

Datensatzdokumentation

Diese Dokumentation beschreibt den aus der BIBB-ANI-Erhebung 2016 resultierenden Datensatz. Die Erhebung wurde bei 5.003 Personen in der Grundgesamtheit der 18- bis 64-Jährigen in Deutschland telefonisch durchgeführt (70% Festnetzanschlüsse und 30% Mobilfunkanschlüsse). Grundlegende Angaben zur Erhebung bezüglich Auswahl, Gewichtung u. ä. finden sich im Methoden und Tabellenband (Müller/Wenzelmann 2018b). Weitere erhebungsbezogene Details (z.B. zur Ausschöpfung) finden sich im Feldbericht des Erhebungsinstituts infas Sozialforschung GmbH, der auf Anfrage durch das Projektteam des BIBB zur Verfügung gestellt werden kann.

1. Beschreibung Personendatensatz

Sämtliche Angaben im Datensatz beziehen sich auf die Personenebene. Der Datensatz lässt sich jedoch in einen Datensatz umwandeln, in dem jede Lernaktivität einen Beobachtungsfall darstellt.¹ Dann müssen entsprechende Gewichte berechnet werden, welche die Auswahl der Lernaktivitäten berücksichtigt. Für die von BIBB-Seite durchgeführten Auswertungen ist dies erfolgt (vgl. Müller/Wenzelmann 2018a).

Interviewtypen

Die 5.003 Fällen im Datensatz verteilen sich auf vier unterschiedliche Interviewtypen, die sich durch starke Unterschiede im Ablauf und in der Länge des Interviews kennzeichnen. Die Variable -inttyp zeigt die Verteilung:

- | | |
|--|-------|
| • Zielperson mit konkret erinnerten Lernaktivitäten | 3.005 |
| • Zielperson ohne Lernaktivitäten | 1.459 |
| • Zielperson ohne konkrete Erinnerung an Lernaktivitäten | 108 |
| • Nicht erwerbsnahe Zielperson (Kurzinterview) | 431 |

Die erste Gruppe (mit konkret erinnerten Lernaktivitäten) erhielt ein ausführliches Interview, innerhalb dessen, Details zu Aufwand und Nutzen für bis zu zwei zufällig ausgewählte Lernaktivitäten erhoben wurden. Die Interviews dauerten im Durchschnitt 35 Minuten. Die zweite Gruppe erhielt ein kürzeres Interview, in dem die Fragenblöcke zu Aufwand und Nutzen durch einen Fragenblock zum potenziellen Nutzen beruflicher Weiterbildung ersetzt wurden. Diese Interviews dauerten durchschnittlich 15 Minuten.

108 Personen gaben Lernaktivitäten an, konnten oder wollten jedoch kein konkretes Stichwort zu Titel oder Thema nennen. Diese Personen erhielten im weiteren Verlauf des Interviews nur noch einige Fragen zu sozio-demografischen Aspekten und wurden dann ausgesteuert. Bei der Berechnung der Weiterbildungsbeteiligung werden sie als Nicht-Teilnehmende gezählt. Für diese Personen konnte nicht das für Teilnehmende vorgesehene Interview geführt werden, da ohne eine konkretisierende

¹ Pro Person wurden max. zwei Lernaktivitäten im Detail erfasst (vgl. Abschnitt *Auswahl von Lernaktivitäten*).

Angabe zur Lernaktivität keine Bezugnahme auf diese möglich war. Ebenso wenig konnte das für Nicht-Teilnehmende vorgesehene Interview geführt werden, da die Fragen für Personen, die tatsächlich an Weiterbildung teilgenommen hatten, irritierend gewesen wären. Alles in allem halten wir die Angabe zur Weiterbildungsbeteiligung ohne jede weitere konkretisierende Angabe zu Inhalt oder Titel der entsprechenden Lernaktivität(en) für nicht valide genug, um sie tatsächlich als Teilnahme zu werten. Die Interviews dauerten im Durchschnitt 17 Minuten.

Die letzte Gruppe betrifft erwerbsferne Personen, die dem Arbeitsmarkt aus freien Stücken nicht zur Verfügung stehen, weil sie nicht vorhaben, eine Erwerbstätigkeit innerhalb von zwei Jahren aufzunehmen. Diese Personen (431 Interviews) sind für die vorgelegte Erhebung an sich nicht relevant, da sie per Definition nicht berufsbezogen lernen. Allerdings liegen für diese Gruppe keine Informationen zur Verteilung in der Grundgesamtheit vor, sodass die Definition der Auswahlgesamtheit und die anschließende Gewichtung der Daten unmöglich geworden wären, wenn sie aus der Erhebung von vornherein hätte ausgeschlossen werden sollen. Daher wurde erwerbsferne Personen bewusst in die Stichprobe aufgenommen und zu einem frühen Zeitpunkt im Interview identifiziert. Anschließend wurden den betreffenden Personen lediglich einige Fragen zu sozio-demografischen Merkmalen gestellt, die für Gewichtungszwecke genutzt werden konnten. Ein Interview in dieser Gruppe dauerte durchschnittlich gut 3 Minuten.

Auswahl von Lernaktivitäten

Für jede Zielperson wurde im Verlauf des Interviews die Zahl ihrer Lernaktivitäten im Jahr 2015 erhoben, differenziert nach 6 Lernformaten/Typen. Für maximal zwei dieser Aktivitäten je Typ (bei Typ B nur für eine) wurde dann ein thematisches Stichwort sowie eine typspezifische Zusatzinformation erfasst, welche die Lernaktivitäten der verschiedenen Typen näher charakterisieren. Insgesamt liegen diese Angaben somit für maximal 11 Lernaktivitäten pro Person vor. Diese 11 konkret erinnerten Lernaktivitäten, für welche die Befragten ein Stichwort nennen konnten, bilden den personenbezogenen Auswahlpool, aus welchem zwei Aktivitäten für die Detailbefragung zu Dauer, Kosten und Nutzen der Lernaktivität ausgewählt wurden. Sofern nur eine oder zwei Aktivitäten im Auswahlpool vorlagen, wurden diese gewählt. Die Auswahlvorschriften gehen aus dem Erhebungsinstrument hervor.

Variablen

Die Inhalte der Variablen gehen unmittelbar aus dem Erhebungsinstrument hervor. Im Normalfall trägt eine Variable die gleiche Bezeichnung wie die jeweilige Frage im Fragebogen. Ausnahmen sind z.B. durch Variablen gegeben, die automatisch vom CATI-Programm berechnet werden (z.B. gen_az). Diese tragen in der Regel das Präfix „gen_“. Weitere Präfixe und Suffixe haben folgende Bedeutung:

Präfixe

wb1 / wb2 Zeigt bei allen Variablen, die sich auf eine konkrete Lernaktivität beziehen (Frageblöcke 5 bis 8), an, ob es sich um die im ersten oder zweiten Schleifendurchlauf generierte

– also auf die erste oder zweite im Detail befragte Lernaktivität bezogene – Variablen handelt (wb1 bzw. wb2).

Info Kennzeichnet Infotexte vor den entsprechend nummerierten Variablen, die keine Angaben der Befragten erforderten (z.B. Info0209)

random Kennzeichnet die Variablen, die bei Fragen mit randomisierten Antwortitems die Information zur jeweiligen Reihenfolge enthalten

Suffixe

P Gibt an, dass es sich um eine Nachfrage zur Plausibilisierung der Angabe handelt, die ebenfalls aus dem Fragebogen ersichtlich ist (wb1v840p).

o Kennzeichnet Variablen, welche bei Fragen mit offener Antwortmöglichkeit die Information enthalten, ob eine offene Angabe gemacht wurde (z.B. v1002o)

open kennzeichnet Variablen, welche die eigentliche offene Angabe zu der entsprechend nummerierten Frage enthalten (z.B. v1002open). Diese Variablen sind aus Anonymisierungsgründen nur in einem separaten Datensatz erhältlich.

Auch die Wertelabels gehen aus dem Erhebungsinstrument hervor. Fehlende Werte sind wie folgt gelabelt:

- .f regulär überfiltert Variablen
- .w „weiß nicht“
- .v Angabe verweigert

Der Datensatz enthält neben den Informationen, die unmittelbar aus dem Erhebungsinstrument hervorgehen einige weitere Informationen, die nachfolgend mit Kürzel und Beschreibung aufgelistet sind:

LFD Laufende Nummer

inttyp Interviewverlauf (siehe oben)

wbanz Anzahl der insgesamt berichteten Lernaktivitäten, wird automatisch vom CATI-Programm errechnet

wb1-11 Enthält die Angabe (j/n), ob ein Stichwort für die jeweils ausgewählte Lernaktivität angegeben wurde (wb1- Typ A, Aktivität 1; wb2- Typ A, Aktivität 2; wb3 – Typ B, Aktivität 1; wb4 – Typ C, Aktivität 1, wb5 – Typ C, Aktivität 2; wb6 – Typ D, Aktivität 1; wb7 – Typ D, Aktivität 2; wb8 – Typ E, Aktivität 1; wb9 – Typ E, Aktivität 2; wb10 – Typ F, Aktivität 1; wb11 – Typ F, Aktivität 2)

desgew_personen Designgewicht auf Personenebene (macht die Auswahlvorschriften rückgängig)

gewicht_personen	Kalibrierter Gewichtungsfaktor auf Personenebene (an Randverteilungen in der Grundgesamtheit angepasst)
hoch_personen	Kalibrierter Hochrechnungsfaktor auf Personenebene
bland	Bundesland
gkbik10	Gemeindegrößenklasse BIK 1-10 (Sys:753)
gkpol7	Politische Gemeindegrößenklasse
kldb5	Berufsvercodung nach KldB2010 (5-Steller)
kldb2	Berufsvercodung nach KldB2010 (2-Steller)
isco08	Berufsvercodung nach ISCO2008 (2-Steller)
bf2010	Berufshauptfelddefinitionen des BIBB 2010
bhf2010	Berufshauptfeld, Berufshauptfelddefinitionen des BIBB 2010
bof2010	Berufsoberfeld, Berufsfelddefinitionen des BIBB 2010
iscd97	Bildungsabschluss nach ISCED 97
v1002_isei08	ISEI Erwerbsberuf
mig	Migrationshintergrund
Nuts2	NUTS2-Ebene Gebietsstand 2011
Ror	Raumordnungsregion Stand 2015

Zur Vercodung der Berufe auf Grundlage der Variablen 1002open und v1002bopen liegt eine separate Dokumentation vor, die auf Anfrage gerne zur Verfügung gestellt wird.

Des Weiteren sind dem Datensatz die wichtigsten berechneten Variablen beigelegt, die für die eigenen Auswertungen des BIBB verwendet werden. Die unter Punkt 2 beschriebenen Maßnahmen der Datenaufbereitung/Plausibilisierung kamen bei der Berechnung dieser Variablen zur Anwendung. Folgende Variablen sind enthalten:

gesamtkosten_selbst	Direkte und indirekte nach Refinanzierung selbst getragene Kosten der Individuen für ihre im Jahr 2015 stattfindenden (Teile von) Weiterbildungen
direkte_kosten	Direkte Ausgaben der Individuen für ihre im Jahr 2015 stattfindende (Teile von) Weiterbildungen
indirekte_kosten	Einkommensverlust der Individuen infolge ihrer im Jahr 2015 stattfindenden (Teile von) Weiterbildungen

direkte_refinanziert	Rückerstattungen von direkten Ausgaben an die Individuen für ihre im Jahr 2015 stattfindenden (Teile von) Weiterbildungen
indirekte_refinanziert	Zuschuss zum Lebensunterhalt zur Kompensation von Einkommensverlusten in Zusammenhang mit ihren im Jahr 2015 stattfindenden (Teile von) Weiterbildungen
zeitaufwand_wb	Zeitaufwand der Individuen für ihre im Jahr 2015 stattfindende (Teile von) Weiterbildungen
zeitaufwand_gesamt	Gesamter Zeitaufwand (zzgl. Vor- und Nachbereitung, Fahrten, u.ä.) der Individuen für ihre im Jahr 2015 stattfindende (Teile von) Weiterbildungen
freizeitaufwand_wb	Freizeitaufwand der Individuen für ihre im Jahr 2015 stattfindende (Teile von) Weiterbildungen
freizeitaufwand_gesamt	Gesamter Freizeitaufwand (zzgl. Vor- und Nachbereitung, Fahrten, u.ä.) der Individuen für Ihre im Jahr 2015 stattfindenden (Teile von) Weiterbildungen

Aufgrund von Imputationen, die im Zuge der Auswertungen vorgenommen wurden, enthalten manche der berechneten Variablen negative Werte. Beispielsweise sind die selbst getragenen Kosten negativ, wenn imputierte Angaben zur Refinanzierung die tatsächlichen Ausgaben der betreffenden Person übersteigen. Dies ist darauf zurückzuführen, dass Angaben zur Refinanzierung und zu den direkten Kosten separat imputiert und anschließend zu den selbst getragenen Kosten verrechnet wurden. Die Imputationen wurden mittelwerttreu durchgeführt. Eine Verzerrung resultiert daraus also nicht. Die negativen Werte sind reine Rechengrößen. Bei der differenzierten Betrachtung von Subgruppen mit geringer Fallzahl könnten jedoch im Extremfall negative Durchschnittswerte ausgewiesen werden. Es ist daher bei differenzierten Auswertungen stets auf eine ausreichende Fallzahl zu achten.

Folgende Variablen sind in separaten Datensätzen erhältlich:

- Sondervariablen
- Regionalvariablen
- Volltextvariablen mit offenen Angaben

Sondervariablen:

missant	Anteil der mit „weiß nicht / Verweigert“ beantworteten Fragen
intrn	Interviewerkennung
intdat_t	Befragungstag
intdat_m	Befragungsmonat

intdat_j Befragungsjahr (immer 2016)

Regionalvariablen:

kkz Kreiskennziffer

aabez Arbeitsamtsbezirke

2. Datenaufbereitung / Plausibilisierung

Für die BIBB-seitigen Auswertungen der Daten wurden einige Datenaufbereitungsmaßnahmen / Plausibilisierungen durchgeführt. Diese werden grundsätzlich auch anderen Nutzern der Daten empfohlen.

Unter anderem wurden Doppelzählungen von Lernaktivitäten anhand der genannten Stichworte soweit möglich eliminiert, z. B. wenn offenkundig dieselbe Aktivität in zwei verschiedenen Kategorien genannt wurde (z. B. non-formales Lernen und arbeitsnahe Qualifizierung). Eine weitere wichtige Korrektur betrifft die Zählung von Einzelterminen länger angelegter oder gar unbegrenzter Weiterbildungsaktivitäten. Ein Beispiel sind ärztliche Qualitätszirkel, die fortlaufend an regelmäßigen Terminen stattfinden. Mitunter zählten die Befragten hier jede Sitzung als Einzelaktivität, gaben aber ihre Kosten und den Zeitaufwand für die gesamte Weiterbildung bzw. das ganze Jahr 2015 an. Da für die Hochrechnung die Zahl der Einzelaktivitäten genutzt wird, würde dies zu einer Überschätzung des tatsächlichen Aufwands führen. Soweit erkennbar haben wir solche Einzelzählungen im Zuge der Datenaufbereitung bereinigt und die Anzahl der Aktivitäten auf eine oder zwei reduziert. Dadurch wurde die Gesamtzahl der erfassten 17.902 Aktivitäten um 2.813 reduziert (entspricht 15,7 %). Es ist jedoch davon auszugehen, dass einige Fälle unentdeckt blieben.

Auf Anfrage ist beim BIBB.FDZ eine Dokumentation dieser und anderer Datenaufbereitungsmaßnahmen erhältlich.

Literatur

Müller, N.; Wenzelmann, F. (2018a): Berufliche Weiterbildung: Aufwand und Nutzen für Individuen, Ergebnisse einer BIBB-Erhebung. BIBB Report 2. Bonn 2018 - URL:
<https://www.bibb.de/veroeffentlichungen/de/publication/show/8931>

Müller, N.; Wenzelmann, F. (2018b): BIBB-Erhebung "Berufliche Weiterbildung - Aufwand und Nutzen für Individuen" (ANI 2015), Methoden und tabellarische Auswertungen. URL:
<https://www.bibb.de/veroeffentlichungen/de/publication/show/8931>

Müller, Normann; Wenzelmann, Felix (2018): BIBB-Befragung: Berufliche Weiterbildung – Aufwand und Nutzen für Individuen 2015. suf_1.0; Forschungsdatenzentrum im BIBB (Hrsg., Datenzugang); Bonn: Bundesinstitut für Berufsbildung. doi:10.7803/680.16.1.1.10