Marco Giesselmann*

9.00-11.00 Fortgeschrittene Analyseverfahren (FV) II:  
Fortgeschrittene Analyseverfahren: Analyse von Dyaden  
*PD Dr. Oliver Arránz Becker*

Stream 1: SOEP	Stream 2: pairfam
----------------	-------------------

11:30-13:30 <u>GV II:</u> Übung Panelregression mit dichotomen abhängigen Variablen (Logit)	<u>GV II:</u> Übung Panelregression mit dichotomen abhängigen Variablen (Logit)
---	---

FV II: Übungen Dyadische Verfahren (SOEP & pairfam)

14:00-17:30 <u>GV II:</u> Übung Panelregression mit dichotomen abhängigen Variablen (Logit)	<u>GV II:</u> Übung Panelregression mit dichotomen abhängigen Variablen (Logit)
---	---

FV II: Übungen Dyadische Verfahren (SOEP & pairfam)

Hinweis: Änderungen im Programmablauf sind vorbehalten

### Teilnahme

Die Teilnahmegebühr für den gesamten Workshop beträgt 210€ (Doktorand/innen und Student/innen 140€). Für Modul 1 einzeln 90€ (ermäßigt 60€), Modul 2 einzeln 120€ (ermäßigt 80€).

In der Teilnahmegebühr enthalten sind: Dokumentation des Workshops mit Präsentationen, Beispielprogrammen und Arbeitshilfen sowie Pausenverpflegung.

### Anmeldung und weitere aktuelle Informationen

Online-Anmeldung (ab 15. Juli 2013) und weitere Infos unter:

[www.uni-bielefeld.de/soz/datalab2013](http://www.uni-bielefeld.de/soz/datalab2013)

### Vorkenntnisse

In den computerbezogenen Lehreinheiten wird das Statistik-Softwarepaket Stata verwendet. Teilnehmern/innen ohne Stata-Kenntnisse wird die Einführungsveranstaltung am Montagmorgen empfohlen.

### Zielgruppe

Der Workshop richtet sich insbesondere an Doktorand/innen und Nachwuchswissenschaftler/innen,

- die für ihre theoretischen Fragestellungen geeignete Datensätze zur Sekundäranalyse kennenlernen wollen,
- und eine anwendungsbezogene Einführung in das Arbeiten mit komplexen Datensätzen mit individueller Betreuung suchen.

### Veranstalter

*Datalab2013* ist eine gemeinsame Veranstaltung des Sonderforschungsbereich 882 „Von Heterogenitäten zu Ungleichheiten“ an der Universität Bielefeld, der GESIS und des Sozio-ökonomischen Panels (SOEP) in Kooperation mit den Forschungsdatenzentren pairfam, ZEW und ifo sowie dem Datenservicezentrum für Betriebs- und Organisationsdaten (DSZ-BO).

### Lokale Veranstalter

Arbeitsbereich V Sozialstruktur und Soziale Ungleichheit, Fakultät für Soziologie, Universität Bielefeld (Prof. Dr. Stefan Liebig).  
Organisation: Jelena Hohlweg (DSZ-BO) und Regina Fischer.

### In Kooperation mit

Dr. Marco Giesselmann, Dr. Ingrid Tucci (SOEP)  
Dr. Daniel Fuß, Sabine Keller (pairfam)  
PD Dr. Oliver Arránz Becker (TU Chemnitz)  
Dr. Sandra Gottschalk (FDZ ZEW)  
Heike Mittelmeier, Christian Seiler (ifo Institut)  
Dr. Michael Weinhardt (SOEP-LEE)

**SFB 882**  
From Heterogeneities to Inequalities

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

**SOEP** Das Sozio-ökonomische Panel

**RatSWD**  
Rat für Sozial- und Wirtschaftsdaten