

FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

06/2011

DE

Panel „Arbeitsmarkt und soziale Sicherung“ – Die PASS Campus Files

Datensätze für den Einsatz in der
kognitiven WUI-WYb-Lehre

Antje Kirchner,
Lutz Gschwind



Panel „Arbeitsmarkt und soziale Sicherung“

Die PASS Campus Files

Datensätze für den Einsatz in der wissenschaftlichen Lehre

Antje Kirchner, Institut für Arbeitsmarkt- und Berufsforschung (IAB)
Lutz Gschwind, Universität Konstanz

Datensatz Version: PASS QCF 0607 v1 / PASS QCF 0607 v1

Dokumentation Version: PASS QCF 0607 v1 LCF 0609 v1 dok1.

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

Inhaltsverzeichnis

1	Das Panel Arbeitsmarkt und soziale Sicherung	5
2	Hintergründe	5
2.1	Datenstruktur der CFs	6
2.2	Stichprobenziehung	7
2.3	Merkmalsspektrum, Variablenbenennung und Identifikationsnummern	8
2.4	Anonymisierung: Informationsreduzierende und datenverändernde Verfahren	9
3	Arbeiten mit den CFs.....	10
4	Abschließende Hinweise und Datenzugang	13
Anhang 1: Haushaltsebene		15
Anhang 2: Personenebene		17

Tabellen und Abbildungsverzeichnis

Abbildung 1: Campus File Datensatzstruktur	7
Tabelle 1: Stichprobenumfang auf Haushaltsebene.....	8
Tabelle 2: Stichprobenumfang auf Personenebene	8
Tabelle 3: Beschreibung der in den Beispielen verwendeten Variablen	11
Tabelle 4: Lineare Regression am Beispiel der Wohnfläche	11
Tabelle 5: Multiple lineare Regression am Beispiel der Wohnfläche	12
Tabelle 6: Panel-Regression (Fixed-Effects-Modell) am Beispiel der Wohnfläche	12
Tabelle 7: Registerdaten (hh_register_xcf)	15
Tabelle 8: Interviewdaten (HHENDDAT_xcf)	15
Tabelle 9: Gewichte (hweights_xcf)	16
Tabelle 10: Registerdaten (p_register_xcf)	17
Tabelle 11: Interviewdaten (PENDDAT_xcf).....	17
Tabelle 12: Gewichte (pweights_xcf)	19

Zusammenfassung

Ausschließlich für die Nutzung in der akademischen Lehre an Hochschulen oder Forschungsinstituten hat das IAB Campus Files (CF), also absolut anonymisierte Daten, des Panels Arbeitsmarkt und soziale Sicherung (PASS) erzeugt. Der Datenreport gibt einen kurzen Überblick über die Campus Files und die Einschränkungen im Umgang mit diesen.

Abstract

Exclusively for use in academic teaching at universities or research institutes, the IAB generated absolutely anonymized data, so-called campus files (CF), based on the data of the „Panel Study Labour Market and Social Security“ (PASS). This documentation briefly provides a description of the data and its limitations.

Keywords: Campus File, Panel Arbeitsmarkt und soziale Sicherung (PASS), akademische Lehre.

Ein herzliches Dankeschön für das konstruktive Feedback und den Input während der gesamten Planungs- und Umsetzungsphase geht an dieser Stelle an das gesamte PASS-Team, sowie an S. Bender (FDZ), K. Drasch (FBE1), J. Drechsler (KEM), B. Matthes (FG BAM), E. Roß (Justizariat) und A. Schmucker (FDZ).

1 Das Panel Arbeitsmarkt und soziale Sicherung

Der Datensatz des Panels Arbeitsmarkt und soziale Sicherung (PASS) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) wurde aufgebaut um eine empirische Grundlage für Wissenschaft und Politikberatung im Bereich der Arbeitsmarkt-, Sozialstaats- und Armutsforschung zu schaffen. Im Rahmen der jährlichen Untersuchung konnten beispielsweise in der dritten Welle des Panels (Datenerhebung 2008/2009) mehr als 11.300 Personen und 8.200 Haushalte wiederholt befragt werden (Berg et al., 2010).

Der gesetzliche Auftrag der SGB II-Forschung am IAB ist es, die Wirkungen der Leistungen zur Eingliederung und zur Sicherung des Lebensunterhalts zu untersuchen. Die Studie ist im Rahmen dieser Forschung angesiedelt. Aufgrund des komplexen Stichprobendesigns erlaubt sie jedoch „auch die Beantwortung von Fragen, die über diesen Rahmen weit hinausgehen. Bei der Entwicklung der neuen Studie waren vor allem fünf Kernfragestellungen leitend, die in Achatz et al. (2007, [vgl. auch Trappmann et al., 2010; Bethmann und Gebhardt, 2011]) ausführlich erläutert werden:

1. Welche Wege führen aus dem Bezug von Arbeitslosengeld II?
2. Wie verändert sich die soziale Lage der betroffenen Haushalte im Leistungsbezug?
3. Wie entwickeln sich Bewältigungsprozesse und Handlungsorientierungen betroffener Personen?
4. Wie gestalten sich Kontakte zu den Trägern der Grundsicherung und wie sieht die institutionelle Handlungspraxis aus?
5. Welche Erwerbsverlaufsmuster oder Haushaltsdynamiken führen in den Bezug von Arbeitslosengeld II?“ (Berg et al., 2010:10).

Seit 2010 werden die PASS-Daten als Scientific Use Files (SUF) über das Forschungsdatenzentrum der BA im IAB (FDZ) für die nicht-kommerzielle Forschung angeboten. Darüber hinaus hat nun das IAB absolut anonymisierte Umfragedaten, sogenannte Campus Files (CF) des Panels Arbeitsmarkt und soziale Sicherung (PASS) erzeugt.

Ziel des vorliegenden Datenreports ist eine kurze Beschreibung dieser Campus Files. Der Querschnitt-Campus File (QCF) umfasst ausschließlich Querschnittsdaten der Welle 1 des PASS SUF und enthält ein umfangreicheres Merkmalspektrum als der Längsschnitt-Campus File (LCF). Der Längsschnitt-Campus File beinhaltet auf der einen Seite zwar weniger Merkmale, erlaubt jedoch das Arbeiten im Längsschnitt basierend auf den Wellen 1-3 des PASS SUF.

2 Hintergründe

Die Campus Files sind ausschließlich für die Nutzung in der akademischen Lehre an Hochschulen oder Forschungsinstituten bestimmt und können in diesem Rahmen frei verwendet werden. Die PASS Campus Files des IAB wurden speziell für Lehrzwecke entwickelt und sollen der anschaulichen Vermittlung von Erhebungs-, Datenmanagement- und Analyse-

techniken anhand verschiedener sozialwissenschaftlicher Fragestellungen dienen. Grundsätzlich sind diese Daten keinesfalls für inhaltliche Auswertungen irgendeiner Art oder Publikationen, also beispielsweise auch nicht zur Verfassung von Seminar- oder Bachelorarbeiten, geeignet. Es ist also nicht möglich mit diesen Daten generalisierbare Aussagen über einzelne Merkmale und Zusammenhänge zwischen verschiedenen Merkmalen zu treffen.¹

Mit den PASS CFs können verschiedene Erkenntnisse im Umgang mit den PASS Datensätzen vermittelt werden. Zusammengefasst, lassen sich die Ziele, die mit den PASS CFs erreicht werden sollen, wie folgt formulieren:

- 1) Einführung in den Umgang mit komplexen Datenstrukturen (Datenmanagement, Datenaufbereitung, Verwendung mehrerer Analyseebenen, etc.) am Beispiel des PASS,
⇒ Querschnitt-Campus File (QCF)

- 2) die Vermittlung spezieller Techniken zur Analyse von Paneldaten
⇒ Längsschnitt-Campus File (LCF)

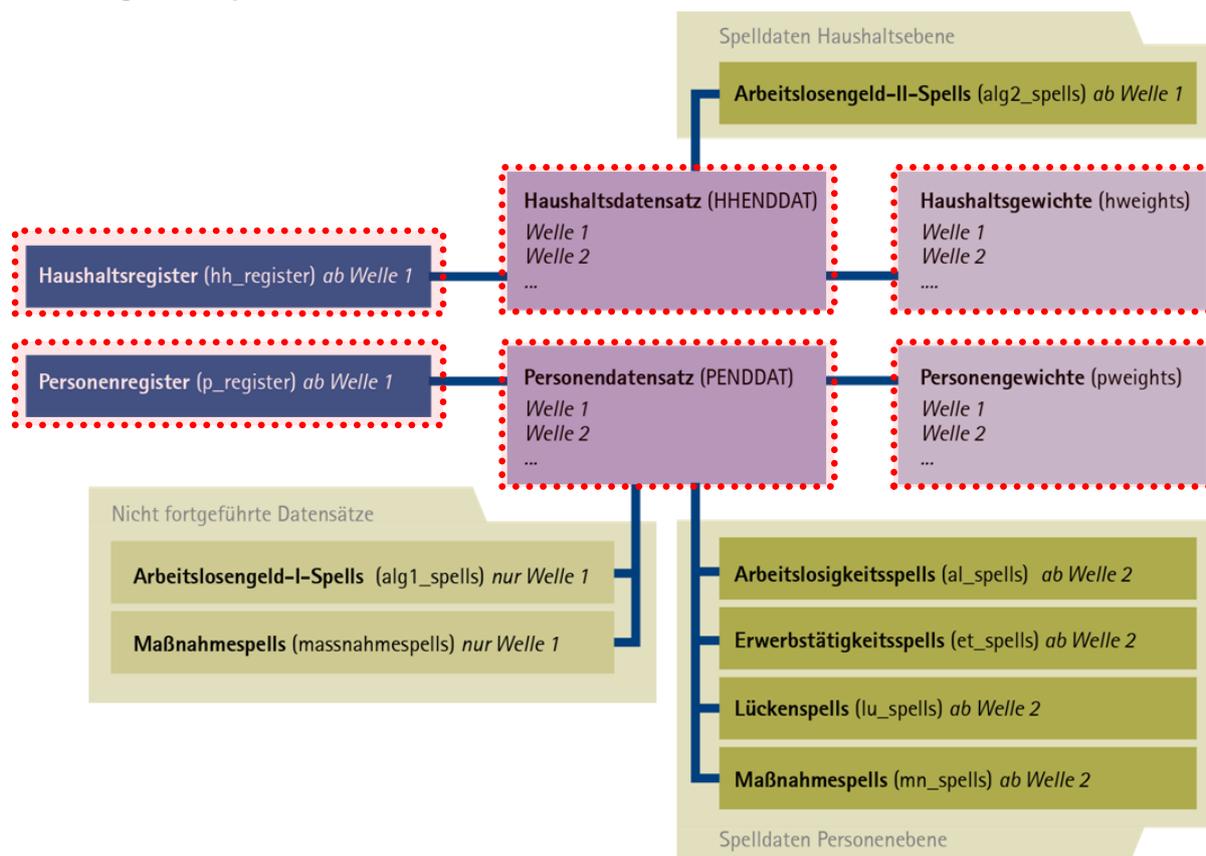
Der wesentliche Unterschied der beiden Campus Files besteht darin, dass der Querschnitt-Campus File ein breiteres Variablenspektrum umfasst als der Längsschnitt-Campus File. Der LCF erlaubt auf der anderen Seite Längsschnittanalysen, die mit dem QCF nicht möglich sind. Auch bezüglich des Stichprobenumfangs unterscheiden sich beide Datensätze, ansonsten sind sie strukturgleich.

2.1 Datenstruktur der CFs

Für beide PASS CFs wird die Komplexität der Datensätze leicht reduziert, ohne jedoch die maßgebliche Organisation der PASS Datensatzstruktur zu verändern. Aus den verschiedenen SUF Datensätzen (vgl. Abbildung 1) werden für die CFs nur die rot markierten Datensätze bereitgestellt. Neben den Personen- und Haushaltsregistern werden für beide Analyseebenen in separaten Datensätzen Gewichte bereitgestellt. Analog zum PASS SUF gibt es bei beiden CFs einen getrennten Personen- und Haushaltsdatensatz.

¹ Für valide inhaltliche Analysen muss unbedingt auf die Daten des Scientific Use Files des PASS zurückgegriffen werden. Informationen zum Datenzugang finden sich unter <http://fdz.iab.de/>.

Abbildung 1: Campus File Datensatzstruktur



2.2 Stichprobenziehung

Für die Campus Files wird zunächst die Datengrundlage des SUF auf der die Ziehung der Stichproben basiert eingeschränkt. Reine Seniorenhaushalte oder Haushalte, in denen mehr als 5 Personen leben, werden ausgeschlossen. Nachdem die Datengrundlage derart eingeschränkt ist, wurde aus der 3. Welle der PASS SUF-Daten eine 30% Stichprobe gezogen.

Auf der Personenebene wurden nach der Stichprobenziehung die Personeninterviews derjenigen verbliebenen Senioren im Personendatensatz entfernt, die nicht in reinen Seniorenhaushalten leben. Die Campus Files enthalten somit ausschließlich Personeninterviews mit Personen im Alter von 15-64 und keine Senioreninterviews und nur Haushalte mit weniger als sechs Personen.

Schließlich wurden die Haushalte der Stichprobe zufällig auf die zwei CFs aufgeteilt, sodass der QCF ca. 10% und der LCF ca. 20% der Haushalte des eingeschränkten SUF umfasst. Für den QCF wurden nur die Informationen der Welle 1 beibehalten. Die Daten aus den übrigen Wellen wurden gelöscht, während beim LCF die Informationen über die ersten drei Wellen des PASS erhalten bleiben.

Die letztendliche Verteilung der Haushalte und Personen auf die Datensätze verhält sich wie in Tabelle 1 und Tabelle 2 dargestellt.

Tabelle 1: Stichprobenumfang auf Haushaltsebene

	SUF	QCF	LCF
Welle 1	12.794	1.104	2.302
Welle 2	8.429	-	1.506
Welle 3	9.535	-	1,732
Insgesamt	30.758	1.104	5.540

Quelle: HENDDAT, Scientific Use File des IAB; HHENDDAT_lcf.dta und HHENDDAT_qcf.dta, Campus Files des IAB

Tabelle 2: Stichprobenumfang auf Personenebene

	SUF	QCF	LCF
Welle 1	18.954	1.623	3.312
Welle 2	12.487	-	2.110
Welle 3	13.439	-	2.349
Insgesamt	44.880	1.623	7.771

Quelle: PENDDAT, Scientific Use File des IAB; PENDDAT_lcf.dta und PENDDAT_qcf.dta, Campus Files des IAB

2.3 Merkmalspektrum, Variablenbenennung und Identifikationsnummern

Aus den PASS SUF Datensätzen wird lediglich eine eingeschränkte Auswahl an Variablen mit den Campus Files zur Verfügung gestellt, die durch verschiedene Maßnahmen absolut anonymisiert wurden. Beide Datensätze lassen nur die Bearbeitung einzelner, ausgewählter Fragestellungen zu. In Bezug auf das inhaltliche Variablenspektrum unterscheidet sich der umfassendere QCF vom LCF.

Als Beispiel lässt sich hier die Auswahl der Items zur Messung von Deprivation aufführen. Während im PASS SUF alle fünf Dimensionen, gemessen über 26 Items enthalten sind, sind im QCF nur neun der einzelnen Items enthalten, im LCF hingegen ausschließlich der zusammengefasste Index und keine einzelnen Items. Die neun im QCF enthaltenen Items umfassen außerdem nur drei der ursprünglich fünf Dimensionen, die mit dem Deprivationsindex abgebildet werden sollen. Im QCF enthalten sind die Dimensionen „Konsum“, „Teilhabe“ und „Finanzen“, während die Dimensionen „Wohnung“ und „Nahrung/Kleidung“ nicht enthalten sind. Im QCF wird außerdem zusätzlich zu dem Deprivationsindex basierend auf den SUF Daten, ein neuer, auf den im QCF enthaltenen Einzelitems basierender Index generiert.

Variablen, die in den CFs enthalten sind, erhalten in der Benennung jeweils den Zusatz „_qcf“ oder „_lcf“, wenn sie sich wesentlich in den Ausprägungen von den im SUF enthaltenen Variablen unterscheiden.

Sämtliche Identifikationsnummern, wie beispielsweise die Personen- oder die Haushaltsnummer wurden für die CFs neu generiert und haben eine Stelle weniger als im SUF. Die CF Haushaltsnummer (hnr) besteht aus einer siebenstelligen Ziffer. Die erste Stelle bezieht sich dabei auf den jeweiligen CF, die zweite Stelle gibt an, wann der Haushalt erstmalig an der

Befragung teilgenommen hat. Stelle drei und vier sind mit Nullen besetzt. Die Stellen fünf bis sieben enthalten eine laufende Haushaltsnummer. Die ID der ursprünglichen Haushaltsnummer (uhnr) wird dem neugenerierten Wert der Haushaltsnummer angepasst. Die Personennummer hat an der achten und neunten Stelle eine laufende Identifikationsnummer innerhalb des jeweiligen Haushaltes. Zuletzt werden auch die IDs der Primary Sampling Unit (psu) und der Stratified Primary Sampling Unit (strpsu) neu generiert.

Anhang 1 und Anhang 2 geben einen Überblick über die in den jeweiligen Datensätzen enthaltenen Variablen, Variablenbenennungen und deren Ausprägungen.

2.4 Anonymisierung: Informationsreduzierende und datenverändernde Verfahren

Grundsätzlich kamen zur Anonymisierung der PASS-Daten verschiedenste Techniken zum Einsatz, die aber zentral zwei Strategien verfolgen: Informationsreduktion und die Datenveränderung (vgl. Südfeld, 1987; Ronning et al., 2005; Zwick, 2008). Beide Strategien können sowohl auf kategoriale als auch auf metrische Variablen angewendet und/oder miteinander kombiniert werden. Eine detaillierte Beschreibung der einzelnen Verfahren findet sich bei Ronning et al. (2005: 55 ff.), weshalb diese hier nur kurz angeführt werden:

Unter die Verfahren zur Informationsreduktion fallen beispielsweise auf der

- | | |
|----------------------------|---|
| Merkmalsträger-Ebene: | das Entfernen auffälliger Merkmalsträger, oder auch die Ziehung von (Sub-) Stichproben |
| Merkmalsbezogenen-Ebene: | die Beseitigung, Ersetzung oder Zusammenfassung von Merkmalen sowie die Vergrößerung und Gruppierung von Merkmalsausprägungen |
| Ausprägungsbezogene-Ebene: | die Unterdrückung einzelner Werte bei seltenen oder einzigartigen Ausprägungskombinationen. |

Unter die Verfahren zur Datenveränderung fallen beispielsweise auf der

- | | |
|----------------------------|---|
| Merkmalsträger-Ebene: | „Klonen“, also Vervielfältigen von Merkmalsträgern oder Zensierungs- (top/bottom-coding) sowie Ersetzungs-Verfahren (Replacement, z.B. Ausreißer mit dem Mittelwert ersetzen) |
| Merkmalsbezogenen-Ebene: | Vertauschungsverfahren (Swapping), Post-Randomisierung oder auch SAFE-Verfahren; Imputationsverfahren oder Stochastische Überlagerung (Überlagerung mit Zufallsfehlern) |
| Ausprägungsbezogene-Ebene: | Mikroaggregationsverfahren. |

3 Arbeiten mit den CFs

Grundsätzlich wurde bei den CFs die gleiche Struktur erhalten wie bei den SUF-Daten. Es kann somit auf die Informationen (vgl. Trappmann et al., 2010) und Arbeitshilfen, die für den SUF zur Verfügung stehen, wie beispielsweise den Datenreport (Berg et al., 2010), den User Guide (Bethmann und Gebhardt, 2011) oder das Codebuch/die Dokumentation (Christoph et al., 2010), zurückgegriffen werden. Auch die Verknüpfung der einzelnen Datensätze erfolgt analog zum SUF.

Die wesentlichen Unterschiede im Arbeiten mit den CFs im Vergleich zum SUF bestehen in

- 1) einer reduzierten Auswahl der bereitgestellten Datensätze (vgl. Abschnitt 2.1);
- 2) einer reduzierten Fallzahl (vgl. Abschnitt 2.2);
- 3) einer eingeschränkten Merkmalsauswahl, teilweise veränderten Variablenbezeichnungen und veränderten Identifikationsnummern (vgl. Abschnitt 2.3 sowie Anhang 1 und Anhang 2)
- 4) Informationsreduktion, z.B. der Veränderung von Kategorien, und Datenveränderung auf allen Ebenen (vgl. Abschnitt 2.4 sowie Anhang 1 und

5) Anhang 2).

Es wird an dieser Stelle erneut darauf hingewiesen, dass die Campus Files keinesfalls für inhaltliche Analysen und Aussagen geeignet sind und daher nicht für solche verwendet werden dürfen. Lediglich die Daten des Scientific Use Files (SUF) genügen diesen Ansprüchen und können für valide inhaltliche Analysen beim IAB beantragt werden. Informationen zum Antragsverfahren finden sich unter <http://fdz.iab.de/>.

Im Folgenden werden kurz Möglichkeiten vorgestellt mit den CF des PASS zu arbeiten.² Die jeweiligen Ergebnisse werden außerdem den entsprechenden Ergebnissen aus dem SUF gegenübergestellt. Da die verwendeten Beispiele an äquivalente Übungen mit den verfremdeten Daten des Sozioökonomischen Panels von 1997 (kksoep), wie sie in Kohler und Kreuter (2008: 3 und 192ff) verwendet werden, angelehnt sind, werden auch diese Ergebnisse dargestellt.

Für deskriptive Analysen wurden die verwendeten Variablen in einem Vergleich der Datensätze auch unter Berücksichtigung der Designgewichte gegenübergestellt (vgl.

² Die Arbeitshilfen, die für den SUF zur Verfügung stehen, können auch für die CF verwendet werden (Bethmann und Gebhardt, 2011). Zur Verwendung der Gewichte im PASS gibt Trappmann (2011a, 2011b) eine Einführung.

Tabelle 3). Da für den Campus File die Auswahl der Haushalte vor der Stichprobenziehung systematisch eingeschränkt wurde, wurden die Berechnungen außerdem auch auf Basis dieser veränderten SUF-Stichprobe durchgeführt und in den Vergleich mit aufgenommen.

Tabelle 3: Beschreibung der in den Beispielen verwendeten Variablen

	SUF		SUF (W 1, CF- Auswahlgrundlage ³)		QCF		LCF	
	(W 1)				(W1)		(W1)	
	ungew.	gew.	ungew.	gew.	ungew.	gew.	ungew.	gew.
Wohnfläche	m ²		m ²		m ²		m ²	
arithmetisches Mittel	83	93	81	93	85	99	82	93
Standardfehler	0,37	1,10	0,39	1,12	1,54	3,40	0,94	1,95
Haushaltseinkommen	€/Monat		€/Monat		€/Monat		€/Monat	
arithmetisches Mittel	1.634	2.126	1.595	2.180	1.526	2.126	1.443	1.950
Standardfehler	16,30	39,96	15,15	36,14	33,03	67,36	21,81	46,76
Haushaltsgröße	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.
1 Personen	3.937	14.926	3.427	10.500	314	8.630	697	10.565
2 Personen	3.875	13.246	3.298	9.680	335	10.518	662	9.572
mehr als 2 Personen	4.982	10.900	4.672	10.438	455	10.350	943	10.441
Wohnform	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.	Anzahl	Anz. i. T.
Eigentümer	3.398	16.464	2.794	12.526	312	13.100	606	13.206
kein Eigentümer	9.396	22.608	8.603	18.092	792	16.401	1.696	17.371

Als Beispiel der Berechnungen einer linearen Regression im Querschnitt dient Tabelle 4. Hier wird die Wohnfläche durch das Haushaltseinkommen erklärt (ungewichtet). Hierfür wird zunächst das Haushaltseinkommen des jeweiligen CFs durch die Multiplikation des bedarfsgewichteten Haushaltseinkommens mit der Wurzel der Haushaltsgröße approximiert.⁴ Anschließend werden die Modelle für alle vier Datensätze (kksoep, SUF, QCF und LCF) berechnet.

Tabelle 4: Lineare Regression am Beispiel der Wohnfläche

	kksoep '97	SUF (W 1)	QCF (W1)	LCF (W1)
<i>Abhängige Variable:</i> Wohnfläche	sqm	wohnfl	wohnfl_qcf	wohnfl_lcf
<i>Unabhängige Variable:</i> Haushaltseinkommen	hhinc	hhincome	hhinc_qcf (gen.)	hhinc_lcf (gen.)
Koeffizient (Standardfehler)	0,017 (0,001)	0,011 (0,000)	0,025 (0,001)	0,021 (0,001)
Konstante	55,77	64,74	47,04	50,99
(Adj.) R ²	0,18	0,19	0,29	0,25
N	3.126	12.385	1.054	2.201

Zur Erweiterung dieses Modells mit nur einer erklärenden Variablen werden anschließend zur Erklärung der Wohnfläche in Tabelle 5 zwei weitere Variablen aufgenommen: Die Haus-

³ Die Einschränkung bezieht sich auf die Haushalte des SUFs, die der Stichprobenziehung für den QCF und LCF zugrunde liegt (vgl. Abschnitt 2.2).

⁴ Zur weiteren Erklärung der Vorgehensweise siehe Buhmann et al. (1988:119).

haltsgröße und eine binäre Variable für die Unterscheidung zwischen Wohnungseigentümern und Mietern.

Tabelle 5: Multiple lineare Regression am Beispiel der Wohnfläche

	kksoep '97	SUF (W 1)	QCF (W1)	LCF (W1)
<i>Abhängige Variable:</i> Wohnfläche	sqm	wohnfl	wohnfl_qcf	wohnfl_lcf
<i>Unabhängige Variable:</i> Haushaltseinkommen	hhinc	hhincome	hhinc_qcf (gen.)	hhinc_lcf (gen.)
Koeffizient (Standardfehler)	0,011 (0,001)	0,005 (0,000)	0,017 (0,001)	0,011 (0,001)
<i>Unabhängige Variable:</i> Haushaltsgröße	hhsiz	HA0100	hhgr_qcf	hhgr_lcf
Koeffizient (Standardfehler)	2,996 (0,485)	9,904 (0,218)	9,329 (1,068)	9,474 (0,664)
<i>Unabhängige Variable:</i> Eigentümer	owner	owner (gen.)	HW0300_qcf	HW0300_lcf
Koeffizient (Standardfehler)	35,760 (1,301)	40,757 (0,647)	32,422 (2,887)	35,325 (1,833)
Konstante	45,40	38,90	28,64	34,13
R ²	0,35	0,50	0,42	0,41
N	3.126	11.144	1.054	2.201

Zuletzt können diese Regressionen auch mit den LCF Daten berechnet werden. Um die Panelstruktur des PASS zu berücksichtigen, werden in Tabelle 6 die Ergebnisse dieser Fixed-Effects-Modelle für den SUF und den LCF dargestellt.

Tabelle 6: Panel-Regression (Fixed-Effects-Modell) am Beispiel der Wohnfläche

	SUF	LCF
<i>Abhängige Variable:</i> Wohnfläche	wohnfl	wohnfl_lcf
<i>Unabhängige Variable:</i> Haushaltseinkommen	hhincome	hhinc_lcf (gen.)
Koeffizient (Standardfehler)	0,001 (0,000)	0,004 (0,000)
<i>Unabhängige Variable:</i> Haushaltsgröße	HA0100	hhgr_lcf
Koeffizient (Standardfehler)	9,699 (0,160)	8,774 (0,441)
<i>Unabhängige Variable:</i> Eigentümer	owner (gen.)	HW0300_lcf
Koeffizient (Standardfehler)	41,003 (0,511)	33,369 (1,241)
Konstante	46,37	46,34
R ² (overall)	0,47	0,39
N	27.264	5.356

Beim Vergleich der Ergebnisse in Tabelle 4 fällt zunächst auf, dass der Zusammenhang zwischen Wohnfläche und Haushaltseinkommen in den beiden Campus Files höher ist. Dies ist auf verschiedene Anonymisierungsmaßnahmen zurückzuführen. Gleichzeitig steigen durch die Verwendung einer kleineren Stichprobe und Zufalls-Überlagerung der beiden metrischen Variablen die Standardfehler. Beide Auffälligkeiten bleiben, wenn auch in schwächerer Form, im multiplen Regressionsmodell erhalten (Tabelle 5). Der Effekt der Dummy-Variable und der Haushaltsgröße geht hingegen zurück, da die Variablen im Rahmen der Anonymisierung neu kategorisiert wurden. Alle Beobachtungen treffen auch auf die Berechnungen der Panelregression zu (Tabelle 6).

Durch verschiedene Maßnahmen zur Anonymisierung während der Erstellung des Campus Files wurden einerseits Zusammenhänge verstärkt, wie beispielsweise im linearen Modell, andererseits aber auch größere Varianz in die Berechnungen eingebracht. Dies zeigt sich u.a. in jeweils erwartungsgemäß höheren oder auch niedrigeren Werten für die erklärte Varianz (R^2), immer in Abhängigkeit davon welche Variablen ins jeweilige Modell eingehen und wie diese Variablen generiert/verändert wurden.

4 Abschließende Hinweise und Datenzugang

Campus Files (CF) sind absolut anonymisierte Datensätze, die vom IAB speziell für die Nutzung in der Lehre an Hochschulen und Forschungsinstituten entwickelt wurden. Aufgrund der umfangreichen Anonymisierungsverfahren, die bei der Generierung der CFs angewandt wurden, dienen die Daten ausschließlich der Vermittlung von Erhebungs-, Datenmanagement- und Analysetechniken. Für valide inhaltliche Analysen sind die CFs nicht geeignet. Hierfür sollten die entsprechenden Scientific Use Files oder schwach anonymisierten Versionen der zugrundeliegenden Datensätze verwendet werden.

Es werden zwei PASS Campus Files zur Verfügung gestellt. Ein Querschnitt-Campus File (QCF) und ein Längsschnitt-Campus File (LCF). Diese werden vom FDZ als kostenlose Downloads angeboten. Nach einer Registrierung und Zustimmung zu den Nutzungsbedingungen können die Daten mit einem Passwort abgerufen werden. Eine Weitergabe der Daten im Rahmen von Lehrveranstaltungen ist möglich, es ist dabei jedoch ausdrücklich auf die Nutzungsbedingungen und Einschränkungen im Arbeiten mit diesen hinzuweisen. Die kommerzielle Nutzung sowie eine kommerzielle Weitergabe sind untersagt.

Weitere Informationen zu den Campus Files finden Sie auf der FDZ-Homepage (http://fdz.iab.de/de/Campus_Files.aspx).

Literatur

- Achatz, Juliane, Hirsland, Andreas und Promberger, Markus (2007): IABPanelbefragung von Haushalten im Niedrigeinkommensbereich. Entwurf für ein Rahmenkonzept, S. 11-32 in Promberger, Markus (Hrsg.), Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung 'Arbeitsmarkt und soziale Sicherung', IAB Forschungsbericht 12/2007, Nürnberg.
- Bethmann, A.; Gebhard, D. (Hrsg.) (2011): User Guide „Panel Study Labour Market and Social Security“ (PASS). Wave 3. FDZ-Datenreport 4/2011 (EN), Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Berg, M.; Cramer, R.; Dickmann, C.; Gilberg, R.; Jesske, B.; Marwinski, K.; Gebhardt, D.; Wenzig, C.; Wetzel, M. (2010): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS) * Welle 3 (2008/2009). FDZ Datenreport, 06/2010, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Buhmann, W.; Rainwater, L.; Schmaus, G.; Smeeding, T. M. (1988): Equivalent Scales, Well-being, Inequality and Polity: Sensitivity Estimates Across Ten Countries using the Luxemburg Income Study (LIS) Data Base. In: The review of income and wealth, Jg. 34, H. 2, S.115-142.
- Christoph, B.; Müller, G.; Gebhardt, D.; Wenzig, C.; Trappmann, M.; Achatz, J. (2010): Codebuch und Dokumentation des 'Panel Arbeitsmarkt und soziale Sicherung' (PASS). Welle 3 (2008/2009). FDZ-Datenreport 6/2010. Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Kohler, U.; Kreuter, F. (2008): Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung. München: Oldenburg.
- Ronning, G.; Sturm, R.; Höhne, J.; Lenz, R. , Rosemann, M .; Scheffler, M.; Vorgrimler, D. (2005): Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Band 4. Statistik und Wissenschaft. Wiesbaden: Statistisches Bundesamt.
- Trappmann, M (2011a): Weighting, S. 51-61 in Bethmann, A. und Gebhard, D. (Hrsg.), User Guide „Panel Study Labour Market and Social Security“ (PASS). Wave 3. FDZ-Datenreport 4/2011 (EN), Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Trappmann, M (2011b): Weights, S. 77-96 in Bethmann, A. und Gebhard, D. (Hrsg.), User Guide „Panel Study Labour Market and Social Security“ (PASS). Wave 3. FDZ-Datenreport 4/2011 (EN), Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Trappmann, M.; Gundert, S.; Wenzig, C.; Gebhardt, D.; (2010): PASS: a household panel survey for research on unemployment and poverty. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 130, H. 4, S. 609-622.
- Südfeld, E. (1987): Anonymisierungsstandards und generelle Abwicklungsregeln für Anforderungen nach anonymisierten Einzelangaben im Statistischen Bundesamt. In: Statistisches Bundesamt (Hg.) Nutzung von anonymisierten Einzelangaben aus Daten der amtlichen Statistik. Forum der Bundesstatistik, Band 5. Stuttgart: Kohlhammer, S. 146-156.
- Zwick, M. (2008): CAMPUS Files. Kostenfreie Public Use Files für die Lehre. Wirt Sozialstat Archiv H. 2, S. 175-188.

Anhang 1: Haushaltsebene

Tabelle 7: Registerdaten (hh_register_xcf)⁵

Variable	QCF	.LCF	Beschreibung	Ausprägungen
hnr_xcf	✓	✓	Haushaltsnummer	laufende siebenstellige ID
uhnr_xcf	✗	✓	zur Identifikation von Splithaushalten benötigte ursprüngliche Haushaltsnummer	laufende siebenstellige ID
hnr1_xcf	✗	✓	Haushaltsnummer in Welle 1	laufende siebenstellige ID
hnr2_xcf	✗	✓	Haushaltsnummer in Welle 2	laufende siebenstellige ID
hnr3_xcf	✗	✓	Haushaltsnummer in Welle 3	laufende siebenstellige ID
sample	✓	✓	Information darüber, aus welcher Stichprobe die Person stammt	1 BA-Stichprobe 2 Microm-Stichprobe
hhgr1	✓	✓	Haushaltsgröße in Welle 1	bis zu 5 Personen
hhgr2	✗	✓	Haushaltsgröße in welle 2	bis zu 5 Personen
hhgr3	✗	✓	Haushaltsgröße in Welle 3	bis zu 5 Personen

Tabelle 8: Interviewdaten (HHENDDAT_xcf)

Variable	QCF	LCF	Beschreibung	Ausprägungen
hnr_xcf	✓	✓	Haushaltsnummer	laufende siebenstellige ID
uhnr_xcf	✗	✓	zur Identifikation von Splithaushalten benötigte ursprüngliche Haushaltsnummer	laufende siebenstellige ID
welle	✗	✓	aufgrund der Datenstruktur (long-format) benötigter Wellenindikator	Nummer der Welle zum Interviewzeitpunkt
psu_xcf	✓	✓	Systemvariable (v.a. für Gewichtung)	ID für die lokale Ebene des Stichproben-Designs
strpsu_xcf	✓	✓	Systemvariable (v.a. für Gewichtung)	Stratifizierte ID der Primary Sampling Unit
hhtyp_xcf	✓	✓	Typisierung des Haushalts	1 1-Personen HH 2 Paar ohne Kinder 3 1-Eltern HH 4 Paar mit Kindern 5 anderer Haushaltstyp
HW0300_xcf	✓	✓	Wohnverhältnis	1 Eigentümer 2 Mieter
wohnl	✓	✓	Wohnfläche	<i>metrisch</i>
HLS0800a	✓	✗	Deprivation: Auto?	1 Haben wir 2 Haben wir nicht
HLS0800b	✓	✗	Deprivation: Warum kein Auto?	1 Haben wir aus finanziellen Gründen nicht 2 Haben wir aus anderen Gründen nicht
HLS1000a	✓	✗	Deprivation: Videorekorder/DVD-Player?	1 Haben wir 2 Haben wir nicht
HLS1000b	✓	✗	Deprivation: Warum kein Videorekorder/DVD-Player?	1 Haben wir aus finanziellen Gründen nicht 2 Haben wir aus anderen Gründen nicht
HLS1100a	✓	✗	Deprivation: Computer mit Internetanschluss?	1 Haben wir 2 Haben wir nicht
HLS1100b	✓	✗	Deprivation: Warum kein Computer mit	1 Haben wir aus finanziellen Gründen nicht

⁵ Hinweis zur nachfolgenden Bezeichnung: im Kürzel „_xcf“ steht das „x“ als Platzhalter für die jeweilige Variablenbezeichnung im entsprechenden Campus File, bspw. QCF „_qcf“ während LCF „_lcf“.

			Internetanschluss?	2 Haben wir aus anderen Gründen nicht
HLS1600a	✓	✗	Deprivation: Eine einwöchige Urlaubsreise/Jahr?	1 Tun wir 2 Tun wir nicht
HLS1600b	✓	✗	Deprivation: Warum keine einwöchige Urlaubsreise/Jahr?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
HLS1700a	✓	✗	Deprivation: Einmal/Monat Freunde zum Essen einladen?	1 Tun wir 2 Tun wir nicht
HLS1700b	✓	✗	Deprivation: Warum nicht einmal/Monat Freunde zum Essen einladen?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
HLS1900a	✓	✗	Deprivation: Einmal/Monat ins Kino/Theater/Konzert gehen?	1 Tun wir 2 Tun wir nicht
HLS1900b	✓	✗	Deprivation: Warum nicht einmal/Monat ins Kino/Theater/Konzert gehen?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
HLS2000a	✓	✗	Deprivation: Festen Betrag sparen?	1 Tun wir 2 Tun wir nicht
HLS2000b	✓	✗	Deprivation: Warum nicht festen Betrag sparen?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
HLS2100a	✓	✗	Deprivation: Abgenutzte Möbel ersetzen?	1 Tun wir 2 Tun wir nicht
HLS2100b	✓	✗	Deprivation: Warum keine abgenutzten Möbel ersetzen?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
HLS2300a	✓	✗	Deprivation: Nicht von KV erstattete Behandlungen?	1 Tun wir 2 Tun wir nicht
HLS2300b	✓	✗	Deprivation: Warum keine nicht von KV erstattete Behandlungen?	1 Tun wir aus finanziellen Gründen nicht 2 Tun wir aus anderen Gründen nicht
depindg_xcf	✓	✗	Gewichteter Deprivationsindex basierend auf CF-Items	<i>metrisch</i>
depindg	✓	✓	Gewichteter Deprivationsindex basierend auf SUF-Items	<i>metrisch</i>
depindug_xcf	✓	✗	Ungewichteter Deprivationsindex basierend auf CF-Items	<i>metrisch</i>
alg2abez	✓	✓	Arbeitslosengeld II Bezug	1 Haushalt bezieht ALGII 2 Haushalt bezieht kein ALGII
oecdincn	✓	✓	bedarfsgewichtetes Haushaltseinkommen	<i>metrisch</i>

Tabelle 9: Gewichte (hweights_xcf)

Variable	QCF	LCF	Beschreibung	Ausprägungen
hnr_xcf	✓	✓	Haushaltsnummer	laufende siebenstellige ID
welle	✗	✓	aufgrund der Datenstruktur (long-format) benötigter Wellenindikator	Nummer der Welle zum Interviewzeitpunkt
dw_xcf	✓	✓	Designgewicht	<i>metrisch</i>
wqhh_xcf	✓	✓	kalibriertes Haushaltsgewicht	<i>metrisch</i>
hpbleib_xcf	✗	✓	Gewicht für die Bleibewahrscheinlichkeit	<i>metrisch</i>

Anhang 2: Personenebene

Tabelle 10: Registerdaten (p_register_xcf)

Variable	QCF	LCF	Beschreibung	Ausprägungen
pnr_xcf	✓	✓	Personennummer	laufende neunstellige ID
uhnr_xcf	✗	✓	zur Identifikation von Spliithaushalten benötigte ursprüngliche Haushaltsnummer	laufende siebenstellige ID
lasthnr_xcf	✗	✓	Letzte Haushaltsnummer	laufende siebenstellige ID
hnr1_xcf	✓	✓	Haushaltsnummer in Welle 1	laufende siebenstellige ID
hnr2_xcf	✗	✓	Haushaltsnummer in Welle 2	laufende siebenstellige ID
hnr3_xcf	✗	✓	Haushaltsnummer in Welle 3	laufende siebenstellige ID
alter1	✓	✓	Alter der Person in Welle 1	<i>metrisch</i>
alter2	✗	✓	Alter der Person in Welle 2	<i>metrisch</i>
alter3	✗	✓	Alter der Person in Welle 3	<i>metrisch</i>
zmhh1_xcf	✓	✓	Zeigervariable in Welle 1	Personennummer der Mutter der Zielperson
zmhh2_xcf	✗	✓	Zeigervariable in Welle 2	Personennummer der Mutter der Zielperson
zmhh3_xcf	✗	✓	Zeigervariable in Welle 3	Personennummer der Mutter der Zielperson
zvhh1_xcf	✓	✓	Zeigervariable in Welle 1	Personennummer des Vaters der Zielperson
zvhh2_xcf	✗	✓	Zeigervariable in Welle 2	Personennummer des Vaters der Zielperson
zvhh3_xcf	✗	✓	Zeigervariable in Welle 3	Personennummer des Vaters der Zielperson
zparthh1_xcf	✓	✓	Zeigervariable in Welle 1	Personennummer des Partners der Zielperson
zparthh2_xcf	✗	✓	Zeigervariable in Welle 2	Personennummer des Partners der Zielperson
zparthh3_xcf	✗	✓	Zeigervariable in Welle 3	Personennummer des Partners der Zielperson

Tabelle 11: Interviewdaten (PENDDAT_xcf)

Variable	QCF	LCF	Beschreibung	Ausprägungen
pnr_xcf	✓	✓	Personennummer	laufende neunstellige ID
hnr_xcf	✓	✓	Haushaltsnummer	laufende siebenstellige ID
uhnr_xcf	✗	✓	zur Identifikation von Spliithaushalten benötigte ursprüngliche Haushaltsnummer	laufende siebenstellige ID
welle	✗	✓	aufgrund der Datenstruktur (long-format) benötigter Wellenindikator	Nummer der Welle zum Interviewzeitpunkt
zpsex	✓	✓	Geschlecht	1 Männlich 2 Weiblich
PD0400	✓	✓ (W1 & W2)	Selbsteinstufung Religiosität	1 überhaupt nicht religiös 2 eher nicht religiös 3 eher religiös 4 sehr religiös
PA0300	✓	✗	Zufriedenheit mit dem Lebensstandard	0 (ganz und gar unzufrieden) 1 2 3

				4 5 6 7 8 9 10 (ganz und gar zufrieden)
PA0800	✓	✓	Soziale Teilhabe	0 (Ausgeschlossen) 1 2 3 4 5 6 7 8 9 10 (Dazugehörig)
PA0900	✓	✓	Soziale Position: Oben-Unten-Skala	0 (unten) 1 2 3 4 5 6 7 8 9 10 (oben)
PA1000	✓	✓	Generelle Lebenszufriedenheit	0 (ganz und gar unzufrieden) 1 2 3 4 5 6 7 8 9 10 (ganz und gar zufrieden)
PEO0100a	✓	✗	Kompetenz: für jedes Problem eine Lösung	1 Trifft voll und ganz zu 2 Trifft eher zu 3 Trifft eher nicht zu 4 Trifft überhaupt nicht zu
PEO0100b	✓	✗	Kompetenz: Auch mit überraschenden Ereignissen zurecht kommen	1 Trifft voll und ganz zu 2 Trifft eher zu 3 Trifft eher nicht zu 4 Trifft überhaupt nicht zu
PEO0100c	✓	✗	Kompetenz: Keine Schwierigkeiten Ziele zu verwirklichen	1 Trifft voll und ganz zu 2 Trifft eher zu 3 Trifft eher nicht zu 4 Trifft überhaupt nicht zu
PEO0100d	✓	✗	Kompetenz: Handeln in unerwarteten Situationen	1 Trifft voll und ganz zu 2 Trifft eher zu 3 Trifft eher nicht zu 4 Trifft überhaupt nicht zu
PEO0100e	✓	✗	Kompetenz: Lösung schwieriger Probleme gelingt immer	1 Trifft voll und ganz zu 2 Trifft eher zu 3 Trifft eher nicht zu

				4 Trifft überhaupt nicht zu
PG1200	x	✓	subjektive Beurteilung des Gesundheitszustandes	1 sehr gut 2 gut 3 zufriedenstellend 4 weniger gut 5 schlecht
PSK0100	✓	x	Angaben über enge Freunde/ Familienmitglieder außerhalb des HH	1 Ja 2 Nein
PSK0200_xcf	✓	x	Anzahl enger Freunde	1 bis zu 5 2 6 bis 10 3 11 bis 20 4 mehr als 20
ekind_xcf	✓	✓	Information über eigene Kinder	1 Zielperson ist ein (Stief)Elternteil eines Kindes 2 Zielperson ist kein (Stief-)Elternteil eines Kindes
schul2_xcf	✓	✓	höchster Schulabschluss	1 kein Abschluss/ Sonder-/Förderschulabschluss 2 Volks-/Hauptschulabschluss bzw. POS 8./9. Klasse 3 Mittlere Reife/Realschulabschluss bzw. POS 10. Klasse 4 Allgem. oder fachgeb. Hochschulreife (Abitur)/EOS 12. Klasse
beruf2_xcf	✓	✓	höchster beruflicher Abschluss	1 kein Abschluss 2 äquivalent zu Ausbildung 3 äquivalent zu Studium
brutto	✓	✓	Bruttoeinkommen	<i>metrisch</i>
erwerb_xcf	✓	✓ (W1)	Angaben zum Erwerbsstatus (alle Befragten)	1 Erwerbstätig (> 16h/Woche) 2 Sonstige
et_xcf	x	✓ (W2 & W3)	Angaben zum Erwerbsstatus (ohne Schüler/Berufsschüler oder Azubis)	1 Erwerbstätig (> 400 Euro/Monat) 2 Nicht-Erwerbstätig
arbeitszeit_xcf	✓	x	Informationen zur Arbeitszeit	1 Teilzeit (1-35 Stunden/Woche) 2 Vollzeit (35 und mehr Stunden/Woche)
isei	✓	x	Klassifikation des Berufs nach ISEI	<i>metrisch</i>
migration_xcf	✓	✓	Migration(shintergrund)	1 Kein Migrationshintergrund (inkl.: Mind. 1 Großelternanteil zugezogen) 2 Selbst zugezogen/Mind. 1 Elternteil zugezogen

Tabelle 12: Gewichte (pweights_xcf)

Variable	QCF	LCF	Beschreibung	Ausprägungen
pnr_xcf	✓	✓	Personennummer	laufende neunstellige ID
welle	x	✓	aufgrund der Datenstruktur (long-format) benötigter Wellenindikator	Nummer der Welle zum Interviewzeitpunkt
wqp_xcf	✓	✓	kalibriertes Personengewicht	<i>metrisch</i>
pbleib_xcf	x	✓	Gewicht für die Bleibewahrscheinlichkeit	<i>metrisch</i>

Impressum

FDZ-Datenreport 6/2011

Herausgeber

Forschungsdatenzentrum (FDZ)
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Str. 104
90478 Nürnberg

Redaktion

Stefan Bender, Iris Dieterich

Technische Herstellung

Iris Dieterich

Rechte

Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit

http://doku.iab.de/fdz/reporte/2011/DR_06-11.pdf

Internet

<http://fdz.iab.de/>

Rückfragen zum Inhalt an:

Antje Kirchner,
Institut für Arbeitsmarkt- und
Berufsforschung (IAB),
Regensburger Str. 104,
90478 Nürnberg
Telefon: 0911 / 179-6050
E-Mail: antje.kirchner@iab.de