



FORSCHUNGSDATENZENTRUM
der Bundesagentur für Arbeit im Institut für
Arbeitsmarkt- und Berufsforschung

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

08|2022 DE IAB-Stellenerhebung: Betriebsbefragung zu
Stellenangebot und Besetzungsprozessen, Welle 2000 bis
2019 mit Folgequartalen von 2006 bis 2021

Mario Bossler, Nicole Gürtzgen, Alexander Kubis, Benjamin Kufner, Martin Popp



Bundesagentur für Arbeit

IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Welle 2000 bis 2019 mit Folgequartalen von 2006 bis 2021

Mario Bossler (IAB), Nicole Gürtzgen (IAB), Alexander Kubis (IAB), Benjamin Kufner (IAB),
Martin Popp (IAB)

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	Beschriftung des Datensatzes.....	5
2.1	Kurzbeschreibung.....	5
2.2	Überblick.....	6
2.3	Grundgesamtheit und Stichprobenziehung der Hauptbefragung.....	8
2.4	Hochrechnungsverfahren der Hauptbefragung	9
2.5	Stichprobenziehung und Hochrechnung der Quartalsbefragung	10
3	Wellenspezifische Besonderheiten	10
3.1	Befragungswelle 2019.....	10
3.1.1	Zusätzliche Quartalsbefragungen	10
3.1.2	Experimental-Daten der Jahreswelle 2019/20.....	11
3.1.3	Revision aufgrund von abweichendem Verhalten von Interviewerinnen bzw. Interviewern	12
3.2	Befragungswelle 2016.....	12
3.2.1	Vignettenstudie zur Wirkung der Mindestlohnhöhe	12
4	Verwendungshinweis für Datennutzerinnen und Datennutzer	13
4.1	Struktur des Datensatzes.....	13
4.2	Verwendung der Gewichte.....	14
4.3	Fragebogenteile mit möglichen Mehrfachantworten.....	15
4.4	Klassifikationen von Wirtschaftszweigen und Berufen	16
4.5	BHP-Verlinkung.....	18
4.6	Stata-Auswertungsbeispiel.....	21
4.6.1	Zeitreihe der offenen Stellen	21
4.6.2	Auswertung einer Frage mit Mehrfachantworten	22

Tabellenverzeichnis

Tabelle 1: Inhaltliche Charakteristika	5
Tabelle 2: Methodische Charakteristika	6
Tabelle 3 : Datenzugang	6
Tabelle 4: Schichtungsmerkmale der Bruttostichprobe 2000 - 2019.....	8
Tabelle 5: Stichprobenentwicklung 2000 - 2019.....	9
Tabelle 6: Technische-Variablen, die kein Variablennamens-Präfix verwenden	13
Tabelle 7: Dateinamen und Variablennamens-Präfixe	14
Tabelle 8: Zu verwendendes Hochrechnungsgewicht nach Modulen und Welle.....	15
Tabelle 9: Beispiel für die Umbuchung einer freien Textfeld-Antwort	16
Tabelle 10: Berufsklassifikationen nach Befragungswelle.....	16
Tabelle 11: Ergänzung der KIdB2010 für Angaben ohne nähere Tätigkeitsbeschreibung (o.n.T)	17
Tabelle 12: Dateinamen der zuspieldbaren BHP-Datensätze	18

Zusammenfassung

Die IAB-Stellenerhebung ist eine quartalsweise durchgeführte und repräsentative Betriebsbefragung über das gesamtwirtschaftliche Stellenangebot sowie Einstellungsprozesse in Deutschland. Die Erhebung ermittelt die Gesamtzahl aller offenen Stellen am Arbeitsmarkt, einschließlich jener Stellen, die nicht der Bundesagentur für Arbeit (BA) gemeldet werden. Das erste Modul des Fragebogens enthält Informationen zu Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. Das zweite Modul erfragt die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente sowie den betrieblichen Umgang mit am Arbeitsmarkt benachteiligten Personen. Das dritte Modul enthält Fragen zum letzten Fall einer Neueinstellung und zum letzten Fall eines gescheiterten Rekrutierungsversuchs. Das Forschungsdatenzentrum der Bundesagentur für Arbeit stellt die Datensätze der Befragungswellen ab 2000 mit allen Modulen einschließlich der Quartalsbefragungen für externe Wissenschaftlerinnen und Wissenschaftler bereit.

Abstract

The IAB Job Vacancy Survey is a quarterly and representative establishment survey on labour demand and recruitment processes in Germany. The survey identifies the overall stock of vacancies in the German labour market, including those vacancies that are not reported to the Federal Employment Agency (FEA). The first module of the questionnaire collects information about the number and structure of vacancies, future personnel requirements, about the current economic situation and the expected development of participating establishments. The second module enquires employer attitudes and firm use of current labour market instruments as well as the employer handling of persons disadvantaged in the labor market. The third module questionnaire asks for information about the last new hiring and the last case of a failed recruitment effort. The Research Data Centre of the Federal Employment Agency offers the data sets of the survey waves from 2000 onwards.

Keywords

Betriebsbefragung, offene Stellen, Stellenbesetzungsprozesse

1 Änderungen am Datensatz gegenüber früheren Versionen

Im Vergleich zu früheren Versionen des Datensatzes wurde der Zeitraum der bereitgestellten Daten um das 4. Quartal 2019 und die anschließenden Folgequartale erweitert (vgl. Bossler et al., 2021). Im 4. Quartal 2019 wurden im Rahmen des Moduls 2 Daten zu Beschäftigungschancen arbeitsloser Personen erhoben (siehe auch Gürtzgen/Popp, 2022). Ab dem 2. Quartal 2020 wurden die Betriebe zum Umgang mit der COVID-19-Krise befragt (siehe auch Bossler et al., 2020a, Gürtzgen/Kubis/Küfner 2020, Gürtzgen/Kubis, 2021).

Die Befragungswelle 2019 weist drei Besonderheiten auf, die bei der Nutzung der Daten zu beachten sind. Diese werden im Abschnitt 3 detailliert erläutert. Erstens wurde die Welle 2019 anlässlich des gesteigerten Informationsbedarfs während der COVID-19-Krise um zwei zusätzliche Quartalsbefragungen im 4. Quartal 2020 und im 1. Quartal 2021 verlängert und den Nutzerinnen und Nutzern im FDZ zur Verfügung gestellt. Zweitens fand parallel zur regulären Befragung im 4. Quartal 2019 ein Survey-Experiment statt. Die daraus resultierenden Daten sind Teil des Datensatzes und sollten ausschließlich für explorative und ungewichtete Zwecke genutzt werden. Drittens wurde im Rahmen von Qualitätssicherungsmaßnahmen in dieser Befragungswelle für einige Beobachtungen ein abweichendes Verhalten von Interviewerinnen bzw. Interviewern festgestellt (Bossler et al., 2022). Die nun im FDZ bereitgestellten Daten wurden im Rahmen der Revision um diese Beobachtungen bereinigt.

Außerdem wurden Inkonsistenzen in früheren Wellen bezüglich der Handhabung von Fragen mit Mehrfachantworten korrigiert (siehe auch 4.3). Alle Antwortmöglichkeiten, die dabei nicht von den Befragten angekreuzt wurden, werden nun einheitlich in den Datensätzen als „Missing“ ausgewiesen.

2 Beschriftung des Datensatzes

2.1 Kurzbeschreibung

Tabelle 1: Inhaltliche Charakteristika

Kategorie	Beschreibung
Themen/Merkmalgruppen	Beschäftigung, Personalbedarf, Aktivitätshemmnisse, abgebrochene Personalsuche, Berufe, für die steigender Bedarf erwartet wird und mit erwarteten Engpässen in den kommenden 3 Jahren, geförderte Beschäftigung, geringfügige Beschäftigung, Teilzeitbeschäftigung, Reform des Arbeitsmarktes, Beschäftigungschancen (langzeit-)arbeitsloser Personen, Weiterbildung der Belegschaft, Ein-Euro-Jobs (Erfahrungen der Betriebe), Arbeitsmarktpartizipation älterer Beschäftigter, arbeitsrelevante Erfahrungen mit Geflüchteten, Umgang mit der COVID-19 Krise, detaillierte Informationen zur letzten Einstellung, detaillierte Informationen zur letzten abgebrochenen Personalsuche

Untersuchungseinheit	Betriebe
Fallzahlen	7.500 bis 15.000 Betriebe
Zeitraum	IV.2000 bis I.2021
Zeitbezug	quartalsweise
Regionale Gliederung	Ost-/Westdeutschland, Bundesländer, weitere Merkmale siehe unter "sensible Merkmale"

Tabelle 2: Methodische Charakteristika

Kategorie	Beschreibung
Erhebungsdesign	Repräsentative Stichprobe der Betriebe, geschichtet nach Betriebsgröße, Branche und West-/Ostdeutschland
An der Erhebung beteiligte Institutionen	Auftraggeber: Forschungsbereich AMPI des IAB Durchführung: Economix Research & Consulting, München
Frequenz der Datensammlung	Jährliche schriftliche bzw. online und quartalsweise telefonische Befragung
Dateiformat/-größe	Stata; alle Datensätze zusammen ca. 264 MB
Dateiorganisation	nach Welle

Tabelle 3 : Datenzugang

Kategorie	Beschreibung
Datenzugang	Datenfernverarbeitung, Gastaufenthalt
Anonymisierungsgrad	Schwach anonymisiert
sensible Merkmal	Wirtschaftszweigklassifikationen: WZ73 3-Steller (2000-2003), WZ03 3-Steller (2004)/5-Steller (2005-2009) WZ08 3-Steller/5-Steller (ab 2010), Kreisnummer (ab 2007), Regierungsbezirk (ab 2005)
Vorgaben zum Zitieren der Daten und Datendokumentation	<p>Daten: "Gürtzgen, Nicole; Kubis, Alexander; Bossler, Mario; Kufner, Benjamin; Popp, Martin (2022): "IAB-Stellenerhebung – Version 0019 v1". Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB). DOI: 10.5164/IAB.IABSE0019.de.en.v1 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.“</p> <p>Datendokumentation: - Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin; Popp, Martin (2022): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Wellen 2000 bis 2019 mit Folgequartalen von 2006 bis 2021. FDZ-Datenreport, 08/2022 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2208.de.v1 - Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin; Popp, Martin (2022): The IAB Job Vacancy Survey: Establishment survey on labour demand and recruitment processes, waves 2000 to 2019 and subsequent quarters 2006 to 2021. FDZ-Datenreport, 08/2022 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2208.en.v1</p>

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

2.2 Überblick

Die IAB-Stellenerhebung ist eine Betriebsbefragung, die repräsentativ das gesamtwirtschaftliche Stellenangebot in Deutschland abbildet (siehe auch Bossler et al., 2020b). 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 es, valide und unverzerrte Aussagen über die gesamtwirtschaftliche Arbeitsnachfrage zu treffen. 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 (mit Unterbrechungen in den Jahren 2005 -2010) können Betriebe an der Befragung auch online über das Internet teilnehmen. Dieser Webfragebogen ist mit dem schriftlichen Fragebogen identisch. Das Modul 1, auch Hauptfragebogen genannt, enthält Informationen zur Zahl und Struktur offener Stellen, dem erwarteten künftigen Arbeitskräftebedarf, zur wirtschaftlichen Lage und zur Entwicklung der befragten Betriebe. In Modul 2, auch Sonderfragebogen genannt, wird seit 2000 im SGBII-Kontext die betriebliche Einschätzung und Nutzung aktueller arbeitsmarktpolitischer Instrumente erfragt. Zudem werden Daten zum Umgang der Betriebe mit am Arbeitsmarkt benachteiligten Personen erhoben. Das Modul 3, auch Zusatzfragebogen genannt, enthält Fragen zum letzten Fall einer Neueinstellung sowie 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, Zahlen zu Bewerberinnen und Bewerbern, Besetzungsschwierigkeiten sowie über gemachte Kompromisse. Das Modul 3 wird nur von Betrieben beantwortet, die im letzten Jahr Mitarbeiterinnen und Mitarbeiter neu eingestellt oder dies zumindest versucht haben.

Module 1 und Modul 3 sind von Welle zu Welle im Wesentlichen gleich, abgesehen von redaktionellen Änderungen und der Verlagerung von Schwerpunkten. Modul 2 ändert sich hingegen mit jeder Befragungswelle. Von 2005 bis 2013 thematisierte das Modul hauptsächlich mit der betrieblichen Nutzung und Einschätzung von Ein-Euro-Jobs. Es 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 teilnehmenden Betrieben der Hauptbefragung statt. Diese telefonischen Quartalsbefragungen ergänzen die schriftliche Hauptbefragung im jeweils 4. Quartal. Erfragt und aktualisiert werden die wichtigsten Eckdaten aus dem ersten Modul. Hierunter fällt u.a. die Zahl der Beschäftigten eines Betriebes, seine Einschätzung der Beschäftigungsentwicklung und die Zahl der offenen Stellen. In der Befragungswelle 2019 fanden zusätzliche Quartalsbefragungen im 4. Quartal 2020 und im 1. Quartal 2021 statt.

Im Forschungsdatenzentrum der Bundesagentur für Arbeit werden die Befragungswellen ab 2000 mit allen Modulen 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.

2.3 Grundgesamtheit und Stichprobenziehung der Hauptbefragung

Grundgesamtheit der Befragung im 4. Quartal eines jeweiligen Jahres sind sämtliche Betriebe mit mindestens einem oder einer sozialversicherungspflichtig Beschäftigten im Juni (bis Welle 2004) bzw. Dezember (ab Welle 2005) des Vorjahres unter Ausschluss der Privathaushalte. Ein „Betrieb“ bezieht sich dabei auf eine wirtschaftliche Einheit nach dem Betriebsnummernkonzept des Betriebsnummern-Service der Bundesagentur für Arbeit (Bundesagentur für Arbeit, 2020).

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 4 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 4: Schichtungsmerkmale der Bruttostichprobe 2000 - 2019

Merkmal	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

Um von der Bruttostichprobe zur bereinigten Versandstichprobe zu gelangen, werden all jene Betriebe ausgeschlossen, die bei vergangenen Befragungen des IAB geäußert haben, nicht mehr zur Teilnahme aufgefordert werden zu wollen. Diese bildet die Basis für die in 5 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 der größte Teil des Rücklaufes zwischen Mitte Oktober und Anfang Dezember erfolgt.

Tabelle 5: Stichprobenentwicklung 2000 - 2019

Welle	Adressbestand Stichtag	Durchschn. Auswahlsatz	Brutto- stichprobe	Versand- stichprobe	Netto- stichprobe	Rücklaufquote
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 %
2016	31.12.2015	4,0 %	110.140	85.000	11.588	13,6 %
2017	31.12.2016	5,1 %	120.102	109.988	14.614	13,3 %
2018	31.12.2017	5,2 %	134.848	109.988	14.528	13,2 %
2019	31.12.2018	5,1 %	114.613	109.986	13.906	12,6 %

2.4 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 der jeweiligen Auswahlwahrscheinlichkeit.
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 (siehe hierzu Hutter 2020), wird die Kalibrierung nach jeweils 18 Monaten mit der tatsächlichen Anzahl der sozialversicherungspflichtigen

Beschäftigten wiederholt und die Hochrechnungsgewichte somit revidiert. Der FDZ-Datensatz enthält ausschließlich die in dieser Weise revidierten Hochrechnungsgewichte; 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 in der Regel innerhalb des jeweiligen 95%-Konfidenzintervalls.

Für die rückwirkende Anwendung des neuen Hochrechnungsverfahrens auf die Wellen 2000 bis 2013 wurden die sich über die Jahre verändernden Betriebsgrößenklassen der ursprünglichen Stichprobenmatrizen (vgl. Tabelle 4) 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. Von 2010 bis 2014 existieren bereits in der Stichprobenmatrix 23 Ausprägungen der Klassifikation der Wirtschaftszweige, Ausgabe 2008. Seit 2015 werden analog zur Stichprobenmatrix 24 Ausprägungen der Wirtschaftszweige, Ausgabe 2008, für die Hochrechnung verwendet.

2.5 Stichprobenziehung und Hochrechnung der Quartalsbefragung

Aus den teilnehmenden Betrieben an der schriftlichen Hauptbefragung im vierten Quartal werden jeweils zufällig Betriebe für eine kurze telefonische Nachbefragung in den drei Folgequartalen gezogen. Die angestrebte Zielgröße beträgt 9.000 Betriebe in jedem Quartal. Die Nettostichprobe der schriftlichen Hauptbefragung stellt damit die Bruttostichprobe für die telefonischen Nachbefragungen in den Folgequartalen dar.

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

3 Wellenspezifische Besonderheiten

3.1 Befragungswelle 2019

3.1.1 Zusätzliche Quartalsbefragungen

Die COVID-19-Krise hat den Bedarf an kurzfristiger und valider Information zum Stellenangebot sowie zum betrieblichen Management im Allgemeinen deutlich gesteigert. Um diesem Informationsbedürfnis nachzukommen, wurde die Befragungswelle 2019 um zwei zusätzliche Quartalsbefragungen im 4. Quartal 2020 und im 1. Quartal 2021 verlängert. In diesen beiden Quartalen wurde eine Teilstichprobe der befragten Betriebe aus dem 4. Quartal 2019, wie in den regulären Quartalsbefragungen, erneut befragt. Die Befragungswelle 2019 stellt somit eine Panelbefragung aus einer Messung vor Ausbruch der Pandemie im 4. Quartal 2019 sowie weiteren

Messungen im Verlauf der Pandemie dar. Beginnend mit dem 2. Quartal 2020 wurden alle Quartalsbefragungen um zusätzliche Fragen zum betrieblichen Umgang mit der COVID-19-Pandemie erweitert. Die Befragungswelle 2019, um die beiden zusätzlichen Quartalsbefragungen erweitert, wird auch als „COVID-Panel“ bezeichnet.

Zu beachten ist, dass für das COVID-Panel – analog zum regulären Gewichtungungsverfahren – eigenständige Hochrechnungsgewichte bereitgestellt werden (siehe Abschnitt 4.1). Die Hochrechnungsgewichte für das COVID-Panel sind spezifisch gelabelt. Die Gewichte haben das Suffix `_covid` und tragen dementsprechend die Variablennamen `q419_greggew_covid`, `q120_greggew_covid`, `q220_greggew_covid`, `q320_greggew_covid`, `q420_greggew_covid` bzw. `q121_greggew_covid`. Im 1., 2., und 3. Quartal 2020 entsprechen die Gewichte mit Suffix `_covid` den regulären Hochrechnungsgewichten ohne Suffix. Da es sich bei den Quartalsbefragungen im 4. Quartal 2020 und im 1. Quartal 2021 um außerordentliche Zusatzbefragungen handelt, stehen für diese beiden Quartale der Welle 2019 keine regulären Gewichte, sondern ausschließlich Gewichte mit Suffix `_covid` zur Verfügung.

Die vom IAB berichteten Auswertungen zu offenen Stellen im 4. Quartal 2020 und im 1. Quartal 2021 basieren auf der Befragungswelle 2020. Die zusätzlichen Quartalsbefragungen der Welle 2019 wurden hierfür nicht genutzt.

3.1.2 Experimental-Daten der Befragungswelle 2019

Im 4. Quartal 2019 fand neben der regulären Hauptbefragung ein Survey-Experiment (N=3,883) zu den Effekten von erklärenden Hinweisen auf das Antwortverhalten statt (für nähere Informationen zum Experimentaldesign und zu den Ergebnissen siehe Kufner/Sakshaug/Zins, 2021). Für das Experiment wurde eine separate Stichprobe gezogen und - anders als in der Hauptbefragung - eine reine Onlinebefragung durchgeführt. Die reguläre Hauptbefragung im 4. Quartal und alle nachfolgenden Quartalsbefragungen waren nicht Teil des Experiments und wurden vom Experiment nicht beeinträchtigt oder beeinflusst. Die Haupterhebung steht für sich und bildet die zentrale Grundlage für inhaltliche Analysen.

Die Daten dieser Experimentalbefragung sind Teil der Jahreswelle 2019, die an das FDZ geliefert werden. Diese Beobachtungen können über die Indikatorvariable `experiment2019` eindeutig identifiziert werden. Es ist zu beachten, dass für diese Experimental-Beobachtungen keine Hochrechnungsgewichte zur Verfügung stehen, weshalb eine Nutzung für gewichtete Analysen nicht möglich ist (`q419_greggew == .`). Mit Ausnahme der im Rahmen des Survey-Experiments durchgeführten Analysen, fand keine zusätzlichen Qualitätsprüfung der Experimentaldaten statt. Auch wurde nicht untersucht, inwiefern Abweichungen im Erhebungsdesign einen Einfluss auf mögliche Selektionseffekte oder Messeffekte und damit die Vergleichbarkeit mit der Haupterhebung haben. Eine Verlinkung mit administrativen Daten ist nicht möglich (siehe Abschnitt 4.3). Daher empfehlen wir standardmäßig den Ausschluss dieser Experimentaldaten (`drop if experiment2019 == 1`) für Analysen inhaltlicher Art. Dennoch kann die Nutzung der Experimentaldaten für explorative und ungewichtete Zwecke einen Mehrwert bringen und steht Nutzerinnen und Nutzern im FDZ daher zur Verfügung.

3.1.3 Revision aufgrund von abweichendem Verhalten von Interviewerinnen bzw. Interviewern

Im 2. Quartal in 2021 wurden bei routinemäßigen Prüfungen der Datenqualität auffällige Datenmuster und Zeitreihen identifiziert. Im Rahmen einer ausführlichen Untersuchung konnte festgestellt werden, dass durch eine Interviewerin bzw. einen Interviewer systematisch Interviews manipuliert wurden. Diese Person hat bei einer Filterfrage gehäuft angegeben, dass die befragten Betriebe über keine Vakanzen verfügen. Durch diese verfälschte Angabe mussten keine weiteren Fragen zur Struktur der Vakanzen erhoben werden, wodurch die Befragungsdauer stark verkürzt werden konnte. Im Laufe der Untersuchung konnte noch eine weitere Interviewerin beziehungsweise ein weiterer Interviewer identifiziert werden, die im 1. Quartal 2020 abweichendes Verhalten gezeigt hat. Alle Interviews der beiden Personen wurden auch retrospektiv aus der gesamten Befragungswelle 2019 ausgeschlossen. In den früheren Befragungswellen waren keine Interviews der beiden Personen enthalten.

Im Rahmen einer Revision wurden neue Hochrechnungsgewichte berechnet. Zudem wurden alle IAB-Publikationen sowie die auf der Homepage veröffentlichten Zahlen revidiert, die Beobachtungen der Interviewfälschung enthielten. In den nun im FDZ veröffentlichten Daten sind die Beobachtungen der beiden Interviewerinnen beziehungsweise Interviewer ausgeschlossen und die revidierten Hochrechnungsgewichte enthalten. Zuvor im FDZ veröffentlichte Daten sind von der Interviewfälschung nicht betroffen. Weitere Informationen zur Entdeckung des abweichenden Verhaltens der Interviewerinnen beziehungsweise Interviewer, der Revision der Daten sowie wie zur verbesserten Qualitätssicherung der Telefoninterviews sind im FDZ-Methodenreport von Bossler et al. (2022) festgehalten.

3.2 Befragungswelle 2016

3.2.1 Vignettenstudie zur Wirkung der Mindestlohnhöhe

Im 2. Quartal 2017 wurde die Befragung für eine Vignettenstudie genutzt. Dabei wurde die betriebliche Reaktion auf einen Anstieg oder ein Absinken des gesetzlichen Mindestlohns untersucht (siehe Bossler/Oberfichtner/Schnabel, 2020). Die beiden dazugehörigen Fragebogenvarianten sind in den jeweiligen Codeplänen dokumentiert. Die entsprechenden Daten sind Teil des Datensatzes und die über die Fragebögen variierende Höhe des genannten fiktiven Mindestlohns kann der Variable `q217_mindestlohn` entnommen werden.

4 Verwendungshinweis für Datennutzerinnen und Datennutzer

4.1 Struktur des Datensatzes

Die Daten jeder Befragungswelle sind in jeweils einer Datei im Stata-Format abgelegt. Sie enthält die Informationen des 1., 2. und 3. Moduls sowie ab 2005 die Angaben aus den telefonischen Folgebefragungen in den Folgequartalen. Die Daten werden also nicht separat für Kalenderjahre, sondern das 4. Quartal eines Erhebungsjahrs wird zusammen mit den – im Regelfall drei - Folgequartalen bereitgestellt. Diese Unterteilung ist deshalb sinnvoll, weil sich die teilnehmenden Betriebe an den Folgebefragungen aus den antwortenden Betrieben an der Hauptbefragung im 4. Quartal des Ursprungsjahrs rekrutieren. So entsteht in jeder Befragungswelle ein unbalanciertes Panel mit vier Beobachtungszeitpunkten bzw. sechs Beobachtungszeitpunkten in der Befragungswelle 2019.

Bis auf wenige Ausnahmen (Tabelle 6) 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 7 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 6: Technische-Variablen, die kein Variablennamens-Präfix verwenden

Variablenname	Bedeutung
key	Pseudonymisierter Betriebsidentifikator
jahr	Befragungswelle
hf	Modul 1 (Hauptfragebogen) wurde beantwortet
sf	Modul 2 (Sonderfragebogen) wurde beantwortet
zf	Modul 3 (Zusatzfragebogen) wurde beantwortet
fragebogen	Kombination aus hf, zf und sf
gebiet04	Region (West-/Ostdeutschland mit Berlin)
b1, b1_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
wz08_24	seit 2015: 24 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 Nutzerinnen und Nutzern nur auf Antrag zur Verfügung gestellt

Tabelle 7: Dateinamen und Variablennamen-Präfixe

Dateiname	Quartal	Variablennamen-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_18	4. Quartal 2018 (Schriftliche Haupterhebung)	q418_
	1. Quartal 2019 (Telefonische Folgebefragung)	q119_
	2. Quartal 2019 (Telefonische Folgebefragung)	q219_
	3. Quartal 2019 (Telefonische Folgebefragung)	q319_
iabse_19	4. Quartal 2019 (Schriftliche Haupterhebung)	q419_
	1. Quartal 2020 (Telefonische Folgebefragung)	q120_
	2. Quartal 2020 (Telefonische Folgebefragung)	q220_
	3. Quartal 2020 (Telefonische Folgebefragung)	q320_
	4. Quartal 2020 (Telefonische Folgebefragung – Zusatz)	q420_
	1. Quartal 2021 (Telefonische Folgebefragung – Zusatz)	q121_

4.2 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 Regressionsanalysen wird üblicherweise auf die Verwendung von Hochrechnungsgewichten verzichtet, wenn sämtliche Schichtungsvariablen (vgl. 4) in die Schätzgleichung mit einfließen (vgl. Bossler/Geis/Stegmaier, 2018; Winship/Radbill, 1994).

Für Variablen des Moduls 1 sowie des Moduls 2, soweit sie sich nicht auf den zuletzt eingestellten Ein-Euro-Jobber beziehen, ist das **Betriebsgewicht** `greggew` zu verwenden. Für Fragen des Moduls 2 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 8 fasst die zu verwendenden Gewichte zusammen. Für die telefonischen Nachbefragungen in den Folgequartalen steht für jedes Quartal ein eigenes Betriebsgewicht `greggew` zur Verfügung, welches durch das quartalspezifische Präfix vom Betriebsgewicht der Hauptbefragung unterschieden wird. Wie in

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.

Abschnitt 3.1.1. beschrieben, stehen für das COVID-Panel eigene Hochrechnungsgewichte zur Verfügung.

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 8: Zu verwendendes Hochrechnungsgewicht nach Modulen und Welle

Module	Zu verwendendes Gewicht
Modul 1, telefonische Folgebefragungen	greggew
Telefonische Folgebefragung – COVID-Panel	greggew_covid
Der letzte Fall einer Einstellung	2000-2003: ab 2004: greggew*f10 greggew*f11_04
Abgebrochene Personalsuche	greggew*f422
Modul 2 (außer Fragen zum letzten Ein-Euro-Jobber)	greggew
Angaben zur zuletzt in einem Ein-Euro-Job eingestellten Person	greggew*p29

4.3 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 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 Nutzerinnen und Nutzer zur Verfügung dargestellt, da über die Formulierung Rückschlüsse auf einzelne Betriebe möglich sein können. Häufiger vorkommende Antworten werden jedoch kategorisiert in einer entsprechenden Variablen 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 9 dargestellten Werte.

Tabelle 9: 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 umgebucht)

* nicht im FDZ-Datensatz enthalten

4.4 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 4 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 Wellen 2000 bis 2003 entsprechend umcodiert), ab 2010 die Klassifikation WZ2008. Die in der Adressendatei der Bundesagentur für Arbeit vorliegenden Original-3 und 5-Steller der Klassifikationen 1973, 1993, 2003 und 2008 werden als sensible Merkmale zur Wahrung des Datenschutzes externen Nutzerinnen und Nutzern nur auf Antrag zur 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 10 zeigt. In einigen Fällen waren die Angaben zum beruflichen Tätigkeitsfeld zu ungenau, um eine passende Codierung zu finden; hier mussten außerhalb der offiziellen Codierung eigene Codes vergeben werden, die in Tabelle 11 aufgelistet sind.

Tabelle 10: Berufsklassifikationen nach Befragungswelle

Welle	Berufscodierung	Variablen
Klassifikation 1 (deutsches System)		

Welle	Berufscodierung	Variablen
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 – Erste Fassung (KldB2010), 5-Steller	kb10_*
Klassifikation 2 (internationales System)		
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 11: 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.

4.5 BHP-Verlinkung

Ab dem 4. Quartal 2010 steht für alle verknüpfbaren Betriebe ergänzend das komplette Variablenspektrum des Betriebs-Historik-Panels, Version BHP 7520 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 12: Dateinamen der zuspielbaren BHP-Datensätze

Dateiname	Inhalt
iabse_0019_v1_bhp_7520_m06_v1_1975 ... iabse_0019_v1_bhp_7520_m06_v1_2019	BHP-Kerndatensatz
iabse_0019_v1_bhp_7520_m06_inflow_v1 iabse_0019_v1_bhp_7520_m06_outflow_1	BHP-Erweiterung „Beschäftigtenströme“
iabse_0019_v1_bhp_7520_m06_entry_v1 iabse_0019_v1_bhp_7520_m06_exit_v1	BHP-Erweiterung „Betriebsdynamik“
iabse_0019_v1_bhp_7520_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 eindeutigen) 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 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. Es ist vereinzelt mit 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_3st/wz08_5st/ieb_w08` bestellt wurden, ist zu beachten, dass sich die entsprechenden Werte im Befragungsdatensatz auf den Zeitpunkt der Stichprobenziehung (vgl. Tabelle 5) beziehen, also auf das Vorjahr der Feldphase. Die Kreisnummern in den Befragungsdaten sind zum

Ziehungszeitpunkt gültig, während die Kreisnummern in den administrativen Daten den Gebietsstand zum 31.12.2020 widerspiegeln.

Die Angaben aus den BHP-Daten können von den Daten aus der Befragung zur IAB-Stellenerhebung abweichen. Die Erforderlichkeit des BHPs bzw. jedes sensiblen Merkmales muss bei der Beantragung gesondert begründet werden. Es werden alle unten aufgeführten Merkmale bereitgestellt.

Basismerkmale

- Ordnungsnummer (key)
- Welle der IAB-Stellenerhebung (SE_jahr)
- Jahr (jahr)
- Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)
- Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)
- Bundesland des Arbeitsorts (ao_bula)
- Anzahl der Beschäftigten insgesamt (az_ges)
- Anzahl der Vollzeitbeschäftigten (az_ges_vz)
- Anzahl der geringfügig Beschäftigten (az_gf)
- Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Thematische Variablenblöcke

- Struktur der Beschäftigten allgemein (az_ges az_f az_reg az_gf az_azubi az_atz az_vz az_tz az_f_vz az_f_tz az_reg_vz)
- Struktur der Beschäftigten nach Schul- und Berufsbildung (az_gq az_mq az_hq az_gq_vz az_mq_vz az_hq_vz)
- Altersstruktur der Beschäftigten (az_15_19 az_15_19_vz az_20_24 az_20_24_vz az_25_29 az_25_29_vz az_30_34 az_30_34_vz az_35_39 az_35_39_vz az_40_44 az_40_44_vz az_45_49 az_45_49_vz az_50_54 az_50_54_vz az_55_59 az_55_59_vz az_60_64 az_60_64_vz az_ab65 az_ab65_vz alter_mw alter_mw_vz)
- Aktivitäten Forschung und Entwicklung (az_ingnat)
- Anzahl atypisch Beschäftigter (az_leih az_bfr)
- Struktur der Beschäftigten nach Staatsangehörigkeit (az_d az_d_vz az_eu)
- Struktur der Beschäftigten nach Blossfeld Berufsqualifikation (az_bfAgr az_bfEmb az_bfEdi az_bfEVB az_bfQMB az_bfQDI az_bfQVB az_bfTec az_bfSemi az_bfIng az_bfProf az_bfMan)
- Struktur der Beschäftigten nach Tätigkeitsniveau (az_niv1 az_niv2 az_niv3 az_niv4)
- Gehaltsstruktur der Vollzeitbeschäftigten (az_zens te_med te_p25 te_p75 te_imp_mw te_imp_med te_imp_p25 te_imp_p75 te_imp_mw_f te_imp_med_f te_imp_med_m te_imp_mw_gq te_imp_med_gq te_imp_mw_mq te_imp_med_mq te_imp_mw_hq te_imp_med_hq te_imp_med_uq te_imp_mw_d te_imp_med_d te_imp_med_a)
- Erweiterungsmodul – Beschäftigtenströme (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_vz_f ein_tz_f ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_15_19 ein_20_24 ein_25_29 ein_30_34 ein_35_39 ein_40_44 ein_45_49 ein_50_54 ein_55_59 ein_60_64 ein_ab65 aus_ges aus_f aus_reg aus_gf aus_azubi aus_vz aus_tz

aus_vz_f aus_tz_f aus_reg_vz aus_temp aus_bw aus_bw_f aus_15_19 aus_20_24
aus_25_29 aus_30_34 aus_35_39 aus_40_44 aus_45_49 aus_50_54 aus_55_59 aus_60_64
aus_ab65 aus_senio_1 aus_senio_2 aus_senio_3)

- Erweiterungsmodul – Betriebsdynamik (eintritt besch inflow betnr_vor besch_vor status_vor austritt besch outflow betnr_nach besch_nach status_nach)

Sensible Merkmale BHP

Zusätzlich können folgende sensible Merkmale aus dem BHP einzeln beantragt werden:

- Kreis des Arbeitsorts (ao_kreis)
- Klassifikation der Wirtschaftszweige 1973 – 3-Steller (w73_3)
- Klassifikation der Wirtschaftszweige 1993 – 3-Steller (w93_3)
- Klassifikation der Wirtschaftszweige 1993 – 5-Steller (w93_5)
- Klassifikation der Wirtschaftszweige 2003 – 3-Steller (w03_3)
- Klassifikation der Wirtschaftszweige 2003 – 5-Steller (w03_5)
- Klassifikation der Wirtschaftszweige 2008 – 3-Steller (w08_3)
- Klassifikation der Wirtschaftszweige 2008 – 5-Steller (w08_5)
- Zeitkonsistente Klassifikationen der Wirtschaftszweige, 3-Steller (w73_3_gen, w93_3_gen w08_3_gen)

4.6 Stata-Auswertungsbeispiel

4.6.1 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=2000/2019 {
    local WelleKurz =stroofreal(`Welle'-2000,"%02.0f")
    quietly use ${orig}/iabse_`WelleKurz',clear
    quietly label language de

    /*Ausschluss der Experimentaldaten */
    if `Welle' ==2019 {
        drop if experiment2019 ==1 & jahr == 2019
    }

    // 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 {
        // Quartalsspezifisches 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
```

4.6.2 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_19,clear

/*Ausschluss der Experimentaldaten */
drop if experiment2019 ==1

label language de

/* Neueinstellungsgewicht bilden */
generate gregneu = q419_greggew*q419_f11_04

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

/* Anteil der Neueinstellungen mit Schwierigkeiten */
svy: tabulate q419_zf8

/* Falls ein Grund angegeben, alle nicht angekreuzten auf "nein" setzen,
sonst auf Missing belassen */
generate byte beantwortet_zf8grund = q419_zf8aiu==1 | q419_zf8bliu==1 | ///
q419_zf8cliu==1 | q419_zf8eiu==1 | q419_zf8niu==1
foreach var of varlist q419_zf8aiu q419_zf8bliu q419_zf8cliu q419_zf8eiu q419_zf8niu
{
    replace `var'=0 if `var'==. & beantwortet_zf8grund==1
}

/* Sicherstellen, dass nur bei Schwierigkeiten ein Grund genannt wurde */
assert beantwortet_zf8grund == 0 if q419_zf8 != 1

/* Gründe nacheinander ausgeben */
foreach var of varlist q419_zf8aiu q419_zf8bliu q419_zf8cliu q419_zf8eiu q419_zf8niu
{
    svy: tabulate `var' if beantwortet_zf8grund
}
log close
```

Literatur

- Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin; Olbrich, Lukas; Schwanhäuser, Silvia (2022): Revision and new data quality concept due to deviant interviewer behavior in the IAB Job Vacancy Survey. (FDZ-Methodenreport, 05/2022), Nürnberg, 19 S.
- Bossler, Mario; Kubis, Alexander; Kufner, Benjamin; Popp, Martin (2021): IAB-Stellenerhebung: Betriebsbefragung zu Stellenangebot und Besetzungsprozessen, Welle 2000 bis 2018 mit Folgequartalen von 2006 bis 2019. (FDZ-Datenreport, 09/2021 (de)), Nürnberg, 19 S.
- Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin (2020a): IAB-Stellenerhebung im ersten Quartal 2020: Mit dem Corona-Shutdown ging zuerst die Zahl der offenen Stellen zurück. (IAB-Kurzbericht, 12/2020), Nürnberg, 12 S.
- Bossler, Mario; Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin; Lochner, Benjamin (2020b). The IAB Job Vacancy Survey: design and research potential. Journal for Labour Market Research, Jg. 54, H. 1, S. 1-12.
- Bossler, Mario; Oberfichtner, Michael; Schnabel, Claus (2020): Employment adjustments following rises and reductions in minimum wages: New insights from a survey experiment. In: Labour, Vol. 34, No. 3, S. 323-346.
- Bossler, Mario; Geis, Gregor; Stegmaier, Jens (2018): [Comparing survey data with an official administrative population: Assessing sample-selectivity in the IAB Establishment Panel](#). In: Quality & Quantity, Jg. 52, H. 2, S. 899-920.
- Bundesagentur für Arbeit (2020): Steckbrief Betriebsnummervergabe. <https://www.informationsportal.de/wp-content/uploads/Steckbrief-Betriebsnummer.pdf> [2021-06-08]
- 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
- Gürtzgen, Nicole; Popp, Martin (2022): Betriebliche Vorbehalte gegenüber Langzeitarbeitslosen sinken leicht in Krisenzeiten. Unveröffentlichtes Manuskript. Im Erscheinen als IAB-Kurzbericht (vorauss. Nr. 17/2022), Nürnberg.
- Gürtzgen, Nicole; Kubis, Alexander (2021): Über ein Viertel der mit Corona-Hilfen unterstützten Betriebe sieht dennoch eine Insolvenzgefahr (Serie "Corona-Krise: Folgen für den Arbeitsmarkt"). In: IAB-Forum, 06.04.2021, o. Sz.
- Gürtzgen, Nicole; Kubis, Alexander; Kufner, Benjamin (2020): Großbetriebe haben während des Covid-19-Shutdowns seltener als kleine Betriebe Beschäftigte entlassen (Serie "Corona-Krise: Folgen für den Arbeitsmarkt"). In: IAB-Forum, 03.07.2020, o. Sz.
- Hutter, Christian (2020): A new indicator for nowcasting employment subject to social security contributions in Germany. In: Journal for labour market research, Vol. 54, No. 1, Art. 4, S. 1-10.

Küfner, Benjamin; Zins, Stefan; Sakshaug, Joseph (2021): More Clarification, Less Item Nonresponse in Establishment Surveys? A Split-Ballot Experiment. In: Survey research methods, Vol. 15, No. 2, S. 195-206.

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

Impressum

FDZ-Datenreport 08|2022

Veröffentlichungsdatum

15. September 2022

Herausgeber

Forschungsdatenzentrum
der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt- und Berufsforschung
Regensburger Straße 104
90478 Nürnberg

Nutzungsrechte

Diese Publikation ist unter folgender Creative-Commons-Lizenz veröffentlicht:
Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Bezugsmöglichkeit dieses Dokuments

https://doku.iab.de/fdz/reporte/2022/DR_08-22.pdf

Dokumentation Version

IABSE0019_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2208.de.v1

Datensatz Version

IAB SE 0019 v1, DOI: 10.5164/IAB.IABSE0019.de.en.v1

Bezugsmöglichkeit aller Veröffentlichungen der Reihe „FDZ-Datenreport“

<https://fdz.iab.de/forschung/publikationen/fdz-datenreporte/>

Website

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

Rückfragen zum Inhalt

Alexander Kubis
Telefon: 0911 179-8978
E-Mail: alexander.kubis@iab.de