

Forschungsdatenzentrum

der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt-
und Berufsforschung

FDZ

FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

05/2018

DE

IAB-Stellenerhebung

Betriebsbefragung zu Stellenangebot und
Besetzungsprozessen

Wellen 2000 bis 2015 mit Folgequartalen von 2006
bis 2016

Alexander Kubis,
Andreas Moczall,
Martina Rebien



Bundesagentur für Arbeit

IAB-Stellenerhebung

Betriebsbefragung zu Stellenangebot und Besetzungsprozessen
Wellen 2000 bis 2015 mit Folgequartalen von 2006 bis 2016

Alexander Kubis,
Andreas Moczall,
Martina Rebien

Datensatz: IAB-Stellenerhebung 2000-2015 v1
DOI: 10.5164/IAB.IABSE0015.de.en.v1

Dokumentation: IAB-Stellenerhebung 2000-2015 DE v1 dok1
DOI: 10.5164/IAB.FDZD.1805.de.v1

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	Änderungen am Datensatz gegenüber früheren Versionen.....	5
2	Beschreibung des Datensatzes	6
2.1	Überblick	7
2.2	Grundgesamtheit und Stichprobenziehung der Hauptbefragung	8
2.3	Hochrechnungsverfahren der Hauptbefragung	9
2.4	Stichprobenziehung und Hochrechnung der Quartalsbefragungen	10
3	Verwendungshinweise für Datennutzer	10
3.1	Struktur des Datensatzes	10
3.2	Verwendung der Gewichte	11
3.3	Fragebogenteile mit möglichen Mehrfachantworten	12
3.4	Klassifikationen von Wirtschaftszweigen und Berufen	13
3.5	BHP-Verlinkung	16
3.6	STATA-Auswertungsbeispiele	17

Zusammenfassung

Die IAB-Stellenerhebung ist eine quartalsweise durchgeführte Betriebsbefragung, die repräsentativ das gesamtwirtschaftliche Stellenangebot in Deutschland abbildet. Sie ermittelt die Gesamtzahl aller offenen Stellen am Arbeitsmarkt, einschließlich jener Stellen, die nicht der Bundesagentur für Arbeit (BA) gemeldet werden. Der Hauptfragebogen enthält Informationen zu Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. Der Zusatzfragebogen enthält Fragen zum letzten Fall einer Neueinstellung und zum letzten Fall eines gescheiterten Rekrutierungsversuchs. Der Sonderfragebogen erfragt die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente. Im Forschungsdatenzentrum der Bundesagentur für Arbeit werden die Befragungswellen ab 2000 mit allen Fragebogenteilen einschließlich der Quartalsbefragungen für externe Wissenschaftlerinnen und Wissenschaftler angeboten.

Abstract

The IAB Job Vacancy Survey is a quarterly establishment survey representing the entire unfilled labour demand in Germany. It identifies the entire number of vacancies on the German labour market, including those vacancies that are not reported to the Federal Employment Agency (FEA). The main questionnaire enquires information about the number and structure of vacancies, future personnel requirements, about the current economic situation and the expected development of participating establishments. The additional questionnaire enquires information about the last new hiring and the last case of a failed recruitment effort. The special questionnaire enquires employer attitudes and firm use of current labour market instruments. The Research Data Centre of the Federal Employment Agency offers the data sets from the survey waves 2000 on.

Keywords: Betriebsbefragung, offene Stellen, Stellenbesetzungsprozesse

Wir danken Benjamin Börschlein für die Mithilfe bei der Erstellung der Datendokumentation der IAB-Stellenerhebung.

1 Änderungen am Datensatz gegenüber früheren Versionen

1. In den Wellen 2000 bis 2004 wurden bei den Such- und Besetzungswegen die (für externe Nutzer nicht zur Verfügung gestellten) offenen Antworten bei „Sonstiger Suchweg“ nun umcodiert in die entsprechenden *iu-Variablen („inklusive Umbuchungen“), vgl. Abschnitt 3.3.
2. Bei sämtlichen offenen Textfeldern ist nun die entsprechende „Sonstiges“-Variable auch 2000 bis 2004 entsprechend auf 1 gesetzt.
3. Für die Wellen ab 2010 steht nun erstmals eine Verlinkung mit den administrativen Daten des Betriebs-Historik-Panels zur Verfügung. Nähere Informationen hierzu siehe Abschnitt 3.5.
4. Neue technische Merkmale: `gesamtausfueller`, `erh_form` und `antwort` (vgl. Tabelle 4).

2 Beschreibung des Datensatzes

Tabelle 1: Kurzbeschreibung der IAB-Stellenerhebung

Inhaltliche Charakteristika	
Themen/Merkmalgruppen	Beschäftigung, Personalbedarf, Aktivitätshemmnisse, abgebrochene Personalsuche, Berufe für die steigender Bedarf erwartet wird mit erwarteten Engpässen in den kommenden 3 Jahren, geförderte Beschäftigung, geringfügige Beschäftigung, Teilzeitbeschäftigung, Reform des Arbeitsmarktes, Beschäftigungschancen arbeitsloser Personen, Weiterbildung der Belegschaft, Ein-Euro-Jobs (Erfahrungen der Betriebe), detaillierte Informationen zur letzten Einstellung, detaillierte Informationen zur letzten abgebrochenen Personalsuche
Untersuchungseinheit	Betriebe
Fallzahlen	7.500 bis 15.000 Betriebe
Zeitraum	IV.2000 bis III.2016
Zeitbezug	quartalsweise
Regionale Gliederung	Ost-/Westdeutschland, Bundesländer, weitere Merkmale siehe unter "sensible Merkmale"
Methodische Charakteristika	
Erhebungsdesign	repräsentative Stichprobe der Betriebe, geschichtet nach Betriebsgröße, Branche und West-/Ostdeutschland
An der Erhebung beteiligte Institutionen	Auftraggeber: Forschungsbereich A1 des IAB Durchführung: Economix Research & Consulting, München
Frequenz der Datensammlung	Jährliche schriftliche und quartalsweise telefonische Befragung
Dateiformat/-größe	STATA; alle Datensätze 200 MB
Dateiorganisation	nach Welle
Datenzugang	
Datenzugang	Datenfernverarbeitung, Gastaufenthalt
Anonymisierungsgrad	Schwach anonymisiert
sensible Merkmale	Wirtschaftszweigklassifikationen: WZ73 3-Steller (2000-2003), WZ03 3-Steller (2004)/5-Steller (2005–2009) WZ08 5-Steller (ab 2010), Kreisnummer (ab 2007), Regierungsbezirk (ab 2005), Konsistente Wirtschaftszweige
Zusatzmerkmale	Es können weitere Betriebsmerkmale aus dem Betriebs-Historik-Panel (BHP) sowie die BHP-Erweiterungsmodule zur Zuspiegelung an die IAB-Stellenerhebung beantragt werden. Welche Variablen der IAB-Stellenerhebung zugespielt werden können, entnehmen Sie bitte der Liste der BHP-Merkmale .
Vorgaben zum Zitieren der Daten und Datendokumentation	Daten: "Die Datengrundlage dieses Beitrags bildet die IAB-Stellenerhebung, Welle(n) JJJJ - JJJJ. Der Datenzugang erfolgte über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und/oder mittels kontrollierter Datenfernverarbeitung beim FDZ." Datendokumentation: - Kubis, Alexander; Moczall, Andreas; Rebien, Martina (2018): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Wellen 2000 bis 2015 und Folgequartale ab 2006. FDZ-Datenreport, 05/2018 (de), Nürnberg - Kubis, Alexander; Moczall, Andreas; Rebien, Martina (2018): The IAB Job Vacancy Survey. Establishment survey on labour demand and recruitment processes Waves 2000 to 2015 and subsequent quarters since 2006. FDZ-Datenreport, 05/2018 (en), Nuremberg.

Details zu verschiedenen Zugangswegen, zu den Voraussetzungen der Nutzung sowie zur Antragstellung des Datensatzes können der FDZ-Internetseite unter <http://fdz.iab.de/> entnommen werden.

2.1 Überblick

Die IAB-Stellenerhebung ist eine Betriebsbefragung, die repräsentativ das gesamtwirtschaftliche Stellenangebot in Deutschland abbildet. Sie ermittelt die Gesamtzahl aller offenen Stellen am Arbeitsmarkt, einschließlich jener Stellen, die nicht der Bundesagentur für Arbeit (BA) gemeldet werden. Allein diese Gesamtzahl erlaubt valide und unverzerrte Aussagen über die gesamtwirtschaftliche Arbeitsnachfrage. Denn die Meldequote, also der Anteil der BA-gemeldeten an allen offenen Stellen, unterliegt im Zeitverlauf starken Schwankungen und unterscheidet sich systematisch zwischen Wirtschaftszweigen, Regionen und Berufen.

Die Erhebung wird seit 1989 im vierten Quartal jedes Jahres schriftlich mit einem mehrteiligen Fragebogen durchgeführt. Seit 2002 können Betriebe an der Befragung auch online über Internet teilnehmen. Dieser Webfragebogen ist mit dem schriftlichen Fragebogen identisch. Der **Hauptfragebogen** enthält Informationen zu Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. Der **Zusatzfragebogen** enthält Fragen zum letzten Fall einer Neueinstellung und zum letzten Fall eines gescheiterten Rekrutierungsversuchs. Diese regelmäßige und detaillierte Abbildung von Stellenbesetzungsprozessen ist weltweit ein Alleinstellungsmerkmal der IAB-Stellenerhebung. Sie liefert damit Informationen zu Merkmalen der besetzten Stelle, der eingestellten Person, zu Such- und Besetzungswegen einschließlich des Einschaltungsgrades der Arbeitsvermittlung, zu Such- und Besetzungszeiten, Bewerberzahlen, Besetzungsschwierigkeiten sowie über gemachte Kompromisse. Der Zusatzfragebogen wird nur von Betrieben beantwortet, die im letzten Jahr Mitarbeiter neu eingestellt oder dies versucht haben.

Haupt- und Zusatzfragebogen sind von Welle zu Welle im Wesentlichen gleich, abgesehen von redaktionellen Änderungen und der Verlagerung von Schwerpunkten. Der **Sonderfragebogen** ändert sich hingegen mit jeder Befragungswelle. Im Sonderfragebogen wird seit 2000 die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente erfragt. Von 2005 bis 2013 beschäftigte er sich hauptsächlich mit der betrieblichen Nutzung und Einschätzung von Ein-Euro-Jobs. Er ging in diesem Zeitraum deshalb nur an die Wirtschaftszweige im öffentlichen Sektor, in welchen die überwiegende Zahl der Ein-Euro-Jobs angeboten und durchgeführt wurde.

Seit 2005 finden im 1., 2. und 3. Quartal des jeweiligen Folgejahres einer schriftlichen Befragungswelle kurze telefonische Nachbefragungen zu den Kernfragen unter den Teilnehmern der Hauptbefragung statt. Diese telefonischen **Quartalsbefragungen** ergänzen die schriftliche Hauptbefragung im jeweils 4. Quartal. Erfragt und aktualisiert werden die wichtigsten Eckdaten aus dem Hauptfragebogen. Hierunter fällt u.a. die Zahl der Beschäftigten eines Betriebes, seine Einschätzung der Beschäftigungsentwicklung und die Zahl der offenen Stellen.

Im Forschungsdatenzentrum der Bundesagentur für Arbeit werden die Befragungswellen ab 2000 mit allen Fragebogenteilen einschließlich der Quartalsbefragungen für externe Wissenschaftlerinnen und Wissenschaftler angeboten. Seit September 2015 stehen die Variablen-

und Wertelabels des Datensatzes wie auch die vollständige Dokumentation einschließlich der Codepläne und Datenreports auch in englischer Sprache zur Verfügung.

2.2 Grundgesamtheit und Stichprobenziehung der Hauptbefragung

Grundgesamtheit der Befragung im 4. Quartal eines jeweiligen Jahres sind sämtliche Betriebe mit mindestens einem sozialversicherungspflichtig Beschäftigten im Juni (bis Welle 2004) bzw. Dezember (ab Welle 2005) des *Vorjahres* unter Ausschluss der Privathaushalte. „Betrieb“ bezieht sich dabei auf eine wirtschaftliche Einheit nach dem Betriebsnummernkonzept des Betriebsnummern-Service der Bundesagentur für Arbeit (Bundesagentur für Arbeit 2013).

Aus dieser Grundgesamtheit wird jedes Jahr erneut eine disproportional geschichtete **Zufallsbruttostichprobe** gezogen. Geschichtet wird nach der Region, Betriebsgrößenklasse sowie nach Wirtschaftszweig, sodass eine dreidimensionale Stichprobenmatrix entsteht. Tabelle 2 fasst die Schichtungsmerkmale zusammen. In den Wellen 2005 bis 2013 wurde darüber hinaus eine Ergänzungsstichprobe im öffentlichen Sektor für die Erhebung der Ein-Euro-Jobs nach gleichem Schema gezogen. Nur in der Welle 2005 wurde außerdem aufgrund des geringen Rücklaufs nach Beginn der Feldphase eine weitere Ergänzungsstichprobe zur Erhöhung der Fallzahlen gezogen. In beiden Fällen wurden vor der Ziehung der Ergänzungsstichproben die bereits gezogenen Betriebe aus der Ziehungsgesamtheit entfernt.

Tabelle 2: Schichtungsmerkmale der Bruttostichprobe 2000–2015

Merkmale	Unterteilungen
Region	2000–2003: alte Bundesländer mit West-Berlin, neue Bundesländer mit Ost-Berlin ab 2004: Westdeutschland, Ostdeutschland mit Berlin
Betriebsgröße (Anzahl aller Beschäftigter)	2000–2003: 1–9, 10–19, 20–49, 50–199, 200–499, 500+ 2004: 1–9, 10–19, 20–49, 50–99, 100–199, 200–499, 500–999, 1000+ 2005–2013: 1–9, 10–19, 20–49, 50–199, 200–499, 500–999, 1000+ ab 2014: 1–9, 10–19, 20–49, 50–249, 250–499, 500–999, 1000+
Wirtschaftszweig	2000–2002: 22 Wirtschaftszweige auf Basis der Klassifikation 1973 2003: 14 Wirtschaftszweige auf Basis der Klassifikation 1973 2004–2009: 28 Wirtschaftszweige auf Basis der Klassifikation 2003 2010–2014: 23 Wirtschaftszweige auf Basis der Klassifikation 2008 ab 2015: 24 Wirtschaftszweige auf Basis der Klassifikation 2008

Aus dieser Bruttostichprobe werden all jene Betriebe ausgeschlossen, die bei vergangenen Befragungen des IAB geäußert haben, nicht mehr zur Teilnahme aufgefordert werden zu wollen, um zur bereinigten **Versandstichprobe** zu gelangen. Diese bildet die Basis für die in Tabelle 3 angegebene Rücklaufquote.

Sämtliche Betriebe der Versandstichprobe werden Anfang Oktober schriftlich per Post kontaktiert und zur Teilnahme aufgefordert. Die Sendung umfasst ein Anschreiben, eine Datenschutzerklärung, den Fragebogen sowie ein Blatt mit Ausfüllhinweisen und Antworten auf häufig gestellte Fragen. Wenige Wochen später wird eine zweite Postsendung versandt, deren Anschreiben an die Teilnahme erinnert und ansonsten den gleichen Umfang und Inhalt hat. Weiterhin ist auf den Anschreiben vermerkt, telefonische Rückfragen an eine ganztägig besetzte Hotline richten zu können. Die Feldphase dauert in der Regel bis Anfang Januar, wobei das

Gro des Rücklaufes Mitte Oktober/Anfang November bzw. Mitte November/Anfang Dezember erfolgt.

Tabelle 3: Stichprobenentwicklung 2000–2015

Welle	Adressbestand Stichtag	Durchschn. Auswahlsatz	Brutto- stich- probe	Versand- stichprobe	Netto- stich- probe	Rücklauf- quote
2000	31.12.1999	1,3%	28.486	28.266	7.578	26,8%
2001	30.06.2000	1,3%	27.994	27.827	7.347	26,4%
2002	30.06.2001	1,3%	27.418	27.147	5.773	21,3%
2003	30.06.2002	1,8%	37.789	37.421	7.310	19,5%
2004	30.06.2003	2,7%	56.926	56.699	11.707	20,6%
2005	31.12.2004	3,9%	78.032	69.702	11.742	16,8%
2006	31.12.2005	3,9%	75.290	69.231	13.537	19,6%
2007	31.12.2006	3,8%	75.128	73.635	14.381	19,5%
2008	31.12.2007	3,8%	77.543	75.035	13.652	18,2%
2009	31.12.2008	3,8%	77.537	74.998	15.288	20,4%
2010	31.12.2009	3,7%	77.739	75.000	15.124	20,2%
2011	31.12.2010	3,6%	77.685	74.660	15.139	20,3%
2012	31.12.2011	3,6%	79.181	75.006	13.807	18,4%
2013	31.12.2012	3,6%	77.214	75.486	14.019	18,6%
2014	31.12.2013	3,6%	76.767	75.073	12.750	17,0%
2015	31.12.2014	3,9%	85.556	83.639	12.852	15,4%

2.3 Hochrechnungsverfahren der Hauptbefragung

Einzelheiten und Hintergründe zum verwendeten Hochrechnungsverfahren liefert der IAB-Forschungsbericht Nr. 4/2016 (Brenzel et al. 2016). Zur Berechnung der Betriebsgewichte werden folgende Schritte durchgeführt:

1. Berechnung der **Designgewichte** für jede Zelle der Stichprobenmatrix als Inverse des jeweiligen Auswahlsatzes.
2. Multiplikation der Designgewichte mit dem Gewicht aus einer **Non-Response-Modellierung**. Dies erfolgt auf Basis eines logistischen Regressionsmodells, bei dem die Betriebsgrößenklasse, der Wirtschaftszweig, das durchschnittliche Tagesentgelt sowie das Durchschnittsalter der Beschäftigten den administrativen Daten der Bundesagentur für Arbeit entnommen und als Hilfsvariablen verwendet werden. Das Non-Response-Gewicht ist die Inverse der geschätzten Antwortwahrscheinlichkeit.
3. **Kalibrierung** der Gewichte aus Schritt 2 als Startgewichte eines generalisierten Regressionsschätzers (GREG) mit dem Ziel, Eckwerte der Ankervariablen „Zahl der Betriebe“ und „Anzahl der sozialversicherungspflichtigen Beschäftigten“ in allen Zellen der Stichprobenmatrix exakt zu erreichen.

Da die Ankervariable „Anzahl der sozialversicherungspflichtig Beschäftigten“ am aktuellen Rand durch Fortschreibung geschätzt werden muss, wird die Kalibrierung nach jeweils 18 Monaten mit der tatsächlichen Zahl wiederholt und die Hochrechnungsgewichte somit revidiert. Der FDZ-Datensatz enthält ausschließlich die in dieser Weise revidierten Hochrechnungsgewichte.

wichte; daher kann jede neue Welle frühestens 18 Monate nach Abschluss der Erhebungsphase im FDZ eingestellt werden. Zuvor veröffentlichte Hochrechnungsergebnisse des IAB in Presseinformationen und IAB-Kurzberichten basieren hingegen in der Regel auf den vorläufigen Hochrechnungsgewichten und können deshalb mit den im FDZ zur Verfügung gestellten Hochrechnungsgewichten nicht exakt repliziert werden; die Unterschiede bewegen sich jedoch immer innerhalb des jeweiligen 95%-Konfidenzintervalls.

Für die rückwirkende Anwendung des neuen Hochrechnungsverfahrens auf die Wellen 2000 bis 2013 wurden die über die Jahre sich verändernden Betriebsgrößenklassen der ursprünglichen Stichprobenmatrizen (vgl. Tabelle 2) auf eine einheitliche Abgrenzung von sechs Betriebsgrößenklassen (1–9, 10–19, 20–49, 50–249, 250–499, 500+) umgebucht. Die Wirtschaftszweige wurden von 2000 bis 2009 einheitlich auf 16 Ausprägungen der Klassifikation der Wirtschaftszweige, Ausgabe 2003, umgebucht; ab 2010 existieren bereits in der Stichprobenmatrix die nach wie vor verwendeten 23 Ausprägungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008.

2.4 Stichprobenziehung und Hochrechnung der Quartalsbefragungen

Aus den Teilnehmern an der schriftlichen Hauptbefragung im vierten Quartal werden zufällig Betriebe für eine kurze telefonische Nachbefragung in den drei Folgequartalen gezogen, bis in jedem Quartal mindestens 9.000 Betriebe vorliegen. Die Nettostichprobe der schriftlichen Hauptbefragung stellt damit die Bruttostichprobe für die telefonischen Nachbefragungen dar.

Für jedes der Folgequartale wird ein eigenes Betriebsgewicht berechnet, welches sich aus dem Betriebsgewicht der Hauptbefragung im 4. Quartal des Ursprungsjahres und dem Ergebnis einer weiteren Non-Response-Modellierung für das jeweilige Folgequartal zusammensetzt.

3 Verwendungshinweise für Datennutzer

3.1 Struktur des Datensatzes

Die Daten jeder Befragungswelle sind in jeweils einer Datendatei im STATA-Format abgelegt. Sie enthält die Informationen des Hauptfragebogens, Zusatzfragebogens und Sonderfragebogens sowie ab 2005 die Angaben aus den telefonischen Folgebefragungen in den drei Folgequartalen. Es sind also nicht jeweils die vier Quartale eines Kalenderjahres enthalten, sondern das 4. Quartal eines Erhebungsjahrs sowie das 1., 2. und 3. Quartal des Folgejahres. Diese Unterteilung ist deshalb sinnvoll, weil sich die Teilnehmer an den Folgebefragungen im 1. bis 3. Quartal des Folgejahres aus den antwortenden Betrieben an der Hauptbefragung im 4. Quartal des Ursprungsjahres rekrutieren. So entsteht in jeder Befragungswelle ein unbalanciertes Panel mit vier Beobachtungszeitpunkten.

Bis auf wenige Ausnahmen (Tabelle 4) enthalten alle Variablennamen jeweils ein Präfix aus der Nummer des jeweiligen Quartals sowie der Jahreszahl minus 2000. Die Variable mit dem Betriebsgewicht `greggew` (vgl. Abschnitt 3.2) heißt im 3. Quartal 2012 demzufolge `q312_greggew`. Tabelle 5 zeigt die Aufteilung der Quartale auf die Datendateien sowie die jeweils zugehörigen Präfixe der Variablennamen. Die Bedeutung der einzelnen Variablennamen und deren Ausprägungen sind der Variablen-tabelle sowie den Codeplänen der jeweiligen

Befragungswelle auf der Webseite des Forschungsdatenzentrums unter „[Betriebsdaten: IAB-Stellenerhebung: Arbeitshilfen](#)“ zu entnehmen. Aus Platzgründen wird in den Codeplänen auf das jeweilige Präfix verzichtet.

Tabelle 4: Technische Variablen, die kein Variablennamens-Präfix verwenden

Variablenname	Bedeutung
key	Pseudonymisierter Betriebsidentifikator
jahr	Befragungswelle
hf	Hauptfragebogen wurde beantwortet
zf	Zusatzfragebogen wurde beantwortet
sf	Sonderfragebogen wurde beantwortet
fragebogen	Kombination aus hf, zf und sf
gebiet04	Region (West-/Ostdeutschland mit Berlin)
bl, bl_agg	Bundesland (NUTS 1), Bundesland (aggregiert)
wz03_16/wz03_28	2000–2009: 16/28 Wirtschaftszweige auf Basis der Klassifikation 2003
wz08_23	seit 2010: 23 Wirtschaftszweige auf Basis der Klassifikation 2008
wz*_3st/wz*_5st	Original-Wirtschaftszweigklassifikationen aus der BA-Adressdatei (3-/5-Steller) ¹⁾
kreis	seit 2007: Kreisschlüssel (NUTS 3) ¹⁾
rbezirk	seit 2005: Regierungsbezirk (NUTS 2) ¹⁾
gesamtausfueller	seit 2013: Es antwortete die übergeordnete Einheit, nicht der angeschriebene Betrieb
erh_form	Erhebungsform (schriftliche oder mündliche Befragung)
antwort	seit 2011: Antwortart (in Papierform oder online)

¹⁾ Sensibles Merkmal: Wird externen Nutzern nur auf Antrag zur Verfügung gestellt

Tabelle 5: Dateinamen und Variablennamens-Präfixe

Dateiname	Quartal	Variablennamens-Präfix
iabse_00	4. Quartal 2000	q40_
iabse_01	4. Quartal 2001	q41_
iabse_02	4. Quartal 2002	q42_
iabse_03	4. Quartal 2003	q43_
iabse_04	4. Quartal 2004	q44_
iabse_05	4. Quartal 2005 (Schriftliche Haupterhebung)	q45_
	1. Quartal 2006 (Telefonische Folgebefragung)	q16_
	2. Quartal 2006 (Telefonische Folgebefragung)	q26_
	3. Quartal 2006 (Telefonische Folgebefragung)	q36_
(...)		
iabse_15	4. Quartal 2015 (Schriftliche Haupterhebung)	q415_
	1. Quartal 2016 (Telefonische Folgebefragung)	q116_
	2. Quartal 2016 (Telefonische Folgebefragung)	q216_
	3. Quartal 2016 (Telefonische Folgebefragung)	q316_

3.2 Verwendung der Gewichte

Bei deskriptiven Auswertungen sind die entsprechenden Hochrechnungsgewichte zu verwenden, um von der realisierten Stichprobe auf die zugrundeliegende Grundgesamtheit schließen

zu können.¹ Bei multivariaten Auswertungen wird üblicherweise auf die Verwendung von Hochrechnungsgewichten verzichtet, wenn sämtliche Schichtungsvariablen (vgl. Tabelle 2) in die Schätzgleichung mit einfließen (vgl. Winship/Radbill 1994).

Für Variablen des Hauptfragebogens sowie des Sonderfragebogens, soweit sie sich nicht auf den zuletzt eingestellten Ein-Euro-Jobber beziehen, ist einfach das **Betriebsgewicht** `greggew` zu verwenden. Für Fragen des Zusatzfragebogens zur letzten Neueinstellung ist das **Neueinstellungsgewicht** (Betriebsgewicht multipliziert mit der Zahl der Neueinstellungen, seit 2004 Zahl der sozialversicherungspflichtigen Neueinstellungen), für Fragen zum letzten Suchabbruch das **Abbruchgewicht** (Betriebsgewicht multipliziert mit der Zahl der Suchabbrüche) zu verwenden. Tabelle 6 fasst die zu verwendenden Gewichte zusammen. Die telefonischen Nachbefragungen in den Folgequartalen verwenden für jedes Quartal ein eigenes Betriebsgewicht `greggew`, welches durch das quartalspezifische Präfix vom Betriebsgewicht der Hauptbefragung unterschieden wird.

Bei diesen Gewichten handelt es sich dem Grunde nach um Stichprobengewichte (sampling weights), das heißt, in STATA ist nach Möglichkeit die Gewichtsart `pweights` anzugeben und solche Auswertungsbefehle zu verwenden, die die Angabe dieser Gewichtsart unterstützen (`svyset`, `svy: tabulate`, `regress`). Für die Befehle `tabulate` (ohne `svy:`) und `summarize` kann hilfsweise auch `iweight` verwendet werden; allerdings können dann keine Standardfehler, etwa bei χ^2 -Tests, interpretiert werden. Soll bei deskriptiven Auswertungen auf statistische Signifikanz zwischen Teilgruppen getestet werden, so darf dies wegen fehlender Gewichtungsmöglichkeit nicht mit dem `ttest`-Befehl geschehen; stattdessen ist hilfsweise eine univariate Regression mit der Gruppenvariable als Regressor unter Angabe der `pweights`-Gewichtung durchzuführen und die Signifikanz des Unterschiedes am P-Wert der Gruppenvariable abzulesen (vgl. STATA-Beispiel 1).

Tabelle 6: Zu verwendendes Hochrechnungsgewicht nach Fragebogenteil und Welle

Fragebogenteil	Zu verwendendes Gewicht
Hauptfragebogen, telefonische Folgebefragungen	<code>greggew</code>
Der letzte Fall einer Einstellung	2000-2003: <code>greggew*f10</code> ab 2004: <code>greggew*f11_04</code>
Abgebrochene Personalsuche	<code>greggew*f422</code>
Sonderfragebogen (außer Fragen zum letzten Ein-Euro-Jobber)	<code>greggew</code>
Angaben zur zuletzt in einem Ein-Euro-Job eingestellten Person	<code>greggew*p29</code>

3.3 Fragebogenteile mit möglichen Mehrfachantworten

An vielen Stellen im Fragebogen sind **Mehrfachantworten** möglich. Da a priori nicht klar ist, ob ein nicht angekreuztes Feld als „nein“ oder als fehlender Wert zu interpretieren ist, sind

¹ Bei Verwendung von Hochrechnungs- oder Gewichtungsfaktoren in deskriptiven Analysen sind auch immer analoge ungewichtete Ergebnisse auszuweisen. Die gewichteten und die dazugehörigen ungewichteten Ergebnistabellen müssen dabei immer direkt untereinander stehen, weil dies die Datenschutzprüfung erleichtert und somit beschleunigt.

angekreuzte Felder im Datensatz mit dem Wert 1, nicht angekreuzte Felder als fehlender Wert (.) codiert. In der Praxis wird bei Auswertungen üblicherweise so vorgegangen, sämtliche nicht angekreuzte Felder dann in 0 („nein“) umzucodieren, wenn mindestens ein Feld innerhalb der Itematterie angekreuzt wurde (vgl. STATA-Beispiel 2).

Bei einigen Fragen können Antwortende „Sonstiges“ ankreuzen und nähere Angaben zu Sonstigem in einem **freien Textfeld** machen. Der Inhalt dieser freien Textfelder wird zur Wahrung des Datenschutzes nicht an externe Forscher zur Verfügung dargestellt, da über die Formulierung Rückschlüsse auf einzelne Betriebe möglich sein können. Häufiger vorkommende Antworten werden jedoch kategorisiert in einer entsprechenden Variable mit der Endung *u* (Umbuchung) zur Verfügung gestellt, deren Ausprägungen den Wertelabels im (Test-) Datensatz zu entnehmen sind.

Sofern eine Antwort im freien Textfeld bereits als angebotene Antwortmöglichkeit im Fragebogen existiert, wird sie entsprechend umbucht. Bei Fragen mit Antwortmöglichkeit „Sonstiges“ existiert jede Antwortvariable deshalb zweimal, einmal mit dem ursprünglichen Inhalt und einmal inklusive dieser Umbuchungen. Die Variablen „**inklusive Umbuchungen**“ erhalten das Suffix *iu*. Wird beispielsweise beim letzten Fall einer Neueinstellung als beschrittener Suchweg nur „sonstiger Suchweg“ angekreuzt und im freien Textfeld die Antwort „wurde uns empfohlen“ eingetragen, so enthalten die Variablen die in Tabelle 7 dargestellten Werte.

Tabelle 7: Beispiel für die Umbuchung einer freien Textfeld-Antwort

Variablenname	Bezeichnung	Wert
zf97	Suchweg: über eigene Mitarbeiter/persönliche Kontakte	. (wurde nicht angekreuzt)
zf9n	Sonstiger Suchweg	1 (wurde angekreuzt)
zf9txt*	Sonstiger Suchweg: freies Textfeld	„wurde uns empfohlen“
zf9u	Umbuchung in...	11 (persönliche Kontakte)
zf97iu	Suchweg: über eigene Mitarbeiter/persönliche Kontakte (inklusive Umbuchungen)	1 (wurde angekreuzt oder durch Umbuchungen generiert)
zf9niu	Sonstiger Suchweg	0 (wurde nicht angekreuzt oder wurde umbucht)

* nicht im FDZ-Datensatz enthalten

3.4 Klassifikationen von Wirtschaftszweigen und Berufen

Der **Wirtschaftszweig** eines Betriebes wird nicht abgefragt, sondern im Rahmen der Ziehung der Bruttostichprobe aus der Adressendatei der Bundesagentur für Arbeit entnommen. Wie Tabelle 2 zeigt, erfolgte dies bis 2003 anhand der Wirtschaftszweig-Klassifikation 1973 (WZ73), von 2004 bis 2009 anhand der Klassifikation WZ2003 und ab 2010 anhand der Klassifikation WZ2008. Im Datensatz wird hingegen von 2000 bis 2009 durchgängig die Klassifikation WZ2003 sowohl mit 16 als auch 28 Zweigen zur Verfügung gestellt (die Welle 2000 bis 2003 entsprechend umcodiert), ab 2010 die Klassifikation WZ2008. Die in der Adressendatei der Bundesagentur für Arbeit vorliegenden Original-5-Steller der Klassifikationen 2003 und 2008 werden als sensible Merkmale zur Wahrung des Datenschutzes externen Nutzern nur auf Antrag Verfügung gestellt.

An mehreren Stellen wird im Fragebogen die Angabe von Berufen verlangt. Da die aufwändige **Berufscodierung** den Antwortenden nicht zuzumuten ist, wird stattdessen nach der Berufsbezeichnung im Klartext gefragt. Diese Freitextantwort wird zur Wahrung des Datenschutzes nicht zur Verfügung gestellt. Die Berufsbezeichnungen werden vom Befragungsinstitut in die entsprechenden Codes zweier Klassifikationen codiert, wie Tabelle 8 zeigt. In einigen Fällen waren die Angaben zu ungenau, um eine passende Codierung zu finden; hier mussten außerhalb der offiziellen Codierung eigene Codes vergeben werden, die in Tabelle 9 aufgelistet sind.

Tabelle 8: Berufsklassifikationen nach Befragungswelle

Welle	Berufscodierung	Variablen
<u>Klassifikation 1 (deutsches System)</u>		
2000-2011	Klassifikation der Berufe, Ausgabe 1992 (KldB92), 3-Steller	c[1-5], bkz[1-3], f418_c[1-5], f423_c[1-5], f424c_[1-5], code, codeab, c430-c432
seit 2012	Klassifikation der Berufe, Ausgabe 2010 (KldB10), 5-Steller	kb10_*
<u>Klassifikation 2 (internationales System)</u>		
2000-2004	keine	
2005-2011	International Standard Classification of Occupations, 1988 Edition (ISCO-88)	isco*, iscl-3, f418isc*, f423isc*, f424isc*, i430-i432
seit 2012	International Standard Classification of Occupations, 2008 Edition (ISCO-08)	isco08_*

Tabelle 9: Ergänzung der KldB2010 für Angaben ohne nähere Tätigkeitsbeschreibung (o.n.T.)

Kategorie	Code	Bezeichnung
Technische Mitarbeiter o.n.T.	99001	Produktionsarbeitskraft, Produktionshelfer, Gehilfe
	99002	Werksfachkraft, technische Fachkraft, Geselle
	99003	Techniker, Vorarbeiter, Meister
	99004	Ingenieur, Diplomingenieur
	99005	Technische Mitarbeiter o.n.A.
Kaufmännische Mitarbeiter o.n.T.	99101	Kaufm. Gehilfen
	99102	Kaufmann, Disponent
	99103	Betriebswirt
	99104	Diplomkaufmann
	99105	Kaufmännische Mitarbeiter o.n.A.
Dienstleistungskräfte o.n.T.	99201	Servicepersonal, Supporter, Terminleger
	99202	Servicefachkraft, Berater, Inspektor, nichtwissenschaftl. Mitarbeiter
	99203	Fachberater, Personaldienstleister
	99204	Senior Consultant
	99205	Dienstleistungskräfte o.n.A.
Führungskräfte o.n.T.	99301	(Anforderungsniveau 1 kommt bei Führungskräften nicht vor)
	99302	(Anforderungsniveau 2 kommt bei Führungskräften nicht vor)
	99303	Teamleiter, Leitungspersonal, Standortchef, Abteilungsleiter
	99304	Direktor, Manager, Geschäftsführer
	99305	Führungskräfte o.n.A.
IT-Kräfte o.n.T.	99401	IT-Mitarbeiter (Anforderungsniveau 1)
	99402	IT-Fachkraft
	99403	IT-Spezialist
	99404	IT-Experte
	99405	IT-Beruf, IT-Mitarbeiter o.n.A.
Lehrkräfte o.n.T.	99501	Schulassistent
	99502	Lehrkraft
	99503	Lehrer
	99504	Dozent, Lehrkraft mit Hochschulabschluss
	99505	Lehrer o.n.A.
Arbeitskräfte in Ausbildung o.n.T.	99611	Azubi
	99621	Praktikant / Volontär
	99691	Sonstige Arbeitskräfte in Ausbildung
Arbeitskräfte o.n.T.	99901	Helfer, Aushilfe, Anlernkraft, ungelernte Mitarbeiter
	99902	Fachkraft, Arbeiter, Assistent
	99903	Spezialist
	99904	Experte
	99905	Arbeitskräfte o.n.A.

3.5 BHP-Verlinkung

Ab dem 4. Quartal 2010 steht für alle verlinkbaren Betriebe ergänzend das komplette Variablenspektrum des Betriebs-Historik-Panels, Version BHP 7516 v1, zur Verfügung mit Ausnahme des Erweiterungsmoduls zu Insolvenzen. Die BHP-Daten sind gesondert zu bestellen, ebenso die [BHP-Zusatzmodule](#). Hierzu sei auf die [Datensatzbeschreibung samt Merkmalspektrum](#) auf den Internetseiten des FDZ verwiesen.

Tabelle 10: Dateinamen der zuspieldbaren BHP-Datensätze

Dateiname	Inhalt
iabse_0015_v1_bhp_7516_m06_1975_v1 ...	BHP-Kerndatensatz
iabse_0015_v1_bhp_7516_m06_2016_v1	
iabse_0015_v1_bhp_7516_m06_inflow_v1 iabse_0015_v1_bhp_7516_m06_outflow_v1	BHP-Erweiterung „Beschäftigtenströme“
iabse_0015_v1_bhp_7516_m06_entry_v1 iabse_0015_v1_bhp_7516_m06_exit_v1	BHP-Erweiterung „Betriebsdynamik“
iabse_0015_v1_bhp_7516_m06_wgen_v1	BHP-Erweiterung „generierte konsistente Wirtschaftszweige“

Bei der Verwendung ist folgendes zu beachten:

1. Die Verlinkung erfolgt ausschließlich mittels der (über alle Jahre eindeutige) Variable `key`. Die im Betriebs-Historik-Panel ursprünglich enthaltene Variable `betnr` ist im BHP-Erweiterungsdatensatz der IAB-Stellenerhebung nicht enthalten.
2. Für eine geringe Anzahl von Antwortern war eine Verlinkung nicht möglich, weil sie für mehr als eine Betriebsstätte geantwortet haben; sie sind im BHP-Datensatz nicht enthalten. Weiterhin sind bei einigen `keys` aus dem gleichen Grund die zugespielten BHP-Daten nicht sinnvoll interpretierbar; bei ihnen hat die Variable `gesamtausfueller` den Wert 1.
3. Die Datenschutzerklärung, die an alle zur Teilnahme aufgeforderten Betriebe versendet wird, ist seit dem 4. Quartal 2010 so abgefasst, dass mit der Teilnahme an der Befragung implizit die Zustimmung zur Datenverlinkung erteilt wird. Mit Ausnahme der in Nummer 2 genannten Betriebe ist somit für sämtliche Betriebe eine Verlinkung möglich.
4. Im BHP-Erweiterungsdatensatz bezeichnet die Variable `jahr` den Zeitpunkt, auf den sich die jeweiligen BHP-Variablen beziehen. Die Befragungswelle, in welcher der jeweilige Betrieb befragt wurde, ist anhand der zusätzlichen Variable `SE_jahr` zu erkennen.
5. Wie auch beim IAB-Betriebspanel ist vereinzelt mit großen Abweichungen etwa bei der Betriebsgröße zwischen Befragungs- und administrativen Daten zu rechnen, etwa wenn die Angaben sich auf eine größere oder kleinere Einheit (z. B. Unternehmen versus Betriebsstätte) beziehen als in der für die Adressziehung verwendete Betriebsdatei der Bundesagentur für Arbeit vermerkt ist.
6. Sofern die sensiblen Merkmale `kreis/ao_kreis` oder `wz08_5st/ieb_w08` bestellt wurden, ist zu beachten, dass sich die entsprechenden Werte im Befragungsdatensatz auf den Zeitpunkt der Stichprobenziehung (vgl. Tabelle 3) beziehen, also auf das Vorjahr der Feldphase. Die Kreisnummern in den Befragungsdaten sind zum Ziehungszeitpunkt gültig, während die Kreisnummern in den administrativen Daten den *aktuellen* Gebietsstand widerspiegeln.

3.6 STATA-Auswertungsbeispiele

3.6.1 Zeitreihe der offenen Stellen

```
/* Ausgeben einer Zeitreihe der offenen Stellen von 2000 bis zu aktuellen Rand,
mit Konfidenzintervallen */

capture log close
log using ${log}/05a_STATABeispiel1_de,text replace
set more off

// Schleife über jede Befragungswelle
forvalues Welle=2000/2015 {
local WelleKurz =stroofreal(`Welle'-2000,"%02.0f")
quietly use ${orig}/iabse_`WelleKurz',clear
quietly label language de

// Anzahl der Quartale in dieser Welle ermitteln
if `Welle' <2005 {
    local Quartale 4
}
else {
    local Quartale 4 1 2 3
}

// Schleife über jedes Quartal
foreach Quartal of local Quartale {
// Quartalspezifisches Variablenpräfix bilden
if `Quartal' ==4 {
    local Jahr =`Welle'
}
else {
    local Jahr =`Welle' +1
}
local JahrKurz =`Jahr' -2000
local Praefix q`Quartal' `JahrKurz'

quietly {
// Keine Angabe heißt keine offenen Stellen
replace `Praefix'_f20 =0 if `Praefix'_f20 ==.
replace `Praefix'_f40 =0 if `Praefix'_f40 ==.

// Summe aus sofort und später zu besetzenden Stellen
generate `Praefix'_Stellen =`Praefix'_f20 +`Praefix'_f40

// Hochgerechnete Zahl der offenen Stellen berechnen inkl. 95%-
// Konfidenzintervall durch Berechnung des gewichteten Mittel-
// werts multipliziert mit der hochgerechneten Anzahl Betriebe
mean `Praefix'_Stellen [pweight =`Praefix'_greggew]
matrix Ergebnis =r(table)
quietly summarize `Praefix'_greggew
local Betriebe =r(sum)
local b =Ergebnis[1,1] *`Betriebe'
local ll =Ergebnis[5,1] *`Betriebe'
local ul =Ergebnis[6,1] *`Betriebe'
} // quietly

// Ergebniszeile ausgeben
display as text "Offene Stellen im " as result `Quartal' as text ". Quartal " as result `Jahr' as text ": " _
display as result %7.0f `b' as text ", 95% Konfidenzintervall: [" as result %7.0f `ll' " ; " %7.0f `ul' as text "]"
}
}
log close
```

3.6.2 Auswertung einer Frage mit Mehrfachantworten

```
/* Anteil der schwierigen Neueinstellungen und Gründe auswerten */

capture log close
log using ${log}/05b_STATABeispiel2_de,text replace
set more off

use ${orig}/iabse_13,clear
label language de

/* Neueinstellungsgewicht bilden */
generate gregneu = q413_greggew*q413_f11_04

/* Gewicht für alle folgenden Auswertungen */
svyset [pweight=gregneu]

/* Anteil der Neueinstellungen mit Schwierigkeiten */
svy: tabulate q413_zf8

/* Falls ein Grund angegeben, alle nicht angekreuzten auf "nein" setzen,
sonst auf Missing belassen */
generate byte beantwortet_zf8grund = q413_zf8aiu==1 | q413_zf8bliu==1 | ///
q413_zf8cliu==1 | q413_zf8eiu==1 | q413_zf8niu==1
foreach var of varlist q413_zf8aiu q413_zf8bliu q413_zf8cliu q413_zf8eiu q413_zf8niu {
    replace `var'=0 if `var'==. & beantwortet_zf8grund==1
}

/* Sicherstellen, dass nur bei Schwierigkeiten ein Grund genannt wurde */
assert beantwortet_zf8grund == 0 if q413_zf8 != 1

/* Gründe nacheinander ausgeben */
foreach var of varlist q413_zf8aiu q413_zf8bliu q413_zf8cliu q413_zf8eiu q413_zf8niu {
    svy: tabulate `var' if beantwortet_zf8grund
}
log close
```

Literatur

Bundesagentur für Arbeit (2013): Betriebsnummernvergabe. <https://www3.arbeitsagentur.de/web/wcm/idc/groups/public/documents/webdatei/mdaw/mtax/~edisp/l6019022dst-bai391187.pdf> [14.12.2016]

Brenzel, Hanna; Czepek, Judith; Kiesel, Hans; Kriechel, Ben; Kubis, Alexander; Moczall, Andreas; Rebien, Martina; Röttger, Christof; Szameitat, Jörg; Warning, Anja; Weber, Enzo (2016): Revision der IAB-Stellenerhebung. Hintergründe, Methode und Ergebnisse. (IAB-Forschungsbericht, 04/2016), Nürnberg

Winship, Christopher; Radbill, Larry (1994): Sampling Weights and Regression Analysis. *Sociological Methods & Research* 23 (2): 230–257

Impressum

FDZ-Datenreport 5/2018

Herausgeber

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

Redaktion

Dana Müller, Dagmar Theune

Technische Herstellung

Dagmar Theune

Rechte

Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit

http://doku.iab.de/fdz/reporte/2018/DR_05-18.pdf

Internet

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

Rückfragen zum Inhalt an:

Dr. Alexander Kubis
Institut für Arbeitsmarkt- und
Berufsforschung (IAB)
Regensburger Str. 104
90478 Nürnberg
Telefon: 0911 / 179-8978
E-Mail: Alexander.Kubis@iab.de