



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

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

11|2025 DE Codebuch und Dokumentation des Websurveys im „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) Datenreport Websurvey 2024

Christian Dickmann, Birgit Jesske, Sabrina Torregroza (infas Institut für angewandte Sozialwissenschaft GmbH) - Jonas Beste, Sandra Dummert, Corinna Frodermann, Sonja Malich, Valentina Prospero, Claudia Wenzig, Mark Trappmann, Stefan Zins, Sebastian Bähr, Matthias Collischon, Stefanie Gundert, Marcel Müller, Jens Stegmaier, Anja Wunder (Institut für Arbeitsmarkt- und Berufsforschung (IAB))



Bundesagentur für Arbeit

Codebuch und Dokumentation des Websurveys im „Panel Arbeitsmarkt und soziale Sicherung“ (PASS)

Datenreport Websurvey 2024

Christian Dickmann, Birgit Jesske, Sabrina Torregroza (infas Institut für angewandte Sozialwissenschaft GmbH) - Jonas Beste, Sandra Dummert, Corinna Frodermann, Sonja Malich, Valentina Prospero, Claudia Wenzig, Mark Trappmann, Stefan Zins, Sebastian Bähr, Matthias Collischon, Stefanie Gundert, Marcel Müller, Jens Stegmaier, Anja Wunder (Institut für Arbeitsmarkt- und Berufsforschung (IAB))

Inhaltsverzeichnis

1	Einleitung	4
1.1	Ziele und Fragestellung des Websurveys	4
1.2	Instrumente und Befragungsprogramm	4
2	Stichprobenumfang	5
3	Datensatzstruktur	8
4	Generierte Variablen	9
4.1	Vercodung offener Restkategorien	9
4.2	Einfache Generierungen	9
5	Datenaufbereitung	12
5.1	Strukturprüfungen und entfernte Interviews	12
5.2	Filterbereinigung und vergebene Missing Codes	15
5.3	Plausibilitätsprüfungen	16
5.4	Anonymisierung	16
6	Gewichtung Websurvey 2024	17

Tabellenverzeichnis

Tabelle 1:	Responserate nach Teilstichprobe	6
Tabelle 2:	Responserate nach Erhebungssprache der Panelhaupterhebung Welle 18	7
Tabelle 3:	Vercodung offener Angaben	9
Tabelle 4:	Einfache Generierungen (alphabetische Reihenfolge)	10
Tabelle 5:	Übersicht über die Schritte der Datenaufbereitung	12
Tabelle 6:	Übersicht über die vergebenen Missing Codes	15
Tabelle 7:	Übersicht über die anonymisierten Variablen	16
Tabelle 8:	Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle	18
Tabelle 9:	Logit-Modell zur Teilnahme am Websurvey	19
Tabelle 10:	Kenngrößen der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten für die Nettostichprobe	20
Tabelle 11:	Kenngrößen der Verteilung der Gewichte (Gesamtstichprobe, Personen)	21

Zusammenfassung

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. Dieser Datenreport dokumentiert die Aufbereitung des Websurveys des PASS 2024.

Abstract

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of re-ports 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. This data report documents the data preparation of the PASS websurvey 2024.

Datenverfügbarkeit

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich.

Nähere Informationen dazu auf der Internetseite <https://fdz.iab.de/>.

1 Einleitung

1.1 Ziele und Fragestellung des Websurveys

Mit dem Panel „Arbeitsmarkt und soziale Sicherung“ (PASS) hat das Institut für Arbeitsmarkt- und Berufsforschung (IAB) 2006 einen Datensatz für die Arbeitsmarkt-, Sozialstaats- und Armutsforschung in Deutschland aufgebaut, der eine neue empirische Grundlage für Wissenschaft und Politikberatung schafft.

Um einen Einstieg in die selbstadministrierte Erhebung im Webmodus zu vollziehen, wurde im Anschluss an Welle 16 im Jahr 2022 erstmalig ein Websurvey im Rahmen des „Panel Arbeitsmarkt und soziale Sicherung“ (PASS) durchgeführt. Das Ziel war dabei insbesondere die Erprobung optimaler Abläufe bei der Durchführung von Webbefragungen in der PASS-Stichprobe. Während für Welle 17 kein Websurvey durchgeführt wurde, erfolgte ein zweiter Websurvey im Anschluss an Welle 18. Im Mittelpunkt stand dabei die Erhebung zusätzlicher Fragestellungen, die im Rahmen der Hauptbefragung aus Zeitgründen nicht erhoben werden konnten: eine Vignettenstudie zur Qualität von Arbeit und Transferleistungsbezug sowie ein umfassendes Fragemodul zum Lebensstandard und zur Ressourcenallokation innerhalb von Haushalten.

Als zusätzliche Erhebung im Nachgang zu den jährlichen Panelerhebungswellen im PASS ist der Websurvey grundsätzlich an die dort erfolgreiche Realisierung eines Personeninterviews gekoppelt. Damit wird für jede Welle des Websurveys jeweils eine neue Stichprobe bestimmt. Die panelbereiten Teilnehmerinnen und Teilnehmer der Hauptwelle werden einzeln kontaktiert und die Durchführung der Erhebung erfolgte unabhängig von den anderen Haushaltsmitgliedern. Basis der Stichprobe waren für Welle 18 alle Teilnehmerinnen und Teilnehmer der Hauptstichprobe, für die ein gültig realisiertes Interview im PASS Welle 18 vorlag (11.235 Personen). Eine ausführliche Beschreibung des Designs für den Websurvey und das Vorgehen für die Erhebung findet sich im Methodenbericht PASS Websurvey Haupterhebung 2024 (Jesske et al. 2025). Dieser Datenreport gibt einen Überblick über den Websurvey Welle 18, für den vom 23. Oktober bis zum 09. Dezember 2024 5.186 Personen befragt wurden.

In Kapitel 1 findet sich ein Überblick über die Ziele und Fragestellungen des Websurveys Welle 18 mit einer Kurzbeschreibung der Instrumente und des Befragungsprogramms in Kapitel 1.2. Der Umfang der Stichprobe des Websurveys wird in Kapitel 2 dargelegt. In den darauffolgenden Kapiteln geht es um die eigentlichen Daten und deren Aufbereitung. In Kapitel 3 wird die Datensatzstruktur und in Kapitel 4 werden die generierten Variablen beschrieben. Weiterhin werden die Schritte der Datenaufbereitung und die dabei getroffenen Entscheidungen in Kapitel 5 erläutert. In Kapitel 6 wird dann das Gewichtungungsverfahren vorgestellt.

1.2 Instrumente und Befragungsprogramm

Im PASS Websurvey wurden die Zielpersonen – im Unterschied zur Panelhaupterhebung – ausschließlich auf Ebene der Personen kontaktiert und Interviews realisiert. Die Gruppe der Personen, die in der Panelhaupterhebung den Seniorenfragebogen beantwortet haben, erhielten auch im Websurvey eine gekürzte Variante des Fragebogens.

Die insgesamt 29 Fragen im Websurvey verteilen sich auf die folgenden Module:

- Soziodemografische Informationen
- Qualität der Arbeit und Transferleistungen
- Lebensstandard im Allgemeinen und im Haushalt
- Ressourcenverwaltung und Entscheidungsmacht
- Rolleneinstellungen
- Einstellungen zur Kindergrundsicherung
- Einstellungen Lebensstandard

Das Modul „Qualität der Arbeit und Transferleistungen“ war als Vignettenstudie konzipiert. In diesem Modul wurden den Zielpersonen fünf unterschiedliche fiktive Stellenangebote in Tabellenform vorgelegt. Diesen Stellenangeboten wurde eine Arbeitslosenunterstützung bzw. das bedingungslose Grundeinkommen gegenübergestellt. Die Befragten sollten für jedes Stellenangebot jeweils einschätzen, wie wahrscheinlich sie die Stelle annehmen würden, wenn sie ihnen angeboten würde. Der zweite Teil der Webbefragung umfasst Module zum Lebensstandard und Ressourcenallokation innerhalb von Haushalten. Grundsätzlich wurden im Websurvey die Sprachen Arabisch, Russisch und Ukrainisch berücksichtigt. Die einzige Ausnahme bildete das Vignettenmodul, das nur in Deutsch erhoben wurde.

Einige Fragen aus dem Modul soziodemografische Informationen (Geschlecht und Geburtsjahr) wurden auch für die Datenprüfungen herangezogen. Abweichungen an dieser Stelle deuteten möglicherweise darauf hin, dass der Online-Fragebogen von einer anderen Person beantwortet wurde. Im Rahmen der Datenprüfung wurden solche Fälle identifiziert und die Interviews als ungültig deklariert (vgl. Kapitel 5).

2 Stichprobenumfang

Ausgangsbasis für den Stichprobenumfang im PASS Websurvey 2024 bilden die befragten Personen der Panelhaupterhebung aus der Welle 18. Dabei muss berücksichtigt werden, dass nur solche Personen für den Websurvey zur Verfügung stehen, für die die Panelbereitschaft¹ vorliegt.

In der Panelhaupterhebung wurden 11.265 Interviews mit panelbereiten Personen realisiert.

Diese wurden zwischen Feldende der Panelhaupterhebung (Mitte September 2024) und Feldstart des Websurveys (Ende Oktober 2024) um kurzfristige Ausfälle (zum Beispiel Verweigerungen, Todesfälle) reduziert. Die bereinigte Personenzahl (n=11.235) bildet den Stichprobenumfang für die Haupterhebung des PASS Websurvey (Bruttoeinsatzstichprobe).

Insgesamt haben am Websurvey Welle 18 5.416 Personen teilgenommen. 5.186 Online-Fragebögen wurden als vollständig auswertbar deklariert.

Tabelle 1 weist die Responseraten differenziert nach der Teilstichprobe sowie die Brutto- und Nettofallzahlen aus. Die Responserate insgesamt liegt bei 0,46 und liegt damit über dem Niveau der

¹ Die Panelbereitschaft wird durch den oder die Haushaltsbevollmächtigte erteilt und zählt damit für alle Haushaltsmitglieder. Panelbereite Haushalte haben der Speicherung ihrer Adresse zum Zweck von Wiederholungsbefragungen im Rahmen der Studie zugestimmt.

letzten Websurvey Welle. Die Responserate differenziert nach Teilstichprobe variiert von 0,16 (Teilstichprobe 32 Bestandsauffrischung (BA) (syrische/irakische HH)) Welle 17 bis 0,60 für die Teilstichprobe 30 (Bestandsauffrischung (EWO) Welle 17).

Tabelle 1: Responserate nach Teilstichprobe

Teilstichprobe	Bruttostichprobe	Nettostichprobe²	Responserate
	<i>absolut</i>	<i>absolut</i>	
Gesamt	11.235	5.186	0,46
1 BA-Stichprobe	742	259	0,35
2 Microm-Stichprobe	1.341	727	0,54
3 Zugangsstichprobe (BA) Welle 2	120	51	0,43
4 Zugangsstichprobe (BA) Welle 3	182	72	0,40
5 Zugangsstichprobe (BA) Welle 4	121	57	0,47
6 Bestandsauffrischung (EWO) Welle 5	480	254	0,53
7 Bestandsauffrischung (BA) Welle 5	241	92	0,38
8 Zugangsstichprobe (BA) Welle 5	128	58	0,45
9 Zugangsstichprobe (BA) Welle 6	196	100	0,51
10 Zugangsstichprobe (BA) Welle 7	191	92	0,48
11 Zugangsstichprobe (BA) Welle 8	194	92	0,47
12 Zugangsstichprobe (BA) Welle 9	212	93	0,44
13 Zugangsstichprobe (BA) Welle 10	136	81	0,60
14 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 10	83	39	0,47
15 Bestandsauffrischung (EWO) Welle 11	550	296	0,54
16 Zugangsstichprobe (BA) Welle 11	131	56	0,43
17 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 11	85	29	0,34
18 Zugangsstichprobe (BA) Welle 12	172	81	0,47
19 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 12	111	26	0,23
20 Zugangsstichprobe (BA) Welle 13	244	114	0,47
21 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 13	36	12	0,33
22 Zugangsstichprobe (BA) Welle 14	273	131	0,48
23 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 14	23	6	0,26
24 Bestandsauffrischung (BA) (syrische/irakische HH) Welle 14	86	26	0,30
25 Zugangsstichprobe (BA) Welle 15	578	282	0,49
26 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 15	34	15	0,44
27 Bestandsauffrischung (BA) (syrische/irakische HH) Welle 15	88	25	0,28
28 Zugangsstichprobe (BA) Welle 16	451	212	0,47
29 Zugangsstichprobe (BA) (syrische/irakische HH) Welle 16	76	20	0,26
30 Bestandsauffrischung (EWO) Welle 17	1.289	778	0,60
31 Zugangsstichprobe (BA) Welle 17	265	99	0,37

² Die Nettostichprobe enthält alle vollständig auswertbaren Online-Fragebögen

Teilstichprobe	Bruttostichprobe	Nettostichprobe²	Response rate
	<i>absolut</i>	<i>absolut</i>	
32 Zugangsstichprobe (BA) (syrische/ irakische HH) Welle 17	43	7	0,16
33 Zugangsstichprobe (BA) (ukrainische HH) Welle 17	592	225	0,38
34 Zugangsstichprobe (BA) Welle 18	441	139	0,32
35 Zugangsstichprobe (BA) (syrische/ irakische HH) Welle 18	124	27	0,22
36 Zugangsstichprobe (BA) (ukrainische HH) Welle 18	606	222	0,37
37 Bestandsauffrischung (EWO) Welle 17/ 2	570	291	0,51

Die Response rate nach Erhebungssprache zeigt, dass Zielpersonen, die in Welle 18 in einer der angebotenen Fremdsprachen³ teilgenommen haben, eine niedrigere Response rate aufweisen, wobei sie für Personen, die auf Russisch teilgenommen haben, am geringsten ausfällt.

Tabelle 2: Response rate nach Erhebungssprache der Panelhaupterhebung Welle 18

Erhebungssprache	Bruttostichprobe	Nettostichprobe	Response rate
	<i>absolut</i>	<i>absolut</i>	
Gesamt	11.235	5.186	0,46
Deutsch	9.355	4.549	0,49
Arabisch	1.065	390	0,37
Russisch	673	188	0,28
Ukrainisch	142	59	0,42

³Ukrainisch wurde im Websurvey der Welle 18 erstmalig angeboten, da diese Gruppe seit der Welle 17 in der Haupterhebung im PASS zusätzlich berücksichtigt wird.

3 Datensatzstruktur

Der Scientific Use File des Websurveys besteht aus drei Datensätzen. Sofern nicht in diesem Kapitel anders erläutert, wurden den Variablennamen ein kleines „w“ vorangestellt, um die Variable als aus dem Websurvey stammend zu kennzeichnen. Generierte Variablen werden in Kapitel 4.2 erläutert.

Im Datensatz *wPENDDAT* liegt für jede durchgeführte Personenbefragung eine Datenzeile vor. Dort enthalten sind die Fragen der in der Fragebogenvorlage dokumentierten Module *Soziodemografische Informationen* (ohne die Angaben zum Alter *wPHH0400* sowie zur verwandtschaftlichen Beziehung zu den anderen Haushaltsmitgliedern *wPHH0500/wPHH0501*), *Ressourcenverwaltung und Entscheidungsmacht*, *Rolleneinstellungen*, *Einstellungen zur Kindergrundsicherung* und *Einstellungen Lebensstandard*. Hinzu kommen die Identifikationsvariablen *pnr*, *hnr* und *uhnr*, sowie eine Wellenkennung und eine Stichprobenkennung. Bei den Identifikationsvariablen *pnr*, *hnr* und *uhnr* sowie bei der Wellenkennung *welle* und der Stichprobenkennung *sample* wurden die aus der Panelhaupterhebung bekannten Variablennamen verwendet. Die Variable *NoWelle17* kennzeichnet das Teilnahmeverhalten der Person in der zugehörigen CAI-Hauptbefragung der Welle 17 und wurde zur Steuerung des Moduls *Rolleneinstellungen* genutzt. Ein Teil der in der Befragung erhobenen Items ist auch Teil des *PENDDAT*-Datensatzes der Panelhaupterhebung. Die Variablen können sich hinsichtlich der Antwortkategorien grundsätzlich allerdings voneinander unterscheiden zwischen Websurvey und Panelhaupterhebung.

Der Datensatz *wVIGNETTEN* enthält die Angaben zum Modul *Qualität der Arbeit und Transferleistungen* für alle Personen, die den Eingangsfiler in dieses Modul erfüllt haben. Der Vignettendatensatz ist grundsätzlich so strukturiert, wie es von den Vignettendatensätzen der Wellen 5, 12 und 15 der Panelhaupterhebung bekannt ist. Für jede Person sind fünf Datenzeilen abgelegt, dies entspricht der Anzahl an Vignetten, die jeder befragten Person vorgelegt wurden. Darin enthalten sind neben der Identifikationsvariablen *pnr* und der Stichprobenkennung *sample* die Vignettennummer *wvignr*, die die Abfragereihenfolge der Vignetten angibt und die Vignetten-ID *wPVIG*, die angibt, welche der 4000 definierten Vignetten vorgelegt wurden. Die Variable *wPQV0200* gibt die Antwort der befragten Person auf die vorgelegte Vignette an, die anderen *wPQV**-Variablen dokumentieren die Dimensionen, anhand derer das Vignettenuniversum aufgespannt wurde. Das Modul *Qualität der Arbeit und Transferleistungen* wurde nur in deutschsprachigen Interviews vorgelegt.

Im Datensatz *wRESSOURCEN* sind platzsparend im Long-Format jene Angaben abgelegt, bei denen die befragte Person nicht nur zu sich selbst, sondern auch zu den anderen Haushaltsmitgliedern Auskunft gibt. Der Datensatz enthält für alle durchgeführten Befragungen so viele Zeilen, wie es der abgefragten Haushaltsgröße entspricht. Für jedes Haushaltsmitglied wurde eine Datenzeile abgelegt. Falls es sich entweder um einen Ein-Personen-Haushalt handelt oder die Frage nach der Haushaltsgröße *wHA0100* inhaltlich nicht beantwortet wurde, wurden ausschließlich Angaben zur befragten Person selbst abgelegt. Dort enthalten sind neben den Identifikationsvariablen *pnr* und *hnr*, der Stichprobenkennung *sample* die Fragen des Moduls *Lebensstandard im Allgemeinen und im Haushalt* sowie die Angaben zum Alter *wpalter* und zur verwandtschaftlichen Beziehung zu den anderen Haushaltsmitgliedern *wPHH0500/wPHH0501*. Um die Befragungsdauer in sehr großen Haushalten zu begrenzen, wurden in Haushalten mit mehr als fünf Mitgliedern, zu denen valide

Alters- und Verwandtschaftsangaben vorlagen, neben der befragten Person selbst noch vier weitere Personen ausgewählt, für die Fragen des Moduls *Lebensstandard im Allgemeinen und im Haushalt* gestellt wurden. Dadurch liegen für 264 der insgesamt 12.115 Haushaltsmitglieder keine Angaben zu den Deprivationsfragen vor.

Die erhobenen Items wurden anonymisiert (siehe Kapitel 5.4), erhobene offene Textangaben wurden codiert (siehe Kapitel 4.1).

In der Labelung des Websurvey-Scientific Use File werden Umlaute und Sonderzeichen aufgelöst, um eine einheitliche Anzeige der Labeltexte unabhängig vom bei den einzelnen Nutzerinnen und Nutzern eingestellten Zeichensatz zu gewährleisten.

4 Generierte Variablen

4.1 Vercodung offener Restkategorien

Zwei Items der Befragung (*wPHH0500* und *wPEW0100*) wurden als geschlossene Items mit offener Restkategorie erhoben. In beiden Fällen wurden – in Anlehnung an das Vorgehen zur Datenaufbereitung der Haupterhebung – zusätzliche Variablen generiert, die sich von der ursprünglichen Version der Variable lediglich dahingehend unterscheiden, dass die offenen Angaben soweit möglich den entsprechenden Kategorien zugecodet wurden. Es wurden keine neuen Kategorien auf Basis der offenen Angaben gebildet. Die Bezeichnung dieser Variablen unterscheidet sich von der ursprünglichen Variable nur dadurch, dass die „0“ in der letzten Ziffer durch eine „1“ ersetzt wurde. Die nachfolgende Tabelle gibt einen Überblick über die offenen Angaben, die vercodet wurden.

Tabelle 3: Vercodung offener Angaben

Fester Variablenname	zugecodet zu Variable	Bezeichnung (Datensatz)
<i>wPHH0500</i>	<i>wPHH0501</i>	Erwerbsstatus (<i>wPENDDAT</i>)
<i>wPEW0100</i>	<i>wPEW0101</i>	Verwandtschaftliche Beziehung (<i>wRESSOURCEN</i>)

4.2 Einfache Generierungen

Die einfachen Generierungen umfassen Variablen, die auf der (Nicht-)Teilnahme an der Panelhaupterhebung der Welle 17 basieren, die als Systeminformation während des Ausfüllens des Websurveyfragebogens gemessen oder die auf Basis der im Websurvey gemachten Angaben generiert worden sind. Die einfachen generierten Variablen sind in der folgenden Tabelle dargestellt. Die einzelnen Variablen werden dabei jeweils kurz beschrieben. Darüber hinaus werden die für die Generierung der Variable notwendigen Quellvariablen aufgeführt.

Tabelle 4: Einfache Generierungen (alphabetische Reihenfolge)

Variable	Label und Beschreibung (Datensatz)	Quellvariablen für die Generierung
NoWelle17	Person wurde (nicht) in Welle 17 befragt (wPENDDAT) Gibt an, ob die befragte Person in Welle 17 der Panelhaupterhebung ein Personeninterview gegeben hat	pnettok17 (p_register)
walterssplitt	Gruppierung Kinderalter (wPENDDAT) Die PLS*50-Fragen wurden split-half entweder für ein 5 oder für ein 13 Jahre altes Kind gestellt. Die Variable gibt an, welches Alter im Fragentext eingeblendet wurde	vom Erhebungsinstitut geliefert
wformauf	Indikator: Formale Auffälligkeit (wPENDDAT) Gibt an, ob ein Fall im Zuge der Strukturprüfungen auffällig war, wegen einer plausiblen Erklärung der Abweichung jedoch als gültig realisiert gewertet wurde. Das Vorgehen bei den Prüfungen wird in Kapitel 5.1 detailliert erläutert	vom Erhebungsinstitut geliefert
wgebhalbj	Geburtshalbjahr, generiert (wPENDDAT) Indikator, ob Geburtstag im ersten oder zweiten Halbjahr des Geburtsjahres liegt	Informationen zum Geburtsmonat
wgeraet	Eingabegerät (wPENDDAT) Aus dem User Agent String des von der befragten Person verwendeten Browsers wurde die Information generiert, ob für das Ausfüllen des Fragebogens ein Gerät mit Android bzw. iOS als Betriebssystem verwendet wurde oder ein Gerät mit einem anderen Betriebssystem. Die Betriebssysteme Android bzw. iOS wurden der Kategorie „Smartphone / Tablet“ zugeordnet, alle anderen Betriebssysteme der Kategorie „PC / Laptop“. 35 befragte Personen haben im Laufe der Befragung das genutzte Endgerät gewechselt mit der Konsequenz, dass sich auch das genutzte Betriebssystem geändert hat. In diesen Fällen wird das Betriebssystem herangezogen, dass bei der Mehrheit der beantworteten Fragen genutzt wurde	vom Erhebungsinstitut geliefert
wintdatende	Datum Befragungsende (letzte Eingabe) (wPENDDAT) Die Variable zeigt im Format YYMMTT an, an welchem Tag die befragte Person die Beantwortung des Fragebogens beendet hat	vom Erhebungsinstitut geliefert
wintdatstart	Datum Befragungsstart (erste Eingabe) (wPENDDAT) Die Variable zeigt im Format YYMMTT an, an welchem Tag die befragte Person die Beantwortung des Fragebogens begonnen hat	vom Erhebungsinstitut geliefert
wintdauer	Gesamtdauer (in Sekunden) (wPENDDAT) Gibt die Gesamtdauer an, die die befragte Person auf das Ausfüllen des Fragebogens verwendet hat. Gemessen wird ausschließlich das aktive Ausfüllen des Fragebogens. Wenn das Ausfüllen des Fragebogens unterbrochen wurde, ist die Dauer der Unterbrechung nicht Teil der gemessenen Gesamtdauer	vom Erhebungsinstitut geliefert
wpalter	Alter (wRESSOURCEN) Für die befragte Person selbst wurden im Interview Geburtsmonat und -jahr in Frage wPD0100 erhoben. Aus diesen Angaben und dem Befragungsdatum wurde das Alter in Jahren generiert. Für die weiteren Haushaltsmitglieder wurde das Alter in Jahren abgefragt	wPD0100 (wPENDDAT)

Variable	Label und Beschreibung (Datensatz)	Quellvariablen für die Generierung
<i>wpsprache</i>	<i>Sprachauswahl (wPENDDAT)</i> Gibt die im Interview genutzte Befragungssprache an. 34 befragte Personen haben im Interviewverlauf die Sprache gewechselt. In diesen Fällen wird die Sprache herangezogen, die bei der Mehrheit der beantworteten Fragen genutzt wurde. Für die Filtersteuerung des Vignettenmoduls war die Befragungssprache relevant, da dieses Modul ausschließlich auf Deutsch erhoben wurde. Im Erhebungsinstrument wurde jene Sprache zur Filtersteuerung herangezogen, die auf dem Screen der Frage <i>wPHH0700</i> unmittelbar vor dem Vignettenmodul genutzt wurde. Dies war in Einzelfällen nicht die in <i>wpsprache</i> dokumentierte Sprache. Abweichungen zwischen der Information, ob in <i>wpsprache</i> die Sprache Deutsch abgelegt ist und dem Umstand, ob das Vignettenmodul beantwortet wurde, lassen sich auf diesen Umstand zurückführen	vom Erhebungsinstitut geliefert
<i>wvigdauer</i>	<i>Gesamtbearbeitungszeit der Vignetten (in Sekunden) (wVIGNETTEN)</i> Gibt die Dauer an, die die befragte Person auf das Ausfüllen des Vignettenmoduls verwendet hat. Gemessen wird ausschließlich das aktive Ausfüllen des Fragebogens. Wenn das Ausfüllen des Fragebogens unterbrochen wurde, ist die Dauer der Unterbrechung nicht Teil der gemessenen Gesamtdauer	vom Erhebungsinstitut geliefert
<i>wzplfd</i>	<i>Durchlaufzähler Haushaltsmitglied (wRESSOURCEN)</i> Durchlaufzähler für die Haushaltsmitglieder, wobei die 1 immer an die befragte Person vergeben werden soll und ab Person 6 die Haushaltsmitglieder genannt werden, die nicht ausgewählt wurden, um die Fragen zur Deprivation zu beantworten	vom Erhebungsinstitut geliefert

5 Datenaufbereitung

Die Datenaufbereitung wird von infas durchgeführt. Wichtige Entscheidungen bspw. zur Bereinigung von Strukturproblemen bei den befragten Personen wurden in Abstimmung mit dem IAB getroffen. Darüber hinaus stand das IAB auch für die Beantwortung von Fragen während der Zeit der Datenaufbereitung zur Verfügung.

Die in den Websurveyinterviews erhobenen Informationen liegen bei infas zunächst als ASCII-Daten vor. In einer ersten Stufe werden die Rohdaten als Querschnittsdatensatz in eine Statistiksoftware eingelesen⁴. Dabei werden die unvollständig realisierten Interviews von den vollständigen Interviews abgetrennt. Als vollständig realisiert gelten nur die Fälle, bei denen der Fragebogen vollständig bis zur letzten Frage des Fragebogens durchlaufen wurde. In einer zweiten Stufe werden ausschließlich die Daten der vollständigen Interviews einer ausführlichen formalen und inhaltlichen Prüfung unterzogen und zum Scientific Use File aufbereitet.

Die sich anschließende Datenprüfung bei infas unterteilt sich in zwei Schritte, die in den folgenden Abschnitten genauer dargestellt werden. Zunächst wurden Strukturprüfungen und ggf. Bereinigungen vorgenommen. Bei schweren Strukturproblemen wurden die entsprechenden Interviews entfernt (siehe dazu Kapitel 5.1). Im Anschluss erfolgten eine detaillierte Überprüfung und ggf. Bereinigung der Filterführung. Dabei wurden spezifische Codes für Missing Values gesetzt (siehe dazu Kapitel 5.2).

Die folgende Tabelle gibt eine Übersicht aller im Rahmen der Datenaufbereitung durchgeführten Arbeitsschritte und ihre Abfolgen:

Tabelle 5: Übersicht über die Schritte der Datenaufbereitung

Nr.	Arbeitsschritt
1	Einlesen der erhobenen Rohdaten in Arbeitsdatensätze
2	Neugenerierung von Steuerungsvariablen auf Basis der filtergeprüften Datensätze der Panelhaupterhebung der Welle 18
3	Strukturprüfungen (siehe Kapitel 5.1)
4	Entfernen problematischer Interviews (siehe Kapitel 5.1)
5	Korrektur der Struktur (siehe Kapitel 5.1)
6	Filterbereinigung (siehe Kapitel 5.2)
7	Zucodung offener Angaben (siehe Kapitel 4.1)
8	Einfache Generierungen (siehe Kapitel 4.2)
9	Generierung der Datenstruktur für Scientific Use File
10	Anonymisierung (siehe Kapitel 5.4)

5.1 Strukturprüfungen und entfernte Interviews

Der Filterbereinigung vorgelagert fand eine Strukturprüfung statt. Diese bestand in einem Abgleich der im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr mit

⁴ Im Rahmen der Datenaufbereitung kamen die Softwarepakete Stata Version 18 sowie SPSS Version 26 zum Einsatz.

den aktuellsten aus der Panelhaupterhebung bekannten Angaben, um Fälle erkennen und ggf. korrigieren zu können, bei denen eine andere als die hierzu eingeladene Person das Websurveyinterview gegeben hat. Bei einer schriftlichen Einladung zur Interviewteilnahme mit Verteilung von individualisierten Zugangscodes zum Befragungsinstrument besteht die Möglichkeit, dass die Zugangscodes innerhalb eines Haushalts bewusst oder versehentlich durch andere Haushaltsmitglieder verwendet werden. Für die wellenübergreifende Betrachtung der erhobenen Daten ist die konsistente Zuordnung und eindeutige Identifizierbarkeit der Personen über die Wellen hinweg zentral. Eine feste Personennummer darf nicht in verschiedenen Wellen unterschiedlichen Personen zugeordnet sein. Wurde auf der Personenebene ein Interview mit der falschen Person geführt, wurde dieses Interview entweder entfernt oder es wurde die feste Personennummer korrigiert. Es wird im Folgenden erläutert, in welchen Konstellationen eine der beiden möglichen Entscheidungen getroffen wurde. Fälle, bei denen eine Korrektur erforderlich war und umgesetzt werden konnte, sind im Datensatz in der Variablen *wformauf* gekennzeichnet.

Beim Abgleich von Geschlecht sowie Geburtsmonat und -jahr wurde ohne pauschalen Ermessensspielraum geprüft. Eine Abweichung wurde also z.B. auch dann im Detail geprüft, wenn lediglich der Geburtsmonat um einen Monat abwich und das Geschlecht sowie das Geburtsjahr identisch waren mit den aus der Panelhaupterhebung bekannten Angaben. Es wurden im Haushaltsnetzwerk *uhn* der von der Abweichung betroffenen Person die Angaben zum Geburtsdatum bzw. zum Alter in Jahren sowie zum Geschlecht aller Personen betrachtet, die in der Panelhaupterhebung jemals erfasst worden sind. Zudem wurden die aus der Panelhaupterhebung bekannten Vornamen zur Prüfung herangezogen. Alle Interviews im Websurvey, bei denen Abweichungen bestanden und die als gültig gewertet wurden, sind in der Variablen *wformauf* mit Code 1 gekennzeichnet. Korrekturen an den im Websurvey erhobenen Angaben zu Geschlecht sowie Geburtsmonat und -jahr wurden in keinem Fall vorgenommen.

Das weitere Vorgehen hing davon ab, ob die im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr einer anderen Person im Haushaltsnetzwerk *uhn* zugeordnet werden konnten. Wenn dies nicht der Fall war und die Abweichung geringfügig war, wurde das Interview ohne weitere Eingriffe als gültig gewertet. Eine Geringfügigkeit wurde beim Geburtsdatum dann angenommen, wenn nur eines der beiden geprüften Merkmale Monat oder Jahr abwich und die Abweichung beim Geburtsdatum als Tippfehler interpretiert werden konnte oder wenn mit den Angaben im Websurvey eine Angabe aus der Panelhaupterhebung repliziert wurde, die vor der dort aktuellsten Angabe in einer früheren Welle gemacht worden war. Beim Geschlecht wurde eine Abweichung als unproblematisch eingestuft, wenn der Vorname für beide Geschlechter verwendet wird oder wenn anhand des Vornamens entschieden werden konnte, dass das in der Panelhaupterhebung erfasste Geschlecht eindeutig falsch ist. Zudem wurde eine Abweichung beim Geschlecht auch dann als unproblematisch eingestuft, wenn die Angaben zum Geburtsdatum nicht abwichen und sich aus der Betrachtung aller Personen im Haushaltsnetzwerk keine Hinweise darauf ergaben, dass das Interview durch eine andere Person gegeben worden sein könnte. Eine weitere unproblematische Abweichung in der Geschlechtsangabe besteht in der Selbsteinordnung als divers im Websurvey, während in der Panelhaupterhebung das männliche oder das weibliche Geschlecht erfasst ist. Wenn die Abweichung beim Geburtsdatum hingegen erheblich war, weil sie

mehrere Altersjahre betrug und sich nicht auf einen Tippfehler zurückführen ließ, wurde das Interview als ungültig realisiert gewertet und entfernt. Gleiches gilt bei einem unklaren Prüfergebnis, wenn z.B. das Geburtsdatum zu Person A und das Geschlecht zu Person B passte.

Wenn die im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr einer anderen Person im Haushaltsnetzwerk *uhnr* zugeordnet werden konnten, bestand grundsätzlich die Möglichkeit, die *pnr* dieses Interviews zu korrigieren. Sofern die identifizierte Person in einem anderen Haushalt *hnr* innerhalb des Haushaltsnetzwerks *uhnr* lebte, wurde zudem der Korrekturbedarf an der Haushaltsnummer *hnr* berücksichtigt. Diese Korrektur ist allerdings von weiteren Voraussetzungen abhängig. Zur Teilnahme am Websurvey eingeladen wurden alle Personen, die in der Panelhaupterhebung der Welle 18 ein vollständiges und gültiges Interview auf der Personenebene gegeben haben und die zum Feldbeginn des Websurveys ihre Teilnahmebereitschaft nicht zurückgezogen haben. Wenn die Person, die ausweislich der Angaben zu Geschlecht sowie Geburtsmonat und -jahr tatsächlich das Interview gegeben hat, in Welle 18 kein in den *PENDDAT*-Datensatz aufgenommenes Interview gegeben hat, wurde das Websurveyinterview entfernt. Auch wenn für die Person, die tatsächlich das Interview gegeben hat, ein gültiges Interview aus der Panelhaupterhebung der Welle 18 vorliegt, wurde ihr Websurveyinterview unter bestimmten Umständen entfernt. Vereinzelt hat eine Person sowohl unter Nutzung des eigenen Zugangscode als auch zusätzlich unter Nutzung des Zugangscode eines weiteren Haushaltsmitglieds ein Interview gegeben. Dann wurde das Interview, das unter Nutzung des Zugangscode eines weiteren Haushaltsmitglieds gegeben wurde, entfernt. Das Interview, das unter Nutzung des eigenen Zugangscode gegeben wurde, wurde in der Variablen *wformauf* mit Code 1 gekennzeichnet. Die Variable *wformauf* kennzeichnet somit nicht ausschließlich Abweichungen bei den im Websurvey gemachten Angaben zu Geschlecht sowie Geburtsmonat und -jahr im Vergleich zur Panelhaupterhebung, sondern auch Auffälligkeiten im Teilnahmeverhalten.

Zusammenfassend gilt, dass die *pnr* und ggf. auch die *hnr* korrigiert wurden, wenn die Person, die den Zugangscode eines anderen Haushaltsmitglieds genutzt hat, ebenfalls Teil der Einsatzstichprobe des Websurveys war und den Websurveyfragebogen nur ein einziges Mal ausgefüllt hat.

Das Interview wurde hingegen nicht entfernt, wenn die auf der Basis der Panelhaupterhebung der Welle 18 generierte Steuerungsvariable *NoWelle17* sich zwischen den beiden Personen, die an der Vertauschung des Zugangscode beteiligt waren, unterschieden haben. Denn in diesem Fall hat die Person, die das Interview gegeben hat, zwar teilweise Fragen erhalten, die ihr bei Nutzung ihres eigenen Zugangscode nicht vorgelegt worden wären. Oder sie hat teilweise Fragen nicht erhalten, die ihr bei Nutzung ihres eigenen Zugangscode vorgelegt worden wären. Dies betraf jedoch ausschließlich die Items der Frage *PEO0400**, die in diesem Fall auf den Missing Code -3 bzw. -4 umgesetzt wurde. Genauere Ausführung zur Vergabe der Missing Codes finden sich in Kapitel 5.2.

Von den 5.186 realisierten Interviews waren 192 formal auffällig. In 161 Fällen wurde die Auffälligkeit als unproblematisch bewertet. In den restlichen 31 Fällen lag ein ungültig realisierter Fall vor, so dass die Datensätze des Scientific Use Files 5.155 Fälle umfassen.

Es wurde darauf verzichtet, den weiteren Haushaltsmitgliedern, zu denen die befragte Person in den im Datensatz *wRESSOURCEN* abgelegten Fragen Auskünfte gegeben hat, eine *pnr* zuzuordnen. Für eine solche Zuordnung müssten Vorname, Alter und Geschlecht genutzt werden. Die Zuordnung wird häufig nicht ohne weiteres möglich sein. Ein Match, bei dem Namen als Matchschlüssel verwendet werden, muss mit dem Problem unterschiedlicher Schreibweisen umgehen. Zudem ist

die Altersangabe kein fixes Merkmal, sondern kann sich seit der Panelhaupterhebung der Welle 18 verändert haben. Darüber hinaus müsste geprüft werden, ob man den Abgleich auch mit Personen vornimmt, die vor Welle 18 irgendwann einmal Teil des Haushalts waren und ihn seither verlassen haben. Weitere Sonderfälle sind denkbar. So könnten Personen ermittelt werden, die zum Zeitpunkt der Panelhaupterhebung noch nicht Teil des Haushalts waren. Oder es könnten mehrere Personen gleiche Merkmale aufweisen, insbesondere bei Haushalten, in denen Zwillinge leben. Es muss daher angenommen werden, dass es eine relevante Anzahl von Fällen gibt, bei denen keine eindeutige Zuweisung einer *pnr* erfolgen kann.

5.2 Filterbereinigung und vergebene Missing Codes

In der Filterbereinigung wurde die im Instrument vorgesehene Filterführung mithilfe eines Statistikprogramms nachvollzogen. Wurden bestimmte Fragen gestellt, obwohl der Wert der relevanten Filtervariable etwas anderes verlangt hätte (wurden also z.B. detaillierte Angaben zu Entscheidungen in Partnerschaften erhoben, obwohl die/der Befragte angegeben hatte, nicht mit einem Partner bzw. einer Partnerin zusammenzuleben), wurden die Variablen auf den Missing Code „-3“ (trifft nicht zu) gesetzt, den sie auch bei einer korrekten Ausführung der Filterführung erhalten hätten⁵. Des Weiteren wurde geprüft, ob in Einzelfällen Items nicht erhoben wurden, obwohl dies laut der entsprechenden Filtervariable notwendig gewesen wäre. In diesen Fällen wurde der spezifische Missing Code „-4“ (Frage irrtümlich nicht gestellt) vergeben. Die Vergabe des Codes „-4“ war in wenigen Einzelfällen bei Frage *PEO0400** notwendig, wenn die Befragung durch ein anderes Haushaltsmitglied durchgeführt wurde und die tatsächlich antwortende Person im Gegensatz zur eigentlichen Befragungsperson kein Interview in Welle 17 der Panelhaupterhebung gegeben hat. Genauere Ausführungen zu solchen Korrekturen der *pnr* finden sich in Kapitel 5.1.

Daneben wurden im Rahmen der Filterprüfung auch die vom Feldinstitut vergebenen Missing Codes sowie die System Missings durch für alle Variablen einheitliche Werte ersetzt. Die vergebenen Werte sind in nachfolgender Tabelle im Überblick dargestellt. „-1“ und „-2“ sind die einheitliche Umkodierung der in der Befragung erhobenen „WN“- und „KA“-Werte. „-3“ ist der generelle „trifft nicht zu“-Code für aufgrund der Filterführung nicht gestellte Fragen. Bei „-5“ und „-6“ handelt es sich um fragespezifische Codes. Hierbei kann es sich sowohl um spezifische Missing Codes als auch um Sonderkategorien für gültige Werte handeln. Diese Codes werden nur im Bedarfsfall vergeben, wobei in dieser Erhebung kein entsprechender Bedarf bestand. Mit dem Code „-11“ wird gekennzeichnet, dass die befragte Person die Frage verlassen hat, ohne eine Antwort anzugeben.

Tabelle 6: Übersicht über die vergebenen Missing Codes

Code	Erläuterung
-1	„weiß nicht“
-2	„keine Angabe“
-3	„trifft nicht zu (Filter)“ (Frage wurde aufgrund der Filterführung nicht gestellt)
-4	„Frage irrtümlich nicht gestellt“ (Frage hätte jedoch gestellt werden sollen)
-5	Fragespezifischer Code Nr. 1, wurde nur bei Bedarf vergeben

⁵ Die Bereinigung erfolgte dabei, wie in solchen Fällen üblich, immer von den zuerst erhobenen zu den später erfragten Items.

Code	Erläuterung
-6	Fragespezifischer Code Nr. 2, wurde nur bei Bedarf vergeben
-11	„Weiterklicken ohne Angabe“

5.3 Plausibilitätsprüfungen

Anders als in der Panelhaupterhebung wird im Websurvey der Welle 18 auf Plausibilitätsprüfungen verzichtet. Die Liste der in der Panelhaupterhebung angewendeten Prüfungen wurde mit dem Fragebogen des Websurveys abgeglichen mit dem Ergebnis, dass im Websurvey keine prüffähigen Sachverhalte erhoben werden. Eine Prüfung im Längsschnitt, also zwischen den im Websurvey und den in der Panelhaupterhebung gemachten Angaben, wurde nicht durchgeführt.

5.4 Anonymisierung

Vom IAB als besondere Dienststelle der Bundesagentur für Arbeit (BA) erhobene Daten stellen immer Sozialdaten dar, weshalb die Ansprüche an den Datenschutz entsprechend hoch sind. Daher wurde es notwendig, zwei Variablen nur in vergrößerter Form in den Scientific Use File aufzunehmen. Die nachfolgende Tabelle gibt einen Überblick über die betroffenen Variablen und die vorgenommenen Anonymisierungsschritte.

Tabelle 7: Übersicht über die anonymisierten Variablen

Varname	Variablenlabel	Vorgehen
wPD0100	Geburtsjahr	Der zusätzlich erhobene Geburtsmonat wurde nicht aufgenommen.
wgebhalbj	Geburtshalbjahr, generiert	Der Monat des Geburtsdatums wurde zu einem Indikator für das 1. bzw. 2. Halbjahr verkürzt.

6 Gewichtung Websurvey 2024

Da nur ein Teil der Personen, die an Welle 18 von PASS teilgenommen haben, auch am Websurvey teilgenommen hat, ist ein zusätzlicher Gewichtungsfaktor (*wteilwahr*) notwendig, um für diesen selektiven Ausfall zu korrigieren. Die PASS-Personen-Hochrechnungsfaktoren aus Welle 18 (*wqp*, *wqmip*, *wqbap*) können mit diesem zusätzlichen Gewichtungsfaktor multipliziert werden, woraus die Hochrechnungsfaktoren für den Websurvey resultieren. Der zusätzliche Gewichtungsfaktor basiert auf den invertierten vorhergesagten Wahrscheinlichkeiten eines Logit-Regressionsmodells zur Schätzung der Teilnahmeneigung am Websurvey.

Unter der Berücksichtigung aller Panelpersonen, die potenziell die Möglichkeit hatten, an der Web-Befragung teilzunehmen (11.235 Fälle), wurde ein Logit-Modell für die Teilnahme geschätzt. Die unabhängige Variable zur Teilnahme am Websurvey setzt sich dabei wie folgt zusammen:

- Ausprägung 1 – teilgenommen: Personen, die den Websurvey selbst und vollständig ausgefüllt haben
- Ausprägung 0 – nicht teilgenommen: Personen, die gar nicht teilgenommen haben sowie Personen, die die Befragung abgebrochen haben (*teilnahme=2*) oder bei denen aufgrund von Plausibilitätsprüfungen vermutet wurde, dass eine andere Person den Websurvey ausgefüllt hat (*wraus=2*)

Die Prädiktorvariablen wurden bereits mit dem ersten PASS-Websurvey in Welle 16 ermittelt und für Welle 18 erneut überprüft. In Welle 16 wurden die Prädiktorvariablen aus einem großen Pool an potenziell relevanten Merkmalen aus dem Personen-, Haushalts- und Methodendatensatz der Welle 16, auf Basis theoretischer Überlegungen sowie vorhandener Literatur ausgewählt.

Dementsprechend wurden die Prädiktorvariablen inklusive der Referenzkategorien bei kategorialen Variablen aus Welle 16 für Welle 18 übernommen, falls sie vorhanden waren, um zwischen den Wellen ein konsistentes und vergleichbares Vorgehen zu haben.

Wie in Welle 16 wurden einige Variablen, die theoretisch einflussreich erschienen, aufgrund von Bedenken bezüglich Multikollinearität aus dem Modell ausgenommen, wie beispielsweise die Staatsbürgerschaft (*PMI0400*). Zudem wurden in Welle 16 wie auch in Welle 18 Ausprägungskategorien von Variablen, wie das Geschlecht (*zpsex*), den Erwerbsstatus (*statakt*), den höchsten Schulabschluss (*schul2*) und die Ortsgrößenklasse (*bik*), aufgrund von Multikollinearität oder aufgrund geringer Fallzahlen, zusammengefasst.

Das Alter der befragten Person (*zpalthh*) wurde wie in Welle 16 in eine kategoriale Variable mit drei Ausprägungen (< 30, 30-74, >74) transformiert.

Berücksichtigt wurden zum einen demographische Variablen, die einen hohen Einfluss auf die Teilnahme vermuten lassen. Hierbei handelt es sich um Altersgruppen (*alterKat*), Geschlecht (*zpsex*), Familienstand (*famstand*), Migrationshintergrund (*migration*), höchsten Schulabschluss (*schul2*), Zahl der eigenen Kinder (*kindzges*), Deprivationsindex (*depindg3*), Ortsgrößenklasse (*bik*), Art des Mietverhältnisses (*HW0300*) sowie Wohnort in alten oder neuen Bundesländern (*blneualt*). Zudem wurden die Anzahl der Kontaktversuche (*ANZKON*) der Welle 18 und eine Variable, die indiziert, ob eine Person ursprünglich aus einer allgemeinen Bevölkerungsstichprobe stammt (*sampleType*), in das Modell miteinbezogen. Eine weitere Variable (*INR*) gibt die Zahl der im Personeninterview der

Welle 18 mit „weiß nicht“ oder „keine Angabe“ beantworteten Fragen wieder. Für die Generierung der Variable *INR* wurden nur Merkmale berücksichtigt, die alle Befragten erhalten haben und die nicht zum Pool der potenziellen Prädiktorvariablen gehörten. Es wurde weiterhin eine Variable aufgenommen, die die Zeit zwischen Durchführung des Personeninterviews und Beginn des Websurveys (24.09.2024) in Tagen misst (*vergangeneZeit*).

Da in Welle 18 keine Fragen zur Internetaffinität gestellt wurden, wurde ersatzweise die Variable verwendet, ob die Person Zugang zu einem PC mit Internetanschluss hat (*HLS1100a*).

Bei der Erstellung der Gewichte für den Websurvey W16 wurden mittels Stepwise Selection (Funktion `step (direction="both")` in R) der Erwerbstatus (*statakt*), der Bürgergeld-Bezug (*alg2abez*), die Interviewsprache der Welle 18 (*psprache*) und die Zahl der bisherigen Wellen (*HHTLN*) aus dem Modell entfernt, da diese im Rahmen des Auswahlalgorithmus die Güte des Modells nicht verbessern konnten. Diese Auswahl wurde für den Websurvey Welle 18 beibehalten.

Fehlende Werte wurden in den finalen Modellen multipel imputiert, um für jede teilnehmende Person eine Teilnahmewahrscheinlichkeit schätzen zu können.

Tabelle 8 enthält eine Übersicht über alle im Modell verwendeten Variablen (Variablenname und Erläuterung), Tabelle 9 enthält die Modellkoeffizienten, Fallzahlen und Anpassungsgüte des verwendeten Modells.

Tabelle 8: Variablenübersicht, Abkürzungen und Referenzkategorien für die Logit-Modelle

Variablenkürzel und Referenzkategorie	Erläuterung
alterKatbis29	Person jünger als 30 Jahre
alterKatab75	Person 75 Jahre und älter
Referenzkategorie	Person 30 – 74 Jahre alt
zpsexW	Person Weiblich
Referenzkategorie	Person Männlich oder Divers
Kindzges	Anzahl eigene Kinder gesamt (inner- und außerhalb HH)
schul2_sNoch_Schüler_In	Schulabschluss Person: noch Schüler*In
schul2_sKein_Abschluss	Schulabschluss Person: ohne Abschluss
schul2_sVolks_Hauptschule	Schulabschluss Person: Volks-/Hauptschule
schul2_sMittlere_Reife	Schulabschluss Person: Realschule/mittlerer Abschluss
schul2_sFachhochschule	Schulabschluss Person: Fachhochschulreife
schul2_sSonstiges	Schulabschluss Person: Sonder-/Förderschule/Ausland/Anderes
Referenzkategorie	Schulabschluss Person: Hochschulreife
famstandVerheiratet..z.l.	Familienstand Person: Verheiratet/eing. Lebensp., zus. lebd.
famstandVerheiratet..g.l.	Familienstand Person: Verheiratet/eing. Lebensp., getr. lebd.
famstandGeschieden	Familienstand Person: Geschieden
famstandVerwitwet	Familienstand Person: Verwitwet
Referenzkategorie	Familienstand Person: Ledig
migrationSelbst	Migrationshintergrund Person: Selbst zugezogen
migrationEltern	Migrationshintergrund Person: Mind. 1 Elternteil zugezogen
migrationGroßeltern	Migrationshintergrund Person: Mind. 1 Großelt. zugez., Elt. in D geb.
Referenzkategorie	Migrationshintergrund Person: Kein Migrationshintergrund
depindg3	Deprivationsindex, gewichtet
bikCatbis2000	BIK-Größenklasse der Gemeinde: unter 2.000 Einwohner

Variablenkürzel und Referenzkategorie	Erläuterung
bikCat2.000_4.999	BIK-Größenklasse der Gemeinde: 2.000 bis u. 5.000 Einwohner
bikCat5.000_19.999	BIK-Größenklasse der Gemeinde: 5.000 bis u. 20.000 Einwohner
bikCat20.000_49.999	BIK-Größenklasse der Gemeinde: 20.000 bis u. 50.000 Einwohner
bikCat50.000_99.999	BIK-Größenklasse der Gemeinde: 50.000 bis u. 100.000 Einwohner
bikCat100.000_499.999	BIK-Größenklasse der Gemeinde: 100.000 bis u. 500.000 Einwohner
Referenzkategorie	BIK-Größenklasse der Gemeinde: 500.000 und mehr Einwohner
blneualtNeu	Neue Bundesländer inkl. Berlin
Referenzkategorie	Alte Bundesländer
HW0300Untermiete	Art des Mietverhältnisses: Untermieter
HW0300Eigentum	Art des Mietverhältnisses: Eigentümer
HW0300Mietfrei	Art des Mietverhältnisses: Wohne mietfrei
HW0300Anderes	Art des Mietverhältnisses: Wohngemeinschaft/Wohnheim
Referenzkategorie	Art des Mietverhältnisses: Hauptmieter
HLS1100aKein PC mit Internet	Person hat keinen Zugang zu einem PC mit Internetanschluss
Referenzkategorie	Person hat Zugang zu einem PC mit Internetanschluss
ANZKON	Anzahl der Kontaktversuche in Welle 18
vergangeneZeit	Tage zwischen Datum der Personenbefragung W18 und Start Websurvey (24.09.2024)
sampleTypeEWO_Microm	Microm-Stichprobe (inkl. EWO-Auffrischung)
Referenzkategorie	BA-Stichproben (inkl. BA-Auffrischungen und BA-Aufstockung)
INR	Item Nonresponse (Weiß Nicht und K.A.) in Personenbefragung der W18

Tabelle 9: Logit-Modell zur Teilnahme am Websurvey

	Teilnahme	
	Coef.	p
(Intercept)	0,4968314	0,000
alterKatbis29	-0,3681976	0,000
alterKatab75	-0,5774036	0,000
zpsexW	0,1938617	0,000
Kindzges	-0,0818329	0,000
schul2_sNoch_Schüler_In	-0,0629681	0,652
schul2_sKein_Abschluss	-0,9896645	0,580
schul2_sVolks_Hauptschule	-0,8406919	0,000
schul2_sMittlere_Reife	-0,4229489	0,000
schul2_sFachhochschule	-0,2805793	0,000
schul2_sSonstiges	-0,7674469	0,580
famstandVerheiratet..z.l.	0,2865658	0,000
famstandVerheiratet..g.l.	0,1660641	0,151
famstandGeschieden	0,0276221	0,191
famstandVerwitwet	-0,0277426	0,205
migrationSelbst	-0,4851747	0,000
migrationEltern	-0,1274539	0,704
migrationGroßeltern	0,0356436	0,939
depindg3	-0,0487390	0,941
bikCatbis2000	-0,3512376	0,678
bikCat2.000_4.999	-0,1929014	0,109

	Teilnahme	
	Coef.	p
bikCat5.000_19.999	-0,1062926	0.719
bikCat20.000_49.999	-0,1421352	0.195
bikCat50.000_99.999	-0,0912534	0.759
bikCat100.000_499.999	-0,1465637	0.215
blneualtNeu	-0,1884084	0.209
HW0300Untermiete	-0,2827080	0.471
HW0300Eigentum	0,0083661	0.890
HW0300Mietfrei	-0,0767651	0.487
HW0300Anderes	-0,0347052	0.000
HLS1100aKein PC mit Internet	-0,8995978	0.031
ANZKON	-0,0125012	0.000
vergangeneZeit	0,0006848	0.110
sampleTypeEWO_Microm	0,2050688	0.098
INR	-0,1414593	0.000
n	11.235	
Log likelihood	-7.081	
Pseudo R²	0,0862	

Um den PASS-Websurvey auf die Wohnbevölkerung in Deutschland ab 15 Jahren hochzurechnen, sind zwei weitere Schritte erforderlich:

Das PASS-Personengewicht (*wqp*) der dazugehörigen PASS-Welle (Panel Haupterhebung Welle 18) wird mit der reziproken vorhergesagten Teilnahmewahrscheinlichkeit (*wteilwahr*) am Websurvey multipliziert. Da dieser Schritt allein einerseits nicht gewährleistet, dass die Größe der Zielpopulation (71,856 Mio.) durch die Gewichte replizierbar ist und andererseits starke Ausreißer in der Verteilung der Gewichte entstehen können, wird im Anschluss daran eine Kalibrierung des neuen Gewichts durchgeführt. Neben der Populationsgröße wird dabei lediglich vorgegeben, dass die Gewichte in den Schranken des Minimums und Maximums des Personengewichts von Welle 18 (*wqp*) dividiert durch die Teilnahmequote des Websurveys bleiben. Es wird damit also insbesondere nicht forciert, dass mit der Gewichtungvariable *wpweight* Verteilungen aus dem Mikrozensus exakt getroffen werden.

Tabelle 10 gibt Aufschluss über Kennwerte der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten (Variable *wteilwahr*) und Tabelle 11 über Kennwerte der Verteilung der berechneten Gewichte (Variable *wpweight*).

Tabelle 10: Kenngrößen der Verteilung der vorhergesagten Teilnahmewahrscheinlichkeiten für die Nettostichprobe

1%-Perzentil	0,1246862
5%-Perzentil	0,2403857
10%-Perzentil	0,3154749
25%-Perzentil	0,435708
50%-Perzentil	0,5427938
75%-Perzentil	0,6267952

90%-Perzentil	0,6849176
95%-Perzentil	0,7133195
99%-Perzentil	0,7486628
Mittelwert	0,5195839
Standardabweichung	0,1423186
Minimum	0,0369354
Maximum	0,7836026
Fallzahl	5.155

Tabelle 11: Kenngrößen der Verteilung der Gewichte (Gesamtstichprobe, Personen)

1%-Perzentil	99,95486
5%-Perzentil	302,1131
10%-Perzentil	527,5805
25%-Perzentil	1332,878
50%-Perzentil	5006,691
75%-Perzentil	17504,78
90%-Perzentil	37780,48
95%-Perzentil	57333,67
99%-Perzentil	107580,6
Mittelwert	13939,09
Standardabweichung	21699,13
Minimum	73,1951
Maximum	158979,5
Fallzahl	5.155

Literatur

- Berg, M., Cramer, R., Dickmann, Gerber, V., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Malich, S., Schwarz, S., Wenzig, C., Trappmann, M., Bähr, S., Bömmel, N., Coban, M., Collischon, M., Gundert, S., Küfner, B., Mackeben, J., Müller, B., Stegmaier, J., Teichler, N. & Wunder, A. (2024): Codebuch und Dokumentation des Panel „Arbeitsmarkt und soziale Sicherung“ (PASS) Bd. 1 Datenreport Welle 17. (FDZ-Datenreport 07/2024). Nürnberg.
- Jesske, Birgit, Schulz, Sabine & Torregroza, Sabrina (2025): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS * 18. Erhebungswelle – 2024 (Haupterhebung). FDZ-Methodenreport 08/2025, Nürnberg.
- Jesske, B., Schulz, S. & Gerber, V. (2024): Methodenbericht Panel Arbeitsmarkt und Soziale Sicherung PASS * 17. Erhebungswelle – 2023 (Haupterhebung). FDZ-Methodenreport 05/2024, Nürnberg.

Impressum

FDZ-Datenreport 11|2025

Veröffentlichungsdatum

14. November 2025

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/2025/DR_11-25.pdf

Dokumentation Version

PASS-SUF0624_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2511.de.v1

Datensatz Version

PASS-SUF0624 v1, DOI: 10.5164/IAB.PASS-SUF0624.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

Jonas Beste
Telefon: 0911 179-2279
E-Mail: jonas.beste@iab.de