

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

01|2026 DE IAB-BETRIEBSANEL (IAB-BP) 1993-2024

Lisa Bellmann, Miriam Gensicke, Katrin Grau, Susanne Kohaut, Iris Möller, Michael Oberfichtner, Barbara Schwengler, Nikolai Tschersich, Matthias Umkehrer



IAB-BETRIEBSANEL (IAB-BP) 1993-2024

Lisa Bellmann (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Miriam Gensicke (Verian),

Katrin Grau (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Susanne Kohaut (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Iris Möller (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Michael Oberfichtner (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Barbara Schwengler (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA)),

Nikolai Tschersich (Verian)

Matthias Umkehrer (Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA))

Die FDZ-Datenreports 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-Datenreports (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 hand, the data can be used to prepare evaluations.

Inhaltsverzeichnis

1	Einleitung	5
2	Datenbeschreibung	6
2.1	Veränderungen zum IAB-BP 9323.....	6
2.2	Steckbrief	6
2.3	Mengengerüst	9
3	Datenerhebung.....	9
3.1	Fragebogenentwicklung und Modulsystem.....	9
3.2	Grundgesamtheit und Stichprobenziehung	10
3.3	Logik der Betriebsnummer.....	12
3.4	Feldarbeit	14
4	Datenaufbereitung	15
4.1	Datenprüfung und Datenbereinigung.....	15
4.2	Der Umgang mit fehlenden Werten.....	16
5	Datenorganisation	17
5.1	Kennzeichnung von Teilgruppen für Auswertungen und Gewichtung.....	17
5.2	Querschnittsauswertung	18
5.3	Längsschnittauswertung	20
6	Verwendungshinweise und Datennutzung.....	21
6.1	Hilfe zum Aufbau eines Panel-Datensatzes.....	21
6.2	Sensible Merkmale und Zuspielvariablen	22
6.3	Datenzugang	25
7	Literatur	26
8	Anhang	27
8.1	Verfahren der Querschnittsgewichtung.....	27

Tabellenverzeichnis

Tabelle 1: Steckbrief	6
Tabelle 2: Mengengerüst	9
Tabelle 3: Basis- und Zusatzmodule	10
Tabelle 4: Branchengliederung nach 20 Wirtschaftszweigen (WZ 2008) für die Stichprobenziehung und die Querschnittsgewichtung seit der Welle 2023.....	11
Tabelle 5: Betriebsgrößenklassen für die Stichprobenziehung und die Querschnittsgewichtung.....	12
Tabelle 6: Übersicht über die Durchführungsart der Interviews.....	14
Tabelle 7: Anzahl der auswertbaren Interviews und Ausschöpfungsrate im Jahr 2024.....	15
Tabelle 8: Inhaltliche Fragen mit hohen (5 Prozent oder mehr) Antwortausfällen (k.A.); nur Querschnittsfälle	16
Tabelle 9: Gruppenkennzeichen in der Variable wellYYYY.....	17

Zusammenfassung

Dieser Datenreport beschreibt das IAB-Betriebspanel (IAB-BP) 1993-2024. Das IAB-Betriebspanel ist eine jährliche repräsentative Arbeitgeberbefragung zu betrieblichen Determinanten der Beschäftigung. Die Erhebung wird seit 1993 in Westdeutschland und seit 1996 auch in Ostdeutschland durchgeführt. Sie stellt die zentrale Quelle für Analysen zur ArbeitskräfteNachfrage auf dem Arbeitsmarkt in Deutschland dar.

Abstract

This data report describes the IAB Establishment Panel (IAB-BP) 1993-2024. The IAB Establishment Panel is an annual representative survey on various topics such as the determinants of labour demand. It has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany, too. The IAB Establishment Panel is the central basis for the analysis of labour demand in Germany.

Keywords

Establishment survey, labour market data, data manual, panel data

Danksagung

Für ihre Mitarbeit und Unterstützung danken wir den Kolleginnen und Kollegen des Forschungsdatenzentrums (FDZ), des Bereichs Betriebe und Beschäftigung (BB) sowie des Erhebungsinstituts Verian (ehemals Kantar Public).

Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich. Nähere Informationen dazu auf der Internetseite <https://fdz.iab.de>.

1 Einleitung

Das IAB-Betriebspanel (IAB-BP) ist eine jährliche repräsentative Arbeitgeberbefragung zu betrieblichen Determinanten der Beschäftigung. Die Erhebung wird seit 1993 in Westdeutschland und seit 1996 auch in Ostdeutschland durchgeführt. Sie stellt die zentrale Quelle für Analysen zur ArbeitskräfteNachfrage auf dem Arbeitsmarkt in Deutschland dar.

Jährlich werden von Ende Juni bis Oktober bundesweit rund 14.900 Betriebe aller Wirtschaftszweige und Größenklassen befragt. Die Befragung wird im Mixed-Mode als Online-Interview zum Selbstausfüllen, in persönlich-mündlichen oder telefonischen Interviews von Verian im Auftrag des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) durchgeführt.

Die besondere Bedeutung dieser repräsentativen Betriebsbefragung, die ein breites Fragenspektrum zu einer Vielzahl beschäftigungspolitischer Themen umfasst, wird in einem Brief der Vorsitzenden des Vorstands der Bundesagentur für Arbeit (BA) angekündigt.

Ergänzt wird das jährliche Standard-Fragenprogramm um jeweils aktuelle Themenschwerpunkte. Mittlerweile existiert das IAB-Betriebspanel in Westdeutschland seit 1993 und in Ostdeutschland seit 1996 und stellt als umfassender Längsschnittdatensatz die Grundlage für die Erforschung der Nachfrageseite des Arbeitsmarktes dar. Die Angaben der Betriebe sollen helfen, die Vermittlungs- und Beratungstätigkeiten der BA näher an der betrieblichen Realität zu orientieren. Die Analysen werden auch zur Entscheidungsfindung von Politik, Tarifparteien und Verbänden genutzt.

Dieser Datenreport dokumentiert die Daten des IAB-Betriebspanels 1993-2024 (IAB-BP 9324). Er enthält Informationen über die Inhalte der Befragung, Erhebungsmethode, Stichprobenziehung, Datenaufbereitung, Anonymisierung, Gewichtung und Datenorganisation.¹ Fragebögen, sowie eine Variablenübersicht sind separat über das Forschungsdatenzentrum (FDZ) der BA im IAB erhältlich (<https://fdz.iab.de/unsere-datenprodukte/betriebsdaten/iab-betriebspanel/>).²

¹ Weiterhin können administrative Betriebsinformationen aus dem Betriebs-Historik-Panel (BHP, siehe <https://fdz.iab.de/unsere-datenprodukte/betriebsdaten/bhp/>) den im IAB-Betriebspanel befragten Betrieben zugespielt werden. Details finden sich in Abschnitt 6.2.

² Für das IAB-Betriebspanel stehen auch englische Variablen- und Wertelabels zur Verfügung. In Stata können die englischen Labels mit dem Befehl "label language en" aktiviert werden.

2 Datenbeschreibung

2.1 Veränderungen zum IAB-BP 9323

Im Betriebspanel der Welle 2024 (IABBP_2024.dta) werden keine Region-Variablen mehr zur Verfügung gestellt. Die Bundesland Variable (bula*) ist weiterhin im Datensatz verfügbar.

Konkret handelt es sich um folgende Variablen:

- by* (Bayern)
- byreg* (Bayern: Regierungsbezirke)
- bystr* (Bayern: Regionalabgrenzung)
- hb* (Bremen/Bremerhaven)
- he* (Hessen: Regierungsbezirke)
- rpbez* (Rheinland-Pfalz Bezirke)
- rp* (Rheinland-Pfalz: Regionen)
- sl* (Saarland: AA-Bezirke)
- sn* (Sachsen: Regierungsbezirke)
- bb* (Brandenburg: Region)

2.2 Steckbrief

Tabelle 1: Steckbrief

Kategorie	Beschreibung
Themen	<p><u>(Modul) Panelvariablen</u></p> <p>(A) Allgemeine Angaben zum Betrieb (B) Altersstruktur und Fachkräftebedarf (C) Berufsausbildung und Ausbildungsstellen (D) Beschäftigungs- und Standortentwicklung (E) Beschäftigungsentwicklung (F) Beschäftigungsform (G) Betriebliche Arbeitszeiten (H) Betriebliche Investitionen und Innovationen im Betrieb (I) Betriebliche Weiterbildung (J) Derzeitige Personalsuche (K) Geschäftspolitik und Geschäftsentwicklung (L) Kapital- und Gewinnbeteiligung (M) Löhne und Gehälter (N) Öffentliche Förderung (O) Personalabgänge im 1. Halbjahr (P) Personaleinstellungen und -abgänge im 1. Halbjahr (Q) Personalstruktur (R) Auswirkungen der Corona-Pandemie</p> <p><u>Zusatzvariablen</u></p> <p>1994 Kontakt mit dem Arbeitsamt 1997 Neugründungen; Rekrutierung von Fachkräften 1998 Investitionen im Ausland 1999 Altersteilzeit und vorgezogener Ruhestand; Umweltschutzgüter; Neuregelung 630 DM-Gesetz 2000 Fachkräftebedarf; ältere Arbeitnehmer; finanzielle Anreize für Beschäftigte 2001 Ausstattung mit Computern und Internet; Kapital- und Gewinnbeteiligung</p>

Kategorie	Beschreibung
	<p>2002 ältere Arbeitnehmer; Gesundheitsschutz; Chancengleichheit; Kontakt mit dem Arbeitsamt</p> <p>2003 Betriebliche Flexibilität</p> <p>2004 Personalstruktur (Führungskräfte), betriebliche Arbeitszeit</p> <p>2005 Personalrekrutierung, Umweltschutzzüge</p> <p>2006 Beschäftigungs- und Standortsicherung</p> <p>2007 Fachkräftebedarf</p> <p>2008 Beschäftigungs- und Standortsicherung; Chancengleichheit</p> <p>2009 Beschäftigungsformen und die Wirtschafts- und Finanzkrise</p> <p>2010 Wirtschafts- und Finanzkrise; Leiharbeit; Kurzarbeit; Auslandsengagement</p> <p>2011 Fachkräftebedarf, (Branchen-) Tarifvertrag</p> <p>2012 Frauen in Führungspositionen, Gesundheitsschutz</p> <p>2013 Befristungsgrund, Fachkräfterekrutierung und Kompromisse, außertarifliche und übertarifliche Entlohnung, Beschäftigungs- und Standortsicherung, Besetzung von Ausbildungsstellen</p> <p>2014 Frauen in Führungspositionen, Mindestlohn, Fachkräftebedarf, vorzeitig aufgelöste Ausbildungsverträge</p> <p>2015 Mindestlohn, Weiterbildungsförderung, Ältere</p> <p>2016 Auszubildende; Automatisierung/Digitalisierung; Integration; Langzeitarbeitslose</p> <p>2017 Beschäftigung Geflüchteter; Mindestlohn; Weiterbildung; Digitalisierung</p> <p>2018 Beschäftigung Geflüchteter, ökologische Nachhaltigkeit, Frauen in Führungspositionen, Mobiles Arbeiten</p> <p>2019 Hochgeschwindigkeitsinternetanschluss, Mindestlohn, Befristungsgründe, Lösung von Ausbildungsverträgen, Zusatzleistung</p> <p>Auszubildende, Einsatz von Robotern</p> <p>2020 Auswirkungen der Corona Pandemie, Kurzarbeit, Veränderung Wertschöpfungskette, Mindestausbildungsvergütung, E-Learning, Home-Office</p> <p>2021 Mobiles Arbeiten, Befristungsgrund</p> <p>2022 Auswirkungen des Kriegs in der Ukraine</p> <p>2023 Industrie- und Wirtschaftsspionage; Fachkräftesicherung</p> <p>2024 Einstellungen mit Kompromissen, Nutzung von grünem Wasserstoff, Auszubildende ohne Schulabschluss, Maßnahmen zur Vereinbarkeit von Beruf und Familie</p>
Untersuchungseinheit	<p>Betriebe mit mindestens einer oder einem sozialversicherungspflichtig Beschäftigten zum Stichtag 30.06. im Jahr vor der Erhebung.</p> <p>Betriebe ohne sozialversicherungspflichtig Beschäftigte werden vom IAB-Betriebspanel nicht erfasst.</p> <p>Seit 2004 sind zudem private Haushalte und exterritoriale Organisationen ausgeschlossen.</p>
Fallzahlen	Zwischen 4.114 und 17.376 Betriebe pro Jahr
Zeitraum	Westdeutschland: 1993 bis 2024; Ostdeutschland: 1996 bis 2024
Zeitbezug	<p>Die Fragen beziehen sich auf unterschiedliche Zeiträume:</p> <ul style="list-style-type: none"> die aktuelle Situation das vergangene Geschäftsjahr das erste Halbjahr des Befragungsjahres. <p>Die Fragen mit Stichtag beziehen sich auf den 30. Juni des Vorjahres oder des Erhebungsjahres.</p>
Regionale Gliederung	Ost/West, Bundesländer, weitere Merkmale unter „Sensible Merkmale“ (siehe Datenzugang)
Gebietsstand	Nicht gebietsstands-bereinigt. Der Gebietsstand bezieht sich jeweils auf den 30. Juni des Vorjahres.
Erhebungsdesign	<p>Arbeitgeberbefragung (repräsentative Zufallsstichprobe, stratifiziert nach Betriebsgröße (10 Kategorien), Branche (20 Kategorien) und 16 Bundesländer. Befragung durch persönlich-mündliche Interviews (CAPI), telefonische Interviews (CATI) sowie Selbstausfüller (CAWI)).</p> <p>Die Stichprobe setzt sich zusammen aus <u>Wiederholerstichprobe</u>:</p> <p>Betriebe mit einem gültigen Interview aus dem Vorjahr.</p> <p>Sie spiegelt den Panelcharakter des IAB-Betriebspanels wider.</p> <p>Sie ist notwendig, damit man über reine Zeitreihenanalysen hinausgehend Panelauswertungen durchführen kann.</p> <p><u>Nachbearbeitungsstichprobe</u>:</p> <p>Sie umfasst alle teilnahmebereiten Betriebe mit einem gültigen Interview aus dem Vorvorjahr. Diese Teilstichprobe erhöht die im Querschnitt auswertbare Anzahl von Fällen.</p>

Kategorie	Beschreibung
	<p>Ergänzungs- und Aufstockungsstichprobe: Sie schließt bundeslandspezifische Aufstockungen sowie für ostdeutsche Flächenländer (ohne Berlin) eine branchenspezifische Aufstockung im Verarbeitenden Gewerbe ein. Sie umfasst auch Betriebe mit einer neuen Betriebsnummer. Ziel dieser Teilstichprobe ist es, Panelausfälle zu ersetzen und die geforderten Fallzahlen in den einzelnen Bundesländern und im Verarbeitenden Gewerbe Ost zu erreichen sowie den wirtschaftlichen Strukturwandel abzubilden.</p>
Beteiligte Institutionen	<p>Auftraggeber: Forschungsbereich "Betriebe und Beschäftigung" des IAB, Ministerien/Senatsverwaltungen der Bundesländer und auswertende Institute Durchführung: Verian (ehemals Kantar Public) in Zusammenarbeit mit dem Institut SÖSTRA Sozialökonomische Strukturanalysen GmbH und dem Institut GEFRA – Gesellschaft für Finanz- und Regionalanalysen GbR. Partnerinstitute: https://iab.de/das-iab/befragungen/iab-betriebspanel/beteiligte-partnerinstitute-in-den-bundeslaendern/</p>
Frequenz der Datensammlung	Jährliche Befragung
Dateiformat/-größe	Stata; Zwischen 1.6 und 10.6 MB pro Jahr
Dateiorganisation	<p>Das IAB-Betriebspanel ist als eine Datei pro Jahr abgelegt. Hinzu kommt eine Datei mit grundlegenden betriebsbezogenen Informationen aus dem BHP für alle Jahre (die Basis-Betriebsdatei). Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen aus dem BHP bereitgestellt, entweder als eine Datei pro Welle (Variablenblöcke) oder als eine Datei für alle Jahre (Erweiterungsmodule). Die Dateien des BHP beschränken sich auf Betriebe, für die eine Zustimmung zur Datenverknüpfung vorliegt.</p>
Datenzugang	Gastaufenthalt oder Datenfernverarbeitung
Anonymisierungsgrad	Pseudonymisiert
Sensible Merkmale	<p>Kreiskennziffer (kkz), Wirtschaftszweigklassifikation 1993 (WZ93), Wirtschaftszweigklassifikation 2003 (WZ03) 5-Steller bzw. Wirtschaftszweigklassifikation 2008 (WZ08) 5-Steller, Gemeindetyp (bik) Daten aus dem BHP: Zeitkonsistente Klassifikation der Wirtschaftszweige (3-Steller) (w08_3_gen). Sensible Merkmale werden nur auf begründeten Antrag bereitgestellt.</p>
Zitierung der Daten und Datendokumentation	<p>Daten: Bellmann, Lisa; Gensicke, Miriam; Grau, Katrin; Kohaut, Susanne; Möller, Iris; Oberfichtner, Michael; Schwengler, Barbara; Tschersich, Nikolai; Umkehrer, Matthias (2026): "IAB-Betriebspanel (IAB-BP) – Version 9324 v1. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB). DOI: 10.5164/IAB.IABBP9324.de.en.v1</p> <p>Der Datenzugang erfolgt über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und anschließend mittels kontrollierter Datenfernverarbeitung beim FDZ.</p> <p>Datendokumentation: Bellmann, Lisa; Gensicke, Miriam; Grau, Katrin; Kohaut, Susanne; Möller, Iris; Oberfichtner, Michael; Schwengler, Barbara; Tschersich, Nikolai; Umkehrer, Matthias (2026): IAB-Betriebspanel (IAB-BP) 1993-2024. FDZ-Datenreport, 01/2026 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2601.de.v1</p>
Datensatzversion	IAB-Betriebspanel 1993-2024 (IAB-BP 9324) - Version 9324 v1; DOI: 10.5164/IAB.IABBP9324.de.en.v1

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

2.3 Mengengerüst

Die Betriebsbefragung setzt sich in jeder Welle zu Teilen aus Erst- und Wiederbefragten zusammen. Tabelle 2 zeigt die Fallzahlen für die Querschnittsfälle sowie die Zusammensetzung der bereitgestellten Datensätze.

Tabelle 2: Mengengerüst

Datensatz	Jahr	Nettostichprobenumfang	Anzahl Erstbefragte	Anzahl Wiederbefragte
IABBP_1993.dta	1993	4.265	4.265	0
IABBP_1994.dta	1994	4.138	238	3.900
IABBP_1995.dta	1995	4.096	511	3.585
IABBP_1996.dta	1996	8.342	4.845	3.497
IABBP_1997.dta	1997	8.850	1.522	7.328
IABBP_1998.dta	1998	9.194	1.946	7.248
IABBP_1999.dta	1999	9.762	2.220	7.542
IABBP_2000.dta	2000	13.931	5.503	8.428
IABBP_2001.dta	2001	15.536	4.167	11.369
IABBP_2002.dta	2002	15.407	3.327	12.080
IABBP_2003.dta	2003	15.856	3.742	12.114
IABBP_2004.dta	2004	15.689	2.564	13.125
IABBP_2005.dta	2005	15.821	2.714	13.107
IABBP_2006.dta	2006	15.449	2.511	12.938
IABBP_2007.dta	2007	15.643	2.883	12.760
IABBP_2008.dta	2008	15.455	2.566	12.889
IABBP_2009.dta	2009	15.522	2.716	12.806
IABBP_2010.dta	2010	15.614	2.766	12.848
IABBP_2011.dta	2011	15.283	1.784	13.499
IABBP_2012.dta	2012	15.556	2.698	12.858
IABBP_2013.dta	2013	15.725	2.494	13.231
IABBP_2014.dta	2014	15.577	2.576	13.001
IABBP_2015.dta	2015	15.500	2.394	13.106
IABBP_2016.dta	2016	15.341	2.327	13.014
IABBP_2017.dta	2017	15.421	2.577	12.844
IABBP_2018.dta	2018	15.263	2.710	12.553
IABBP_2019.dta	2019	15.438	3.861	11.577
IABBP_2020.dta	2020	16.686	5.892	10.794
IABBP_2021.dta	2021	15.217	3.131	12.086
IABBP_2022.dta	2022	14.575	3.095	11.480
IABBP_2023.dta	2023	14.700	4.464	10.236
IABBP_2024.dta	2024	15.857	5.206	10.651

3 Datenerhebung

3.1 Fragebogenentwicklung und Modulsystem

Die inhaltliche und methodische Konzeption des IAB-Betriebspanels wird vom zuständigen Team des IAB erarbeitet. Die Durchführung der Befragung erfolgt durch Verian. Der Panelcharakter des IAB-Betriebspanels wird auch bei der Entwicklung des Fragebogens berücksichtigt. Die Fragebögen enthalten zwei Arten von Fragen-Modulen: In jeder Welle werden sog. Basismodule mit möglichst unveränderten Fragen eingesetzt. Diese betrieblichen Basisinformationen sind seit 2008 für jedes Jahr vorhanden. Die Basismodule werden durch Zusatzmodule ergänzt. In diesen werden in zeitlich definierten Abständen – meistens in einem zweijährigen Turnus – vertiefende Fragen gestellt. Ein Beispiel für die Wellen 2020 bis 2024 findet sich in Tabelle 3.

Tabelle 3: Basis- und Zusatzmodule

	2020	2021	2022	2023	2024
Basismodule					
Beschäftigungsentwicklung	X	X	X	X	X
Geschäftspolitik und -entwicklung	X	X	X	X	X
Investitionen und Innovationen	X	X	X	X	X
Personalstruktur	X	X	X	X	X
Personalbewegung inkl. Personalsuche	X	X	X	X	X
Berufsausbildung	X	X	X	X	X
Weiterbildung Basis	X	X	X	X	X
Arbeitszeiten Basis	X	X	X	X	X
Löhne und Gehälter	X	X	X	X	X
Allgemeine Angaben zum Betrieb	X	X	X	X	X
2-jährige Zusatzmodule					
Weiterbildung		X		X	
Innovationen im Betrieb		X		X	
Organisatorische Änderungen		X		X	
Frauen in Führungspositionen	X		X		X
Altersstruktur der Beschäftigten		X		X	
Gewinn- und Kapitelbeteiligung		X		X	
Betriebliche Arbeitszeiten/Schwankungen	X		X		X

In jedem Jahr gibt es außerdem zusätzlich zu den Basismodul-Fragen sogenannte Schwerpunktfragen. Im Steckbrief (vgl. Tabelle 1) sind die Themen gelistet, die in den Fragebögen außerhalb des Modulsystems aufgenommen wurden.

Die Fragebögen sind über das FDZ erhältlich. Eine Übersicht zu den Fragen über die Befragungswellen hinweg bietet außerdem das Excel-File „Variablenliste“, welches ebenfalls über die Internetseite des FDZ verfügbar ist: <https://fdz.iab.de/unsere-datenprodukte/betriebsdaten/iab-betriebspanel/>.

Zur Qualitätssicherung werden in einem kognitiven Pretest neue Fragen daraufhin getestet, ob sie für eine Betriebsbefragung prinzipiell geeignet erscheinen und ob die Betriebe voraussichtlich über die abgefragten Sachverhalte auskunftsähig sind, also die gewünschten Informationen auf Betriebsebene vorliegen. Dazu werden in rund 80 Betrieben aus unterschiedlichen Branchen und Betriebsgrößenklassen Interviews durchgeführt. Das Pretest-Interview besteht aus zwei Teilen. Die Testperson wird zunächst gebeten, den Fragebogen zu beantworten. Der zweite Schritt beinhaltet dann den eigentlichen kognitiven Test. Die Betriebe werden gebeten, jede Frage zu kommentieren, und außerdem gefragt, ob sie bei der Beantwortung Schwierigkeiten hatten. Von besonderem Erkenntnisinteresse sind die Auskunftsähigkeit der Betriebe sowie Verständnisschwierigkeiten und Interpretationsprobleme. Die Ergebnisse des kognitiven Pretests fließen in die weitere Fragenentwicklung ein. Der Pretest wird von speziell geschulten Projektmitarbeitenden des IAB durchgeführt.

3.2 Grundgesamtheit und Stichprobenziehung

Die Grundgesamtheit des IAB-Betriebspanels umfasst alle Betriebsstätten mit mindestens einer oder einem sozialversicherungspflichtig Beschäftigten zum Stichtag 30. Juni des Vorjahres. Basis der Stichprobenziehung ist die Betriebsdatei der BA. Sie enthält alle Betriebe, die im Rahmen des

Meldeverfahrens zur Sozialversicherung ihre sozialversicherungspflichtig Beschäftigten an die Sozialversicherungsträger melden und eine Betriebsnummer erhalten. Zum Stichtag 30. Juni 2023 enthielt die Grundgesamtheit 420 Tsd. Betriebe für Ostdeutschland und 1.704 Tsd. Betriebe für Westdeutschland. Betriebe ohne sozialversicherungspflichtig Beschäftigte – z. B. Ein-Personen-Betriebe oder Betriebe, die nur geringfügig Beschäftigte, Beamte oder Beamten beschäftigen – werden vom IAB-Betriebspanel nicht erfasst. Seit 2004 sind zudem private Haushalte und extritoriale Organisationen ausgeschlossen. Aus diesem Grund weisen z. B. die volkswirtschaftlichen Gesamtrechnungen der Länder deutlich mehr Personen aus als das IAB-Betriebspanel.

Die Stichprobe ist nach Betriebsgröße (10 Kategorien), Branche (20 Kategorien³) und Bundesland (16 Kategorien) disproportional geschichtet (zur Branchen- und Größenklassenaufteilung siehe Tabelle 4 und Tabelle 5). Für die Branchengliederung wird seit der Erhebungswelle 2009 die Wirtschaftszweigklassifikation WZ 2008 herangezogen. Die Sollbesetzung der einzelnen Zellen wird zum einen durch den Umfang der Basisstichprobe sowie der länder- und branchenspezifischen Aufstockungsstichproben vorgegeben. Zum anderen werden die einzelnen Zellen nach dem „Prinzip der optimalen Schichtung“ näherungsweise beschäftigtenproportional gezogen. Große Betriebe, kleine Bundesländer und kleine Branchen sowie das Verarbeitende Gewerbe in Ostdeutschland sind aus diesen Gründen in der Stichprobe überproportional vertreten. Diese Disproportionalitäten werden im Rahmen der Gewichtung wieder ausgeglichen (siehe Kapitel 5.2, 5.3 sowie im Anhang).

Tabelle 4: Branchengliederung nach 20 Wirtschaftszweigen (WZ 2008) für die Stichprobenziehung und die Querschnittsgewichtung seit der Welle 2023

Code	Branche (Variablen: br*fb*)	WZ 2008
1	Land- und Forstwirtschaft, Fischerei	01-03
2	Bergbau und Gewinnung von Steinen und Erden, Energie und Wasserversorgung, Abwasser- und Abfallentsorgung	05-09, 35-39
3	Nahrungs- und Genussmittel	10-12
4	Verbrauchsgüter	13-18
5	Produktionsgüter	19-24
6	Investitions- und Gebrauchsgüter	25-33
7	Baugewerbe	41-43
8	Großhandel, KFZ-Handel und -reparatur	45-46
9	Einzelhandel	47
10	Verkehr und Lagerei	49-53
11	Information und Kommunikation	58-63
12	Gastgewerbe	55-56
13	Finanz- und Versicherungsdienstleistungen	64-66
14	Wirtschaftliche, wissenschaftliche und freiberufliche Dienstleistungen	68-82
15	Erziehung und Unterricht	85
16	Gesundheitswesen	86
17	Sozialwesen	87-88
18	Sonstige Dienstleistungen	90-93, 95-96
19	Interessenvertretungen	94
20	Öffentliche Verwaltung, Verteidigung, Sozialversicherung	84

³ Vom Erhebungsjahr 2010 bis 2022 galt eine 19er Branchengliederung. Im Erhebungsjahr 2023 wurden die Branchen Gesundheitswesen und Sozialwesen für die Stichprobenziehung getrennt.

Tabelle 5: Betriebsgrößenklassen für die Stichprobenziehung und die Querschnittsgewichtung

Größenklasse	Sozialversicherungspflichtig Beschäftigte (Variablen: grkl*)
1	1 bis 4
2	5 bis 9
3	10 bis 19
4	20 bis 49
5	50 bis 99
6	100 bis 199
7	200 bis 499
8	500 bis 999
9	1000 bis 4.999
10	5.000 und mehr

Es gibt insgesamt drei Teilstichproben:

- **Wiederholerstichprobe:** Sie umfasst alle teilnahmebereiten Betriebe mit einem gültigen Interview aus dem Vorjahr. Die Wiederholerstichprobe spiegelt den Panelcharakter des IAB-Betriebspanels wider. Sie ist notwendig, damit man über reine Zeitreihenanalysen hinausgehend Panelauswertungen durchführen kann.
- **Nachbearbeitungsstichprobe:** Sie umfasst alle teilnahmebereiten Betriebe mit einem gültigen Interview aus dem Vorjahr. Diese Teilstichprobe erhöht die im Querschnitt auswertbare Anzahl von Fällen.
- **Ergänzungs- und Aufstockungsstichprobe:**
 - **Teilstichprobe „Verarbeitendes Gewerbe Ost“:** In Ostdeutschland (ohne Berlin) wurde die Stichprobe wie in den Vorjahren für eine Sonderauswertung durch das Leibniz-Institut für Wirtschaftsforschung Halle (IWH) speziell im Verarbeitenden Gewerbe aufgestockt. Diese Aufstockungsstichprobe bezieht sich ab 2007 nur noch auf die ostdeutschen Länder ohne Berlin.
 - **Teilstichproben für landesspezifische Auswertungen:** In insgesamt 12 der 16 Bundesländer wurde die Basisstichprobe des IAB-Betriebspanels im Jahr 2024 aus Landesmitteln auf mindestens 800 auswertbare Fälle aufgestockt, um landesspezifische Auswertungen zu ermöglichen. Dies betrifft im Jahr 2024 das Land Berlin und sämtliche ostdeutschen Bundesländer (Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen) sowie sechs der westdeutschen Bundesländer (Baden-Württemberg, Bayern, Bremen, Hessen, Rheinland-Pfalz und Saarland).
 - **Ergänzungsstichprobe:** Eine vom IAB neu aus der Betriebsdatei der Bundesagentur für Arbeit gezogene Stichprobe erstmals zu befragender Betriebe zum Ausgleich der Panelmortalität und zur laufenden Anpassung der Stichprobe an den wirtschaftlichen Strukturwandel (neue Betriebsnummern).

3.3 Logik der Betriebsnummer

Die Betriebe für die Bruttostichprobe werden aus der Betriebsdatei der BA gezogen. Es handelt sich demnach bei einem Betrieb um eine regional und wirtschaftlich abgegrenzte Einheit mit mindestens einem sozialversicherungspflichtig Beschäftigten. Für jeden Betrieb wird von der zuständigen Arbeitsagentur eine achtstellige Betriebsnummer vergeben, über die Betriebe ihre

sozialversicherungspflichtig Beschäftigten, Minijobber und Auszubildenden zur Sozialversicherung an die Sozialversicherungsträger melden. Anstelle dieser ursprünglichen Betriebsnummer wird für wissenschaftliche Auswertungen eine pseudonymisierte Betriebsnummer bereitgestellt, die sogenannte „betnr“.⁴

Bei der Vergabe von Betriebsnummern gilt:

- Niederlassungen eines Arbeitgebers in verschiedenen Gemeinden erhalten grundsätzlich eine eigene Betriebsnummer.
- Niederlassungen eines Arbeitgebers innerhalb einer Gemeinde werden zu einem Betrieb mit einer Betriebsnummer zusammengefasst, sofern diese dieselbe wirtschaftliche Tätigkeit ausüben. Niederlassungen eines Arbeitgebers innerhalb einer Gemeinde mit unterschiedlichen wirtschaftlichen Tätigkeiten erhalten eine eigene Betriebsnummer.
- Ein Betrieb selbst kann mehrere Betriebsnummern haben; das gilt insbesondere für größere Betriebe mit unterschiedlichen Funktionsbereichen, die verwaltungstechnisch eigenständig sein sollen.
- Unternehmensverflechtungen spielen keine Rolle. Jedes rechtlich eigenständige Unternehmen bekommt nach den eben genannten Regeln Betriebsnummern zugewiesen.

Betriebsnummern werden (neu) vergeben, wenn

- der Betrieb zuvor noch keine Betriebsnummer hatte (i. d. R., weil der Betrieb zum ersten Mal mindestens eine oder einen sozialversicherungspflichtig Beschäftigten hat),
- sich die wirtschaftliche Tätigkeit des Betriebs geändert hat oder
- sich die Eigentumsverhältnisse ändern.

Immer wenn eine Betriebsnummer neu für die Stichprobe gezogen wird, vergibt das Erhebungsinstitut Verian eine eigene Identifikationsnummer für das IAB-Betriebspanel, die sogenannte „idnum“. In der Regel ist eine „betnr“ eineindeutig einer „idnum“ zugewiesen. In seltenen Fällen kann es jedoch zu folgenden Ausnahmen kommen:

- Einer „idnum“ sind mehrere „betnr“ zugeordnet

Es kann vorkommen, dass über die Jahre hinweg einer „idnum“ mehrere „betnr“ zugeordnet sind. Dies kann beispielsweise dann auftreten, wenn ein Betrieb seinen wirtschaftlichen Schwerpunkt verändert hat oder wenn es zu einem Eigentümerwechsel kam. Der Betrieb bekommt dann von der BA eine neue „betnr“, die „idnum“ ändert sich jedoch nicht, weil Verian nur bei erstbefragten Betrieben die Zuordnung „idnum“ - „betnr“ vornimmt. In den Folgejahren nach dem ersten Interview bleibt die „idnum“ somit unverändert.

- Einer „betnr“ sind mehrere „idnum“ zugeordnet

Seit 2002 konnten vormals befragte Betriebe, die bis dahin für die Stichprobenziehung ausgeschlossen wurden, erneut gezogen werden. Sie erhielten dann eine neue „idnum“, haben aber die gleiche „betnr“.

Seit der Version IAB-BP 9321 werden die Kombinationen von „betnr“ und „idnum“ so bereinigt, dass es für ein gegebenes Jahr nur noch eindeutige Zuordnungen gibt. Die Zuspielung von

⁴ Für weiterführende Informationen zur Betriebsnummer siehe:
<https://www.arbeitsagentur.de/unternehmen/betriebsnummern-service/alles-wichtige>.

administrativen Daten erfolgt dann anhand der „idnum“ (siehe Abschnitt 6.2). Die „betnr“ wird aus dem Datenprodukt entfernt.

Schließlich sei noch darauf hingewiesen, dass Neugründungen nicht über das erste Auftreten einer „idnum“ identifiziert werden können. Zur Identifikation von Neugründungen empfiehlt es sich, entweder die Befragungsdaten über das Merkmal „neuYYYY“ oder das BHP-Erweiterungsmodul „Betriebsdynamik“ auszuwerten. Das Verfahren zur Identifikation von Neugründungen mit Hilfe der BHP-Daten ist in Ganzer et al. (2022) beschrieben. Dabei sei insbesondere auf Hethay & Schmieder (2010) und Hethay-Maier & Schmieder (2013) verwiesen.

3.4 Feldarbeit

Die Datenerhebung wird von geschulten Interviewenden aus dem Stab von Verian durchgeführt. Dabei kommt ein elektronischer Fragebogen zum Einsatz. Dieser ist so gestaltet, dass der Fragebogen sowohl von den Interviewenden als auch vom Betrieb selbst ausgefüllt werden kann. Seit der Welle 2024 werden keine Papierfragebögen mehr verwendet.

Die Kontaktbeschränkungen während der Corona-Pandemie sowie die seitdem deutliche Zunahme an Homeoffice seitens der betrieblichen Ansprechpersonen führten seit der Welle 2020 dazu, dass auch telefonische Interviews zugelassen sind; siehe Tabelle 6.

Tabelle 6: Übersicht über die Durchführungsart der Interviews

	2021	2022	2023	2024	2024	2024
Durchführungsart	Gesamt	Gesamt	Gesamt	Gesamt	Wiederholer	Erstbefragte (inkl. temp. Ausfälle)
ganz telefonisch durchgeführt	48,1%	50,9%	42,7%	32,9%	31,7%	33,5%
ganz persönlich-mündlich durchgeführt	7,3%	6,8%	9,0%	10,2%	10,8%	8,9%
teilweise telefonisch durchgeführt	1,0%	1,8%	0,4%	3,1%	3,9%	1,5%
teilweise persönlich-mündlich durchgeführt	0,1%	0,0%	0,2%	0,6%	0,8%	0,2%
ganz von den Befragten selbst ausgefüllt	43,5%	40,4%	47,6%	53,1%	51,8%	56,0%

Der Einsatz von geschulten Interviewenden führt dazu, dass die ausgefüllten Fragebögen grundsätzlich weniger Fehler aufweisen als die im Selbstausfüllermodus erhobenen und der Anteil fehlender Angaben niedriger ist. Grundsätzlich beinhaltet das computergestützte Erhebungsinstrument eine automatisierte Filterung, vorgegebene Wertebereiche sowie ausgewählte Plausibilitätschecks, um auch im Selbstausfüllermodus fehlerhafte Eingaben möglichst gering zu halten.

Die Befragungsart variiert erheblich in Abhängigkeit von der Betriebsgröße: Der Anteil der vollständig durch Interviewende befragten Betriebe (persönlich-mündlich oder telefonisch) nimmt mit zunehmender Betriebsgröße ab (von 65 Prozent bei den Kleinstbetrieben mit 1 bis 4 Beschäftigten bis 38 Prozent bei den Großbetrieben mit 5.000 und mehr Beschäftigten). Denn je größer die Betriebe sind, desto häufiger können die Auskunftspersonen insbesondere die komplexen quantitativen Angaben zur Beschäftigtenstruktur, zur Personalsuche, zu Einstellungen und Entlassungen, zur Ausbildung, zum Geschäftsvolumen und zu den Investitionen nur nach aufwendigen Vorarbeiten bereitstellen.

Die Feldarbeit für die Welle 2024 begann am 01. Juli 2024, das letzte digitale Interview wurde am 18. November 2024 angenommen. Aus dem in der Feldarbeit eingesetzten Brutto von 72.677 Betrieben konnten insgesamt 16.246 auswertbare Interviews realisiert werden (siehe Tabelle 7). Die Teilnahme an der Befragung erfolgt auf freiwilliger Basis. Für die Befragung ausgewählte Betriebe erhalten ein Anschreiben vom Vorstand der BA, in dem sie um Teilnahme an der Befragung gebeten werden. Des Weiteren enthält das Anschreiben Informationen zur Zielsetzung der Studie und Hinweise zum Datenschutz.

Tabelle 7: Anzahl der auswertbaren Interviews und Ausschöpfungsrate im Jahr 2024

Teilstichproben	eingesetztes Brutto (absolut)	Auswertbare Interviews	
		absolut	in % vom Brutto
Wiederholerstichprobe	14.645	10.541	72,0%
Nachbearbeitungsstichprobe	1.453	499	34,3%
Erstbefragtenstichprobe	56.579	5.206	9,2%
Gesamt	72.677	16.246	22,4%

4 Datenaufbereitung

4.1 Datenprüfung und Datenbereinigung

Ein Teil der Datenprüfung erfolgt automatisch über die vom Befragungsinstitut verwendete Software während der Befragung. Diese stellt sicher, dass alle numerischen Angaben ausschließlich in den zuvor definierten Wertebereichen liegen. Des Weiteren wurden für einige numerische Angaben Plausibilitätsroutinen implementiert.

Parallel zur Feldarbeit werden die bereits erhobenen Daten auf Vollständigkeit, Konsistenz und Plausibilität sowohl im Querschnitt als auch im Längsschnitt geprüft. Dabei werden vier Prüfungsarten unterschieden:

- Filterfehler: Wurde die Filteranweisung in der vorgesehenen Weise beachtet oder haben Betriebe fälschlicherweise eine Frage beantwortet bzw. nicht beantwortet?
- Vollständigkeitsprüfungen: Bei ausgewählten Fragen wird geprüft, ob die Frage beantwortet wurde. Dies betrifft vor allem Fragen, die für die Gewichtung relevant sind, wie die Frage nach der Anzahl der sozialversicherungspflichtig Beschäftigten.
- Plausibilitätsprüfungen: Bei diesen Prüfungen werden Angaben überprüft, die im Regelfall unwahrscheinlich sind, aber in der Praxis durchaus vorkommen können. Ein Beispiel für solch eine Prüfung ist die Überprüfung des Pro-Kopf-Einkommens. Bei verhältnismäßig niedrigen oder verhältnismäßig hohen Werten schlägt die Prüfung an.
- Konsistenzprüfungen: Diese Prüfungen beziehen sich auf logische Beziehungen zwischen verschiedenen Antworten. Konsistenzkriterien werden z. B. verletzt, wenn der Fragebogen

widersprüchliche Angaben enthält. Zu den Konsistenzprüfungen gehört z. B. das Abprüfen der angegebenen Gesamtsumme in einer Frage mit der Summe der Einzelwerte.

Wenn fehlende oder fehlerhafte Angaben anhand des Fragebogens nicht ergänzt bzw. korrigiert werden können, wird im Rahmen einer telefonischen Nachbefragung versucht, zusammen mit der Zielperson eine Klärung herbeizuführen. Im Ergebnis werden fehlende Angaben ergänzt und fehlerhafte Angaben korrigiert. Bei Plausibilitätsprüfungen werden nach Rücksprache mit dem Betrieb unplausible Werte bei einer plausiblen Begründung freigesetzt – der Wert also trotz Verletzung der Prüfbedingung als in Ordnung angesehen.

4.2 Der Umgang mit fehlenden Werten

In einigen Fragestellungen beinhaltet die „Weiß nicht“-Kategorie eine verwertbare Information und kann zu Analysezwecken verwendet werden. So gibt es z. B. Fragen mit der Antwortkategorie „Weiß noch nicht“, „Kann ich noch nicht sagen“ oder „Kann ich nicht sagen“. In solchen Fällen erhält die entsprechende Variable im Datensatz einen eigenen Code.

Bei allen anderen Fragen gibt es keine gesonderte Antwortkategorie „Weiß nicht“ (oder eine inhaltlich gleichgerichtete Kategorie), weshalb diese Kategorie grundsätzlich nicht separat ausgewiesen werden kann. Wurde eine Frage von der Zielperson nicht beantwortet (z. B. weil die Zielperson die Antwort nicht weiß oder über den Sachverhalt keine Auskunft geben will), so erhält die entsprechende Variable im Datensatz den Code „-9“ für „Weiß nicht/Keine Angabe“.

Tabelle 8 gibt einen Überblick über die Fragen mit sehr hohen Anteilen (5 Prozent und mehr) fehlender Werte. Fragen des Kapitels „Weitere Fragen an die Auskunft gebende Person“ sind nicht berücksichtigt. Bei einer Auswertung von Variablen mit vielen fehlenden Angaben sollten dadurch bedingte mögliche Verzerrungen berücksichtigt werden. Die Quoten fehlender Angaben beziehen sich auf die Querschnittsfälle. Sofern die betreffenden Fragen auch im Vorjahr gestellt wurden, sind die damaligen KA-Anteile zum Vergleich mit angegeben.

Tabelle 8: Inhaltliche Fragen mit hohen (5 Prozent oder mehr) Antwortausfällen (k.A.); nur Querschnittsfälle

Frage/Variable	Inhalt	Einheit	KA-Anteil *	
			2024	2023
bf59lohn	Brutto-/Gehaltssumme Juni 2024 (EUR)	EUR	21,2%	22,2%
bf41gvol	Geschäftsvolumen 2023 (EUR)	EUR	17,3%	17,7%
bf17vor	Anteil Vorleistungen Fremdkosten am Umsatz 2023 (Prozent)	Prozent	15,0%	15,5%
bf23inv	Summe aller Investitionen 2023 (EUR)	EUR	8,5%	6,9%
bf24erw	Anteil Erweiterungsinvestitionen 2023 (Prozent)	Prozent	8,5%	7,6%
bf20proz	Erwartete Entwicklung Geschäftsvolumen 2023 → 2024 (Prozent)	Prozent	8,2%	5,7%
bf62e	Geförderte Weiterbildung: Arbeitsplatzwechsel	Anzahl	6,2%	4,1%

Frage/Variable	Inhalt	Einheit	KA-Anteil *	
			2024	2023
bf62g	Geförderte Weiterbildung: Qualitätszirkel, Werkstattzirkel, Lernstatt und Beteiligungsgruppe	Anzahl	6,1%	3,8%
bf62h	Geförderte Weiterbildung: Sonstig Weiterbildungsmaßnahmen	Anzahl	5,3%	4,2%

* In Prozent der Fälle (ungewichtet), die die entsprechende Frage zu beantworten hatten.

5 Datenorganisation

5.1 Kennzeichnung von Teilgruppen für Auswertungen und Gewichtung

Um den Nutzenden des IAB-Betriebspanels die Identifikation verschiedener Teilgruppen für Querschnitts- und insbesondere für Längsschnittausswertungen zu erleichtern, wurden die relevanten Teilgruppen gekennzeichnet. Dazu müssen folgende Informationen in geeigneter Weise in Verbindung gebracht werden:

- Feldergebnis der laufenden Welle (auswertbares Interview mit/ohne sozialversicherungspflichtig Beschäftigte zum Stichtag, Betrieb erloschen, nachbefragbare/endgültige Ausfälle).
- Wurde der Betrieb in der vorherigen Welle befragt (Antwort aus der vorherigen Welle, wiederbefragbare Ausfälle aus der vorherigen Welle, Ergänzungsstichprobe)?
- Identität der befragten Betriebseinheit (Wurde derselbe Betrieb befragt wie beim letzten Mal)?

Dafür wurde folgendes Konzept entwickelt:

- Jeder Fall erhält in jeder Welle eine eindeutige Kennzeichnung, die die o. g. Kriterien berücksichtigt. Dieses sog. Wellenkennzeichen ist in der Variablen „wellYYYY“ abgespeichert, wobei YYYY für das Jahr steht, in dem die Erhebung stattgefunden hat (also well1993 für Welle 1 im Jahr 1993, well1994 für Welle 2 im Jahr 1994 usw.). Diese Kennzeichnung erfolgt mit Hilfe eines Buchstabens (siehe Tabelle 9).

Tabelle 9: Gruppenkennzeichen in der Variable wellYYYY

	Kennbuchstaben	
	mit sozialversicherungspflichtig Beschäftigte am jeweiligen Stichtag (30.06.)	ohne
Fälle mit Interview in der laufenden Welle Erstmals (= zum Ziehungszeitpunkt) befragte Fälle aus der Basisstichprobe und den jeweiligen Ergänzungs- bzw. Aufstockungsstichproben	A	nicht zulässig
Wiederholt befragte Fälle		
Mit Interview im Vorjahr		
selbe Einheit wie im Vorjahr befragt	B	C
andere Einheit als im Vorjahr befragt	D	nicht zulässig
ohne Interview im Vorjahr	E	nicht zulässig

	Kennbuchstaben	
	<i>mit</i> sozialversicherungspflichtig	<i>ohne</i> Beschäftigte am jeweiligen Stichtag (30.06.)
Fälle ohne Interview in der laufenden Welle		
Ausfall, in Zukunft wieder befragbar	H	
Erstbefragte ohne endgültigen Ausfall	I	
Fälle aus früheren Aufstockungen, die nicht mehr weiter befragt werden	W	
Ausfall, nicht mehr befragbar	X	
Betrieb erloschen (lt. Feldergebnis, Editing bzw. BA-Datei)		
in der laufenden Welle	Y	
bereits früher	Z	

Für Querschnittsauswertungen stehen alle Fälle mit den Kennbuchstaben A, B, D und E aus der laufenden Welle zur Verfügung (vgl. Kapitel 5.2).

Die Definition von Panelfällen (vgl. Kapitel 5.3) umfasst alle Fälle mit den Kennzeichen B, C oder Y, Z aus der laufenden Welle, die in der unmittelbar vorangehenden Welle bereits ein Panelfall waren, sowie alle „neuen Betriebsnummern“ unter den mit A gekennzeichneten Fällen aus der laufenden Welle.

5.2 Querschnittsauswertung

Als Querschnittsfälle werden alle Fälle bezeichnet, die am 30. Juni des Vorjahres mindestens eine oder einen sozialversicherungspflichtig Beschäftigten hatten und im aktuellen Befragungsjahr ein gültiges Interview gegeben haben. Wegen des disproportionalen Stichprobenansatzes müssen deskriptive Auswertungen gewichtet durchgeführt werden. Bei einer disproportionalen Stichprobenanlage führen Auswertungen mit ungewichteten Daten zu nicht-repräsentativen Ergebnissen.

Die Gewichtung ist aufgrund des in Bezug auf Betriebsgröße, Branche und Bundesland disproportionalen Stichprobenansatzes sowie zum Ausgleich möglicher Differenzen zwischen Ist- und Sollstärke der einzelnen Schichtungszellen nötig (vgl. hierzu auch Abschnitt 8.1 im Anhang).

Die Gewichtung des IAB-Betriebspansels erfolgt grundsätzlich in Form einer Hochrechnung auf die Grundgesamtheit. Dies gilt für die Anzahl der Betriebe in der Grundgesamtheit, also für Gesamt-, aber auch für Ost- und Westdeutschland und für die einzelnen Bundesländer sowie für die Betriebe des Verarbeitenden Gewerbes in Ostdeutschland. Die Soll-Strukturen ergeben sich aus der Betriebsdatei der BA. Maßgeblich ist die Struktur (Verteilung der Betriebe) zum Zeitpunkt der Stichprobenziehung für die jeweilige Welle (also zum 30. Juni des jeweiligen Vorjahres). Durch eine solche Hochrechnung werden sowohl die Disproportionalitäten als auch die unterschiedlichen Ausschöpfungsraten entlang der Schichtungszellen in einem Schritt ausgeglichen.

Die gewichtete Stichprobe des IAB-Betriebspansels ist betriebsproportional. Sie spiegelt also die Verteilung der Betriebe über die Zellen der Schichtungsmatrix wider. Ihre Struktur unterscheidet sich damit von zahlreichen anderen Betriebsbefragungen, bei denen die Überrepräsentation der Großbetriebe nicht korrigiert wird (und die damit beschäftigten- oder umsatzproportionale, nicht jedoch betriebsproportionale Ergebnisse liefern). Mit Hilfe der gewichteten Daten des IAB-Betriebspansels sind jedoch nicht nur betriebsproportionale, sondern auch beschäftigtenproportionale Auswertungen möglich. Im Rahmen der Querschnittsgewichtung wird

außerdem darauf geachtet, dass die aus der gewichteten Stichprobe hochgerechneten Beschäftigtenzahlen (sozialversicherungspflichtig Beschäftigte zum 30. Juni des Vorjahres) auf der Ebene der Bundesländer den Soll-Vorgaben aus der Beschäftigtenstatistik der BA entsprechen. Für beschäftigtenproportionale Auswertungen ist die gewichtete Zahl der Beschäftigten aus den Betrieben, auf die das betreffende Merkmal zutrifft, ins Verhältnis zur Gesamtzahl der Beschäftigten zu setzen.

Ergebnis der Gewichtung ist ein integrierter betriebs- und beschäftigtenproportionaler Gewichtungsfaktor. Eine betriebsproportionale Gewichtung spiegelt die Verteilung der Betriebe über die Zellen der Schichtungsmatrix wider. Damit sind repräsentative Aussagen darüber möglich, auf wie viel Prozent der Betriebe in Deutschland eine bestimmte Aussage (z. B. Betrieb hat einen Betriebsrat) zutrifft. Die beschäftigtenproportionale Gewichtung spiegelt die Verteilung der Beschäftigten über die Bundesländer wider. Damit sind repräsentative Aussagen darüber möglich, wie viel Prozent der Beschäftigten in Deutschland in Betrieben arbeiten, auf die eine bestimmte Aussage (z. B. Beschäftigte arbeiten in einem Betrieb mit einem Betriebsrat) zutrifft.

Für Querschnittsauswertungen ist der Querschnittsgewichtungsfaktor (Variable „hrYYYYq“, wobei YYYY für das Jahr der jeweiligen Befragungswelle steht) zu nutzen. Querschnittsfälle sind in der Variablen „querYYYY“ gekennzeichnet.

Anwendungsbeispiele

Es gibt zwei Möglichkeiten gewichtete deskriptive Auswertungen zu erstellen. Zum einen kann mit dem im Datensatz vorgegebenen Gewichtungsfaktor „hrYYYYq“ der Anteil der Betriebe mit einer bestimmten Eigenschaft ausgezählt werden. Zum anderen muss, um Aussagen zu Beschäftigtenanteilen zu treffen, ein eigener Hochrechnungsfaktor berechnet werden. Im Folgenden sollen die beiden Gewichtungsmöglichkeiten anhand von Beispielen in Stata erläutert werden.

Gewichtung von betrieblichen Anteilen

Beispiel: Wie hoch ist der Anteil der Betriebe, die tarifgebunden sind?

Hierfür kann der im Datensatz vorhandene Hochrechnungsfaktor direkt verwendet werden.

```
tab tarifbindung // Fallzahlen ungewichtet  
tab tarifbindung [iw=hrYYYYq]
```

Gewichtung von Beschäftigtenanteilen

Beispiel 1: Wie viele Beschäftigte arbeiten in Betrieben, die tarifgebunden sind?

Hierfür muss ein eigener Gewichtungsfaktor berechnet werden, indem die Anzahl der Beschäftigten mit dem vorgegebenen Gewichtungsfaktor multipliziert wird.

```
generate N = bf03ges24 // für Welle 2024  
generate besfrac=N* hrYYYYq  
tab tarifbindung // Fallzahlen ungewichtet  
tab tarifbindung [iw=besfrac]
```

Dieser Gewichtungsfaktor ist geeignet für die Gewichtung von allen Anteilen, bei denen Aussagen über die gesamten Beschäftigten gemacht werden sollen.

Beispiel 2: Wie hoch ist der Anteil der Beschäftigten, die in Betrieben mit Betriebsrat (works council) arbeiten?

```
generate N = bf03ges24 // für Welle 2024
generate besfrac=N* hrYYYYq
tab works council // Fallzahlen ungewichtet
tab works council [iw=besfrac]
```

Dieser Gewichtungsfaktor kann dann genauso wie der betriebsproportionale Gewichtungsfaktor eingesetzt werden.

Beispiel 3: Wie hoch ist der Anteil der Frauen (frauant) an allen Beschäftigten (N) in West- und Ostdeutschland (west)?

```
generate N = bf03ges24 // für Welle 2024
generate N_women = bf06gesfr // für Welle 2024
generate west = wo2024 // für Welle 2024
generate frauant=N_women/N
bysort west: sum frauant // Fallzahlen ungewichtet
bysort west: sum frauant [iw=besfrac]
```

5.3 Längsschnittauswertung

Längsschnitt- bzw. Panelauswertungen ermöglichen es, betriebsindividuelle Entwicklungen über einen längeren Zeitraum nachzuvollziehen. Das IAB-Betriebspanel bietet aufgrund seiner großen Nettostichprobe sowie seiner langen Laufzeit vielfältige Möglichkeiten für solche Auswertungen.

Im IAB-Betriebspanel sind folgende Längsschnitte definiert, die sich auf unterschiedlich lange Zeiträume beziehen:

- Panelfälle 1993 – 2006 (nur für Westdeutschland)
- Panelfälle 1996 – 2006
- Panelfälle 2000 – 2012
- Panelfälle 2003 – 2016
- Panelfälle 2007 – 2017
- Panelfälle 2009 – 2020
- Panelfälle 2012 – 2024
- Panelfälle 2016 – 2024
- Panelfälle 2020 – 2024

Mit diesen Typen von Panelfällen lassen sich Längsschnittauswertungen vornehmen. Die zu untersuchende Zeiträume müssen nicht identisch mit den für die Panelfall-Definition gewählten Zeiträumen sein, sie müssen aber innerhalb des jeweiligen Zeitraums liegen.

Dazu einige Beispiele:

- Mit den Panelfällen 1996–2006 sind z. B. auch Auswertungen für den Zeitraum 1998–2001 oder 1997–2000 möglich.

- Will man einen Längsschnitt von 1995 bis 1998 analysieren, dann muss man auf die Panelfälle 1993–2006 zurückgreifen. Für Längsschnittanalysen von 1998 bis 2001 sollten dagegen – wegen höherer Fallzahlen – besser die Panelfälle 1996–2005 herangezogen werden.

Die Panelfallddefinition umfasst grundsätzlich alle Betriebe, die im Vorjahr ein Panelfall waren – entweder mit einem Interview oder als erloschener Betrieb – und alle Betriebe, die zwischen dem Stichtag des Vorvorjahres und dem Stichtag des Vorjahres neugegründet wurden.

Die Panelfälle, die aktuell noch weitergeführt werden, sind im Datensatz des Jahres 2024 wie folgt identifizierbar:

- Panelfälle 2012–2024: pan12_24 == 1
- Panelfälle 2016–2024: pan16_24 == 1
- Panelfälle 2020–2024: pan20_24 == 1

Seit der Welle 2023 werden keine Panel-Gewichtungsfaktoren mehr bereitgestellt. Die Panel-Gewichte bis einschließlich der Welle 2022 stehen weiterhin in den Daten bis einschließlich der Welle 2022 zur Verfügung (für Informationen zu den Panel-Gewichtungsfaktoren siehe Bächmann et al. 2023). Alternativ zu der Verwendung von Gewichten kann in Regressionsmodellen auch flexibel für die Stratifizierungsmerkmale kontrolliert werden (siehe Bossler et al. 2018).

6 Verwendungshinweise und Datennutzung

6.1 Hilfe zum Aufbau eines Panel-Datensatzes

Das FDZ bietet ein Stata do-file an, welches alle Wellen von 1993 bis zur jeweils aktuellsten Welle in einem Paneldatensatz zusammenführt. Das zugrunde liegende Verfahren wird im FDZ-Methodenreport 12/2017 beschrieben (vgl. Umkehrer 2017).

Das do-file bereitet die einzelnen Querschnitte des IAB-Betriebspans als Panel-Datensatz auf. Es wird ein Datensatz erstellt, der für jeden gewählten Themenblock (Modul) ein konsistent aufbereitetes Merkmalsspektrum umfasst. Mit "konsistent" ist hier die Zusammenfassung der einzelnen Querschnittsvariablen zu einer Variable mit einheitlicher Kodierung gemeint. Die Bedeutung der Inhalte der einzelnen Querschnittsvariablen kann nach wie vor über die Zeit variieren, z. B. aufgrund von Änderungen der befragten Gruppe, der Fragestellung und/oder der ursprünglichen Kodierung der Variablen.

Die Variablennamen haben folgende Bedeutung:

- _b: zeigt an, dass es sich um eine binäre Variable handelt (i. d. R. ja/nein)
- _d: zeigt an, dass es sich um eine diskrete Variable handelt
- _c: zeigt an, dass es sich um eine stetige Variable handelt

6.2 Sensible Merkmale und Zuspielvariablen

Bestimmte, sogenannte sensible Merkmale, die eine De-Anonymisierung von Personen oder Betrieben erleichtern könnten, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf einen Datenzugang explizit und separat für jedes sensible Merkmal begründet wird. Es ist zudem zu beachten, dass die sensiblen Merkmale in der Regel bereits in vergrößerter Form in den Daten enthalten sind. Die sensiblen Merkmale werden nur dann weitergegeben, wenn die Information im vergrößerten Zustand zur Erreichung des Forschungsziels nicht ausreichend ist. Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale sind:

Sensible Merkmale beim IAB-Betriebspanel

- Kreiskennziffer (kkz)
- Wirtschaftszweigklassifikation 1993 (wz93) (3- oder 5-Steller)
- Wirtschaftszweigklassifikation 2003 (wz03) (5-Steller)
- Wirtschaftszweigklassifikation 2008 (wz08) (5-Steller)
- Gemeindetyp (bik) (https://bik-gmbh.de/wp-content/uploads/Karte_BIK-Gemeindegroessenklassen_10_BRD.pdf)

Sensible Merkmale im Betriebs-Historik-Panel (BHP)

Zusätzlich kann folgendes sensible Merkmal aus dem Betriebs-Historik-Panel (BHP) beantragt werden:

- Zeitkonsistente Klassifikation der Wirtschaftszweige (3-Steller) (w08_3_gen group_w08_3)

BHP-Zuspielmerkmale

Für das IAB-Betriebspanel stehen weitere Variablen aus dem BHP zur Verfügung; siehe <https://fdz.iab.de/betriebsdaten/betriebs-historik-panel-bhp-version-7524-v1/>. Beim BHP handelt es sich um administrative Betriebsdaten, die den Befragungsdaten des IAB-Betriebspansels auf Ebene des Betriebes (Variable „idnum“) und dem Jahr zugespielt werden können.

Dabei gilt zu beachten, dass nicht für jeden Betrieb in jedem Jahr eine Information aus dem BHP verfügbar ist. Dies liegt einerseits daran, dass BHP-Zuspielmerkmale nur dann zur Verfügung gestellt werden können, wenn eine Zustimmung zur Verknüpfung von Befragungsdaten mit administrativen Daten vorliegt.⁵ Andererseits können sich Betriebsnummern, die ursprünglich einmal für die Befragung gezogen wurden, über die Zeit ändern, sodass sie in den administrativen Daten nicht mehr gefunden werden können, u. a. weil es bei Wiederholungsbefragungen keinen Betriebsnummernabgleich gibt. Schließlich ist zu beachten, dass die Angaben aus den administrativen Daten grundsätzlich von den Daten aus der Befragung zum IAB-Betriebspanel abweichen können.

Ab der Version IAB-BP 9323 wird standardmäßig eine sogenannte Basis-Betriebsdatei bereitgestellt, die grundlegende Informationen aus dem BHP enthält:

⁵ Der Anteil der Betriebe, für die eine Zustimmung zur Verknüpfung mit den administrativen Daten vorliegt, liegt bis zum Jahr 2009 bei über 98 Prozent, fällt dann aber auf 94 Prozent im Jahr 2024. Im Jahr 2020 ist der Anteil mit 90 Prozent am geringsten. Dieser Rückgang ist auf eine Änderung der Erhebung der Zuspielbereitschaft ab dem Jahr 2020 zurückzuführen, die in erster Linie Betriebe, die ab dem Jahr 2020 an der Befragung teilgenommen haben, betrifft.

Basismerkmale

- Betriebsidentifikator (idnum)
- Jahr (jahr)
- Jahr erstes Auftreten Betriebsnummer (grd_jahr)
- Jahr letztes Auftreten Betriebsnummer (lzt_jahr)
- Anzahl Beschäftigte gesamt (az_ges)
- Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)
- Anzahl geringfügig Beschäftigter (az_gf)
- Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Auf begründeten Antrag hin können aus dem BHP weitere Betriebsmerkmale in Variablenblöcken bereitgestellt werden. Zudem können die Erweiterungsmodule „Beschäftigungsströme“ (Eintritte/Austritte) und „Betriebsdynamik“ (Gründungen/Schließungen) mit gesonderter Begründung beantragt werden. Die Begründung der Notwendigkeit für das Erreichen des Forschungsziels hat dabei separat für jeden Variablenblock bzw. jedes Erweiterungsmodul zu erfolgen. Die Variablenblöcke bzw. Erweiterungsmodule sind:

BHP-Variablenblöcke in „Jahresscheiben“

- Struktur der Beschäftigten allgemein (az_f az_reg az_azubi az_atz 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_ignat)
- 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_bf_agr az_bf_emb az_bf_edi az_bf_evb az_bf_qmb az_bf_qdi az_bf_qvb az_bf_tec az_bf_semi az_bf_ing az_bf_prof az_bf_man)
- Struktur der Beschäftigten nach Tätigkeitsniveau (az_niv1 az_niv2 az_niv3 az_niv4)
- Gehaltsstruktur der Vollzeitbeschäftigten (az_zens 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 te_imp_sum)

BHP-Erweiterungsmodule

- Beschäftigtenströme (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_f_vz ein_f_tz ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_bf_agr ein_bf_emb ein_bf_edi ein_bf_evb ein_bf_qmb ein_bf_qdi ein_bf_qvb ein_bf_tec ein_bf_semi ein_bf_ing ein_bf_prof ein_bf_man 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_f_vz aus_f_tz aus_reg_vz aus_temp aus_temp_f aus_bw aus_bw_f aus_bf_agr aus_bf_emb aus_bf_edi aus_bf_evb aus_bf_qmb aus_bf_qdi aus_bf_qvb aus_bf_tec aus_bf_semi aus_bf_ing aus_bf_prof aus_bf_man 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)
- Betriebsdynamik (eintritt besch inflow besch_vor status_vor austritt besch outflow besch_nach status_nach)

Nähere Informationen zum BHP finden sich im FDZ-Datenreport 13/2025 (Ganzer et al., 2025).

Dateinamen

Die Merkmale aus dem BHP sind in separaten Dateien abgelegt und können über die Variablen „idnum“ und ggf. „jahr“ an die Befragungsdaten gespielt werden. Die Dateien für die Version IAB-BP 9324 v1 sind wie folgt benannt:

- Basis-Betriebsdatei:

IABBP_9324_v1_bhp_basis_v1.dta

- Thematische Variablenblöcke in Jahresscheiben (YYYY=1992...2024):

IABBP_9324_v1_bhp_v1_YYYY.dta

- Erweiterungsmodul Beschäftigtenströme:

IABBP_9324_v1_bhp_inflow_v1.dta;

IABBP_9324_v1_bhp_outflow_v1.dta

- Erweiterungsmodul Betriebsdynamik:

IABBP_9324_v1_bhp_entry_v1.dta;

IABBP_9324_v1_bhp_exit_v1.dta

- Sensibles Merkmal Konsistente Wirtschaftszweige:

IABBP_9324_v1_wgen_w08_v1.dta

AKM-Effekte

Für das IAB-Betriebspanel stehen betriebsspezifische Effekte (AKM-Effekte) für die Jahre 1985 bis 2023 zur Verfügung. Weitere Informationen finden sich im FDZ-Methodenreport 03/2025 (Lochner & Wolter, 2025). Diese können als Ergänzung zum Projekt formlos per E-Mail an iab.fdz@iab.de beantragt werden.⁶

⁶ Die entsprechenden Datei- und Variablennamen lauten: IABBP_9324_v1_aki_estab.dta (feff_1985_1992 feff_1993_2000 feff_2001_2008 feff_2009_2016 feff_2017_2023).

6.3 Datenzugang

Das IAB-Betriebspanel kann im Rahmen von Gastaufenthalten am Forschungsdatenzentrum der BA im IAB (FDZ) und anschließender Datenfernverarbeitung ausgewertet werden. Für die Datennutzung ist ein Antrag beim FDZ einzureichen. Details zur Beantragung des Datenzugangs können der FDZ-Website (<https://fdz.iab.de/>) entnommen werden.

Zudem steht das IAB-Betriebspanel auch verknüpft mit Personendaten des IAB als Linked-Employer-Employee-Datensatz des IAB (LIAB; siehe <https://fdz.iab.de/unsere-datenprodukte/integrierte-betriebs-und-personendaten/liab/>) oder als Teil des Linked Personnel Panel (LPP; siehe <https://fdz.iab.de/unsere-datenprodukte/integrierte-betriebs-und-personendaten/lpp/>) zur Verfügung.

7 Literatur

- Bächmann, A.-C.; Bellmann, L.; Gensicke, M., Kohaut, S.; Möller, I.; Schwengler, B.; Tschersich, N.; Umkehrer, M. (2023): IAB-Betriebspanel (IAB-BP) 1993-2022. FDZ-Datenreport, 16/2023 (de), Nürnberg.
- Bossler, M., Geis, G., Stegmaier, J. (2018): Comparing Survey data with an official administrative population: assessing sample-selectivity in the IAB-Establishment Panel. *Qual Quant* 52, pp. 899-920.
- Cassel, Clare; Särndal, Carl-Erik; Wretman, Jan Hakan (1977): Foundations of Inference in Survey Sampling. New York.
- Deville, Jean-Claude; Särndal, Carl-Erik (1992): Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Vol. 87, No. 418, pp.376 – 382.
- Deville, Jean-Claude, Särndal, Carl-Erik, Sautory, Olivier (1993): Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 1013 – 1020.
- Ganzer, Andreas, Alexandra Schmucker, Jens Stegmaier & Stefanie Wolter (2022): Betriebs-Historik-Panel 1975-2021. (FDZ-Datenreport 12/2022 (de)), Nürnberg, 116 S. DOI:10.5164/IAB.FDZD.2212.de.v1
- Ganzer, Andreas; Schmucker, Alexandra; Wolter, Stefanie (2025): Betriebs-Historik-Panel 1975–2024. FDZ-Datenreport, 13/2025 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2513.de.v1
- Hethay, Tanja; Schmieder, Johannes F. (2010): Using worker flows in the analysis of establishment turnover - Evidence from German administrative data. FDZ-Methodenreport, 06/2010 (en), Nürnberg.
- Hethay-Maier, Tanja; Schmieder, Johannes F. (2013): Does the use of worker flows improve the analysis of establishment turnover? Evidence from German administrative data. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 133, H. 4, S. 477–510.
- Lochner, Benjamin; Wolter, Stefanie (2025): AKM effects for German labour market data 1985-2023. FDZ-Methodenreport 03/2025 (en), Nuremberg. DOI: 10.5164/IAB.FDZM.2503.en.v1
- Umkehrer, Matthias (2017): Zusammenführen der Wellen des IAB-Betriebspanels. Ein Do-File für die grundlegende Aufbereitung eines Paneldatensatzes in Stata. FDZ-Methodenreport 23/2017.

8 Anhang

8.1 Verfahren der Querschnittsgewichtung

Ziel der auf einer Zufallsstichprobe aufbauenden, also design-basierten Inferenz, ist es, für ein interessierendes Zielmerkmal y bestimmte Parameter der Grundgesamtheit zu schätzen (vgl. Deville et al. 1993). Wichtige Parameter sind die Summe oder der Durchschnitt dieses Zielmerkmals in der Grundgesamtheit. Wird zur Schätzung solcher Parameter aus der Grundgesamtheit U ($= 1, \dots, k, \dots, N$) per Zufallsmechanismus eine Stichprobe s ($= 1, \dots, k, \dots, n$) mit auf Basis des Stichprobendesigns strikt positiven Auswahlwahrscheinlichkeiten für jedes Element ($\pi_k = \Pr(k \in s) > 0, \pi_{kl} = \Pr(k \& l \in s) > 0$) gezogen, dann ergibt sich das Design-Gewicht d_k eines Stichprobenelements k als die Inverse seiner Auswahlwahrscheinlichkeit π_k , also $d_k = \pi_k^{-1}$. Die zweite Anforderung der strikt positiven Auswahlwahrscheinlichkeit ist erforderlich, um die Varianz der Schätzer ermitteln zu können (vgl. Cassel et al. 1977). Der Horvitz-Thompson-Schätzer stellt dann die design-gewichtete Schätzung des Parameters dar, also z. B. für die Summe \hat{y} eines Merkmals y der Wert $\hat{y} = \sum s d_k y_k$.

Um neben design-bedingten unterschiedlichen Auswahlwahrscheinlichkeiten auch das Ausfallgeschehen zu berücksichtigen und um die Varianz der Schätzer zu verringern, erfolgt beim IAB-Betriebspanel unter Hinzuziehung bestimmter Hilfsmerkmale eine weitere Anpassung der Gewichtungsfaktoren anhand generalisierter Regressionsmodelle, kurz GREG. Ausgehend von einem Horvitz-Thompson-Schätzer besteht das Ziel bei einer Kalibrierung per GREG darin, die Design-Gewichte vor dem Hintergrund zusätzlich zur Verfügung stehender Informationen über die Summe (oder den Durchschnitt) von Hilfsmerkmalen x so anzupassen und in neue Gewichtungsfaktoren w_k zu überführen, dass die Stichprobe nach der Gewichtung die Summe (oder den Durchschnitt) dieser Hilfsmerkmale x abbildet, d. h. also $\sum s w_k x_k = \sum U x_k$. Gleichzeitig sollten die ursprünglichen Design-Gewichte d_k aber möglichst wenig verändert werden: "Our objective is to derive new weights that modify as little as possible the original sampling weights ($d_k = \pi_k^{-1}$), which have the desirable property of yielding unbiased estimates" (Deville & Särndal 1992: 376).

Die Gewichte w_k sind die Lösung eines Minimierungsproblems unter Nebenbedingungen: Bezeichnet G (w_k/d_k) eine Funktion, die die Distanz zwischen d_k und w_k abbildet, dann besteht das Optimierungsproblem darin, die im Folgenden dargestellte Funktion hinsichtlich w_k zu minimieren, wobei λ den Vektor der Langrange-Multiplikatoren darstellt.

Optimierungsproblem im Rahmen einer Generalisierten Regression⁷

$$\sum_s d_k G\left(\frac{w_k}{d_k}\right) - \lambda' \left(\sum_s w_k x_k - \sum_U x_k \right)$$

mit:

⁷ Hier und in den folgenden Gleichungen bezeichnen Buchstaben in Fettschrift Vektoren, in Normalschrift Skalare.

w_k	= finaler Gewichtungsfaktor
x_k	= Hilfsmerkmale der Elemente der Stichprobe/Grundgesamtheit
d_k	= Design-Gewicht
S	= Stichprobe
U	= Grundgesamtheit
G	= Distanzfunktion
λ	= Langrange Multiplikator

Deville et al. (1993) beschreiben mehrere Distanzfunktionen. Die von Deville et al. (1993: 1014) als lineare Methode bezeichnete Variante⁸ führt in der Anwendung auf die Betriebspanelstichprobe zu den besten Anpassungen mit leicht größeren Faktorenspannbreiten im Vergleich zu der von ihnen als Logit-Methode bezeichneten Vorgehensweise. Aufgrund der besseren Anpassungen wurde deshalb die lineare Methode gewählt. Verwendet man dieses Verfahren, dann ergibt sich für die Schätzung der Summe eines Merkmals der „generalized regression estimator“ (GREG, vgl. Deville, Särndal 1993: 1014):

Schätzer der Summe des Merkmals y auf Basis der generalisierten Regression (GREG)

$$\hat{y}_{reg} = \sum_S w_k y_k = \hat{y}_\pi + (\hat{x} - \hat{x}_\pi)' \hat{B}_s$$

Dabei bezeichnen die mit dem Subskript π bezeichneten Schätzer y und x die Horvitz-Thompson-Schätzer der Summen des Merkmals y und des x -Vektors der Hilfsmerkmale, \hat{x} den Vektor der für die Grundgesamtheit bekannten Summen der x -Merkmale und \hat{B}_s den Vektor der auf Basis der Stichprobe geschätzten Regressionsparameter von y auf die x -Merkmale.

Die GREG-Gewichte lassen sich auf stetige Variablen bzw. auf deren Summen direkt anpassen. Das bedeutet für die Gewichtung der Querschnittsstichprobe, dass simultan die Stichprobe an die Verteilung der Betriebe und an die Verteilung der Beschäftigten angepasst werden kann.

Im Rahmen der Gewichtung wird geprüft, ob die Faktoren in einzelnen Zellen zu hoch oder zu niedrig werden oder die Zellenbesetzung gleich Null ist. In diesen Fällen werden Zusammenfassungen mit einer geeigneten Nachbarzelle vorgenommen.

⁸ Hierbei ist allerdings die Distanzfunktion G selbst nicht linear, sondern quadratisch. Relevant für die Bezeichnung als lineare Methode ist, dass F , die Inverse der ersten Ableitung von G , linear ist.

Impressum

FDZ-Datenreport 01|2026

Veröffentlichungsdatum

11. Februar 2026

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/reportes/2026/DR_01-26.pdf

Dokumentation Version

IAB-BP 9324_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2601.de.v1

Datensatz Version

IAB-BP 9324 v1, DOI: 10.5164/IAB.IABB9324.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

Lisa Bellmann

Telefon: 0911 179-6122

E-Mail: lisa.bellmann@iab.de