

LIAB-Datenhandbuch

Version 1.0

Holger Alda, Dagmar Herrlinger

Inhaltsverzeichnis

1. Einleitung	5
1.1. <i>Die Linked-Employer-Employee-Daten des IAB, LIAB</i>	8
2. Veränderungen des Datensatzes und der Datensatzbeschreibung im Vergleich zu früheren Versionen	17
3. Kurzbeschreibung der LIAB-Versionen, die derzeit im FDZ für Auswertungen zur Verfügung stehen	22
4. Variablenbeschreibung	26
4.1. <i>Allgemeiner Teil</i>	27
4.2. <i>Beschreibung der einzelnen Variablen in den LIAB-Personendatenbanken</i>	29
4.2.1. Abgabegrund	31
4.2.2. Abmeldegrund aus dem Leistungsbezug (Zusammenfassung)	33
4.2.3. Alter der Person	34
4.2.4. Arbeitsamtbezirk des Arbeitsortes	35
4.2.5. Kreiskennziffer des Arbeitsortes	36
4.2.6. Ausbildung (B2-Schlüssel)	36
4.2.7. Beginndatum	39
4.2.8. Stellung im Beruf (B 1 – Schlüssel)	40
4.2.9. Berufskennziffer	41
4.2.10. Status der Betriebsnummer	43
4.2.11. Bewilligungsart	44
4.2.12. Datumsbereinigung	44
4.2.13. Dauer	46
4.2.14. Eintrittsdatum Betrieb	46
4.2.15. Eintrittsdatum in das Erwerbsleben	47
4.2.16. Enddatum	48
4.2.17. Währungskennzeichen	49
4.2.18. Familienstand der Person	49
4.2.19. Geschlecht	50
4.1.20. Personenidentifikator	50
4.1.21. Betriebsidentifikator	51
4.2.22. Kinder	52
4.2.23. Leistungsart	52
4.2.24. Episodenbezogener Zähler	53
4.2.25. Nationalität	54
4.2.26. Personengruppe	55
4.2.27. Quelle	56

4.2.28. Tagesentgelt.....	56
4.2.29. Ost-West-Kennzeichen	59
4.2.30. Arbeitsamt am Wohnort	59
4.2.31. Kreiskennziffer des Wohnortes	60
4.2.32. Wirtschaftszweig nach BA-Klassifizierung 1973.....	60
4.2.33. Wirtschaftszweig nach BA-Klassifizierung 1993.....	61
4.3. Logische Zusammenhänge zwischen Variablen.....	62
4.3.1. Logische Zusammenhänge in den LIAB-Personendatenbanken	63
4.3.1.1. Berufsstellung und Personengruppen	63
4.3.1.2. Variablen quelle, beginn und ende (nur LIAB-Längsschnittmodell).....	64
4.3.1.3. Vollzeitbeschäftigung und Tagesentgelt.....	65
4.3.1.4. Abgabegrund und Dauer (nur Beschäftigtenmeldungen)	66
4.3.2. Logische Zusammenhänge zwischen Variablen des IAB-Betriebspanels und von auf die Betriebsebene aggregierten LIAB-Personendaten	67
4.3.2.1. Ost/Westkennzeichen in Personendatenbanken und im IAB-Betriebspanel.....	68
4.3.2.2. Lohnsumme im IAB-Betriebspanel und in den Personendaten.....	69
4.3.2.3. Wirtschaftszweigklassifizierungen	71
4.3.2.4. Anzahl sozialversicherungspflichtig Beschäftigter.....	71
4.3.2.5. Personalstruktur.....	72
5. Datenquellen	74
6. Datenaufbereitung.....	75
6.1. <i>Das Episodensplitting</i>	76
6.1.1. Kontenbeispiel für das Episodensplitting	77
6.2. <i>Bereinigung von Versicherungsnummern und Konsequenzen für die Aktualisierung der LIAB-Versionen</i>	79
7. Datenanonymisierung.....	80
7.1. <i>Datenschutzbestimmungen</i>	80
7.2. <i>Die schwache Anonymisierung der LIAB-Originaldaten</i>	82
7.2.1. Anonymisierung der Sozialversicherungsnummer.....	82
7.2.2. Anonymisierung der Betriebsnummer.....	83
8. Datenqualität und Probleme.....	85
9. Testdaten und Hinweise für Datenauswertungen	89
9.1. <i>Hinweise und Anregungen für die Datenauswertungen</i>	89
9.1.1. Grundsätzliches zur Verknüpfung der Betriebs- und Personendaten.....	89
9.1.2. Planung und Umsetzung der Forschungsinhalte	90
9.1.3. Umgang mit den Personendatenbanken	91

9.1.4. Varianz der Betriebsmerkmale nach der Verknüpfung mit den Personendaten und Gewichtung	92
9.1.5. Hinweise zu den Variablen in den LIAB-Personendatenbanken	92
9.2. <i>Beispieldaten und -programme</i>	94
9.2.1. Aggregation der Personendaten auf die Betriebsebene	94
9.2.2. Bestimmung von Ein- und Austrittsraten	97
9.3. <i>Beschreibung und Umgang mit den Testdaten sowie weitere Arbeitshilfen</i>	101
9.3.1. Beschreibung der Testdaten des LIAB-Querschnittmodells	101
9.3.2. Beschreibung der Testdaten des LIAB-Längsschnittmodell	103
9.3.3. Der Paneldatensatz des IAB-Betriebspanels	105
10. Abkürzungsverzeichnis	106
11. Literaturverzeichnis für das LIAB-Datenhandbuch	107
12. Literaturliste der LIAB-Anwendungen	109
Anhang	113

Datenverfügbarkeit

Die in diesem Beitrag beschriebenen Datensätze sind der Fachöffentlichkeit zugänglich. Nähere Informationen dazu befinden sich auf der Internetseite: <http://fdz.iab.de/> unter der Rubrik „Integrierte Betriebs- und Personendaten“. Stand der Dokumentation ist das 3. Quartal 2005.

Danksagung

Das ITM-Team Datenmanagement des Instituts für Arbeitsmarkt- und Berufsforschung, IAB, hat die für die Erstellung der LIAB-Personendatenbanken erforderliche Dateninfrastruktur des IAB bereitgestellt. Zudem wird für die in diesem Rahmen erfolgte datentechnische Beratung und Unterstützung gedankt. Die Dokumentationen zum Inhalt der einzelnen Variablen in den Personendatenbanken aus dem zweiten Abschnitt des vierten Kapitels bauen teilweise auf einem von ITM konzipierten Datendokumentationssystem für die Basisdatenquelle der Personendaten auf. Dem Forschungsbereich 6 des IAB „Betriebe und Beschäftigung“ wird für die Bereitstellung der IAB-Betriebspaneldaten und die Unterstützung bei inhaltlichen Fragen gedankt. Anmerkungen, Fragen und Kritik zu diesem Datenreport werden bevorzugt vom Autor entgegengenommen.

1. Einleitung

Das vorliegende LIAB-Datenhandbuch erläutert und beschreibt die Linked-Employer-Employee-Daten des Institut für Arbeitsmarkt- und Berufsforschung (IAB), wie sie über das Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im IAB für wissenschaftliche Auswertungen externen Forschern und Forscherinnen¹ zur Verfügung stehen. Neben der Darstellung des Datengeneseprozesses wird auf den Inhalt und den Füllungsgrad einzelner Variablen mit Informationen über Personen eingegangen. Außerdem werden praktische Hinweise für die Arbeit und den Umgang mit den Daten gegeben. Informationen zum Datenzugang hingegen werden in diesem Handbuch nicht behandelt. Sie sind auf den entsprechenden Internet-Seiten des FDZ (<http://fdz.iab.de/pageText.asp?PageID=76>) abrufbar.

Im Verlauf dieses Buchs werden immer wieder Anregungen für das Schreiben von Auswertungsprogrammen in den Programmiersprachen STATA und wenn möglich auch SPSS gegeben. Sie sind im Anhang in der Reihenfolge ihres Auftauchens im Text abgedruckt. Die Beschränkung auf diese beiden Programmiersprachen sind keine Auswertungsrestriktionen. Für die Arbeit mit den FDZ-Daten ist auch die Anwendung anderer Softwareanalysetools wie beispielsweise SAS oder TDA möglich. Nutzerbefragungen zeigen, dass LIAB-Analysen sehr häufig mit dem Auswertungsprogramm STATA durchgeführt werden. In geringerem Umfang wird auch das Programmpaket SPSS genutzt. Anwender anderer Programmiersprachen wie beispielsweise SAS und TDA hingegen haben bisher relativ selten ihr Interesse an der Arbeit mit LIAB-Daten bekundet. Daher werden die meisten Anwenderbeispiele in der Programmiersprache STATA formuliert. In Abschnitt 9.2.1. befindet sich auch ein Auswertungsbeispiel in SPSS. SPSS verfügt nicht über gerade bei der Arbeit mit verlaufsorganisierten Personendaten² wichtige Systemvariablen (wie beispielsweise „_n“). Die Programmier Techniken zur Erreichung eines gleichen Auswertungsziels wie in den STATA-Programmbeispielen ist im Einzelfall schwierig.

¹ Fortfolgend meinen Bezeichnungen von Personen und Personengruppen immer sowohl männliche wie weibliche.

² Verlaufsorganisierte Personendaten sind in der Zeit möglichst vollständig vorliegende Beschäftigungs- und Leistungsmeldungen von einzelnen Personen, aus denen sich ihre Erwerbsverläufe für einen bestimmten Zeitraum rekonstruieren lassen.

Bei allen Auswertungsvorschlägen und den Programmbeispielen in diesem Datenreport handelt es sich selbstverständlich nur um Anregungen, die keineswegs für eine bestimmte Anwendung die einzig mögliche Umsetzung darstellen oder das Auswertungsspektrum der LIAB-Daten auch nur annähernd abdecken. Im Rahmen aller bisher durchgeführten FDZ-Veranstaltungen und Nutzerbefragungen äußerten die an LIAB-Daten Interessierte immer wieder den Wunsch, Programmierbeispiele für die konkrete Arbeit mit den Daten einsehen zu können. Das wird in diesem Handbuch anhand einiger ausgewählter Beispiele aufgegriffen und umgesetzt. An den relevanten Stellen im Handbuch befinden sich Zeichen wie beispielsweise eine „❶“. Im Anhang sind dann entsprechende Programmcodes in der Reihenfolge ihres Auftauchens in diesem Datenreport dokumentiert.

Das LIAB-Datenhandbuch ist wie folgt gegliedert. Im nächsten Unterabschnitt und im nächsten Kapitel werden nach einer allgemeinen Einführung in das Datenmaterial die LIAB-Datenmodelle und Versionen beschrieben³. Das dritte Kapitel stellt die wichtigsten Informationen zu den derzeit zugänglichen LIAB-Daten zusammen. Das vierte Kapitel beschreibt ausführlich den Inhalt und die Datenqualität der einzelnen Variablen in den LIAB-Personendatenbanken. In einem weiteren Abschnitt werden die logischen Zusammenhänge zwischen einzelnen Variablen behandelt. Sie werden unterteilt in logische Abhängigkeiten zwischen einzelnen Ausprägungen nur der Personendaten und in solche, die das Verhältnis der Betriebs- zu den Personeninformationen bei inhaltlich gleichen oder ähnlichen Sachverhalten beleuchten⁴.

Das fünfte Kapitel (Datenquellen) und das sechste (Datenaufbereitung) vertieft die Informationen zu den einzelnen Variablen durch die Beschreibung des Datengenerierungsprozesses und Bereinigungs-schritten. Kapitel sieben beschreibt die technische Umsetzung der datenschutzrechtlichen Vorgaben durch die schwache Anonymisierung der Originaldaten vor Ort. Im achten Kapitel wird kritisch auf die Verknüpfungsqualität der LIAB-Daten eingegangen⁵. Das neunte Kapitel fokussiert auf das Schreiben von Auswertungsprogrammen. Im ersten Abschnitt werden allgemeine Empfehlungen im Umgang mit den Daten gegeben, beispielsweise wie Auswertungsprogramme organisiert und effektiv aufgebaut werden können. Im zweiten Ab-

³ Sie geben in wesentlichen Teilen Auszüge aus den FDZ-Datenreporten 1/2005 bis 3/2005 wieder.

⁴ So liegt beispielsweise die Angabe für den Arbeitsort (Ost- oder Westdeutschland) in Betriebs- wie Personendaten vor.

⁵ Das Kapitel ist ein Auszug aus dem FDZ-Methodenreport Nr. 1/2005.

schnitt befinden sich ausgewählte Programmierbeispiele anhand fiktiver Forschungsfragen. Im dritten Abschnitt des neunten Kapitels werden die Testdaten und der Umgang mit Ihnen beschrieben. Die letzten Kapitel runden mit dem Abkürzungsverzeichnis, der Liste der derzeit mit den LIAB-Daten erstellten Forschungsarbeiten und dem Literaturverzeichnis dieses LIAB-Datenhandbuch ab.

Zu diesem LIAB-Handbuch gehören zwei Anlagebände. In den LIAB-Codeplänen befinden sich Listen mit allen Merkmalsausprägungen der einzelnen Variablen in den Personendaten. Dort sind auch die Ausprägungen von denjenigen Variablen dokumentiert, die im zweiten Abschnitt des vierten Kapitels aus Platz- und Übersichtsgründen nicht einzeln aufgelistet sind. Die Häufigkeitsauszählungen der Originaldaten befinden sich im LIAB-Codebuch, das als Anhangsband ebenfalls ein Teil dieses LIAB-Datenhandbuchs ist und unter der gleichen Internetadresse wie dieses Buch Interessierten zur Verfügung steht. Der folgende Abschnitt beschreibt ausführlich, wie die LIAB-Daten konstituiert sind und welche grundsätzlichen Verknüpfungsmöglichkeiten sich daraus ergeben.

1.1. Die Linked-Employer-Employee-Daten des IAB, LIAB

Linked-Employer-Employee-Daten im **IAB** (LIAB) sind aufeinander bezogene Arbeitgeber-Arbeitnehmer Daten. Sie bestehen aus zwei Einzelteilen, den Betriebsdaten mit den jährlichen Erhebungswellen des IAB-Betriebspanels und ein bis mehreren Personendatenbanken. Durch die Verknüpfung beider Datenquellen entstehen Linked-Employer-Employee-Daten. Die Personendatenbanken entsprechen in ihren Strukturen, dem Abdeckungsgrad und Aufbau sowie im Inhalt einzelner Variablen gängigen IAB-Personendaten, insbesondere der schwach anonymisierten Version der IAB-Beschäftigtenstichprobe. In diesem LIAB-Datenhandbuch wird bei dem IAB-Betriebspanel nur auf Aspekte eingegangen, die im Zusammenhang mit der Verknüpfung mit Personendaten von Bedeutung sind. Vertiefende Informationen für das IAB-Betriebspanel stellt das FDZ auf Internet-Seiten zur Verfügung.

Abowd/Kramarz (1999) geben einen Überblick über die grundsätzlichen Verknüpfungsmöglichkeiten von Betriebs- und Beschäftigendaten. Sie unterscheiden u.a. zwischen Verknüpfungen ausschließlich administrativer Daten, ausschließlich von Befragungsdaten und diversen Mischformen. Für die hier beschriebenen LIAB-Daten geben Alda/Bender/Gartner (2005) einen umfassenderen Überblick über mögliche Anwendungsgebiete und Anwenderbeispiele.

Die Verknüpfungsmöglichkeiten der LIAB-Daten im Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung ergeben sich aus den beiden Elementen der LIAB-Daten, den Betriebs- und Personendaten. Die Personendaten stammen aus der Beschäftigten-Leistungsempfänger-Historik-Datei (BLH) des IAB. Die Beschäftigtenhistorik als ein Teil dieser Datei deckt dabei die sozialversicherungspflichtige Beschäftigung ab. Für einzelne Arbeitnehmer geben Arbeitgeber mindestens einmal im Jahr eine Meldung ab. Diese Meldungen enthalten u.a. für Beschäftigte

- ihr Alter, das Geschlecht und die Nationalität
- das Beginn- und Endedatum jeder Beschäftigtenmeldung
- den ausgeübten Beruf auf der Dreisteller-Ebene
- den tagesgenauen Lohn zensiert an der oberen und unteren Beitragsbemessungsgrenze
- das Bildungsniveau in Verbindung mit dem beruflichen Qualifikationsniveau

- den Wirtschaftszweig
- Kennziffern zum Arbeitsort
- einen eindeutigen Betriebsidentifikator

Indirekt lassen sich auch Beschäftigte in Freistellungszeiten identifizieren. Diese haben in der Variable Tagesentgelt einen Wert von Null eingetragen. Die Beschäftigtenhistorik-Datei wird im IAB mit Leistungsempfangsdaten, die aus Geschäftsstatistiken der Bundesagentur für Arbeit erstellt wird, zur *Beschäftigten-Leistungsempfänger-Historik-Datei* (BLH) verknüpft. Zeitliche Lücken der Beschäftigtenmeldungen werden durch Meldungen aus der *Leistungsempfänger-Historik-Datei* (LeH) des IAB aufgefüllt. Dann noch verbleibende zeitliche Lücken sind mit den vorliegenden Daten nicht beobachtbar („nicht gemeldete Zeiträume“, NGZ).

In der LeH werden Arbeitslosigkeitszustände nur erfasst, wenn Personen Leistungen der Bundesagentur für Arbeit in Anspruch nehmen. Gemeldete Arbeitslosigkeit ohne Anspruch auf Bezüge (z.B. nach einem Studium, bei Hilfe zum Lebensunterhalt oder nach Selbstständigkeit) ist in der LeH derzeit nicht beobachtbar. Die Leistungsempfänger-Historik-Datei enthält u.a. folgende Informationen:

- Alter, Geschlecht, Nationalität, Familienstand und Kinder (ja/nein) der Person
- Kreiskennziffer des zahlenden Arbeitsamtes
- Art und Höhe der Leistung
- Beginn und Ende jeder Leistungsmeldung
- Bewilligungsart
- Abgabegrund der Meldung
- Beendigungsgrund
- den Beruf auf der Viersteller-Ebene

Leistungsmeldungen erfolgen nur bei leistungsrechtlich relevanten Veränderungen. Eine einzelne Leistungsmeldung kann also im Gegensatz zu den Beschäftigtenmeldungen einen längeren Zeitraum als ein Jahr beschreiben.

Auf der Betriebsseite steht als Datenbasis das IAB-Betriebspanel zur Verfügung. Das IAB-Betriebspanel ist eine im Jahr 1993 erstmals und seitdem jährlich durchgeführte Arbeitgeberbefragung, zunächst in etwa 4000 westdeutschen, seit 1996 auch in etwa 4300 ostdeutschen Betrieben. Ab dem Jahr 2000 sind auf nahezu jeder Bundeslandebene repräsentative Auswertungen möglich, denn seitdem finanzieren auch west-

deutsche Bundesländer entsprechende Aufstockungsstichproben. Im Jahr 2001, dem letzten Beobachtungsjahr der derzeit vorliegenden Personendaten, haben mehr als 15.000 Betriebe an der Befragung teilgenommen.

Das IAB-Betriebspanel enthält u.a. die folgenden Informationen:

- Beschäftigtenzahl (auch für verschiedene Qualifikationsgruppen)
- Anzahl befristet Beschäftigter und Leiharbeiter im Betrieb
- Wochenarbeitszeit für Vollzeitbeschäftigte und Überstunden
- Tarifbindung des Betriebs, Existenz eines Betriebsrat
- Umsatz, Vorleistungen und Exportanteil
- Investitionssumme
- Gesamtlohnsumme im Juni des Befragungsjahres
- technischer Stand der Betriebsanlagen
- Betriebsalter, Rechtsform und Unternehmensstellung
- Beurteilung der betrieblich-wirtschaftlichen Gesamtsituation
- Reorganisationsmaßnahmen und betriebliche Weiterbildungsaktivitäten (in mehrjährigen Abständen)
- Wirtschaftszweigzugehörigkeit

Den Ansatz und Aufbau des IAB-Betriebspanels beschreibt beispielsweise Bellmann (2002). Die Grundgesamtheit der Betriebe in Deutschland ist in der Betriebsdatei des IAB erfasst. Aus ihr wird eine bezüglich der Betriebsgröße disproportionale Stichprobe für das IAB-Betriebspanel gezogen⁶, die Großbetriebe überrepräsentiert. Durch Hochrechnung der Analyseergebnisse erfolgt der Schluss auf die Grundgesamtheit⁷.

Die Tabelle 1 zeigt beispielhaft die Ziehungswahrscheinlichkeiten nach der Betriebsgröße in der ersten Befragungswelle 1993.

⁶ Die Schichtungsmatrix berücksichtigt neben zehn Betriebsgrößenklassen 16 Wirtschaftszweige bis zum Jahr 1999 und 20 Wirtschaftszweige ab dem Jahr 2000.

⁷ Auch deskriptive Analyseergebnisse mit den LIAB-Personendaten können mit den Hochrechnungsfaktoren des IAB-Betriebspanels gewichtet werden.

Tabelle 1: Ziehungswahrscheinlichkeiten und Rücklaufquote des IAB-Betriebspanels in der ersten Befragungswelle 1993 (nur westdeutsche Betriebe)

Beschäftigtenzahl	gezogene Betriebe	Auswahlwahrscheinlichkeit	antwortende Betriebe	Rücklaufquote
1 - 4	1,072	.0011	625	.67
5-9	431	.0015	250	.64
10-19	466	.0030	299	.71
20-49	862	.0089	542	.70
50-99	535	.0153	350	.72
100-199	543	.0304	376	.77
200-499	923	.0862	615	.74
500-999	479	.1504	304	.71
1.000 – 4.999	1,497	.8765	924	.72
5.000 +	115	.9127	71	.73
insgesamt	6,923	.0043	4,356	.71

Quelle: IAB-Betriebspanel 1993, Kölling (2000), S. 294

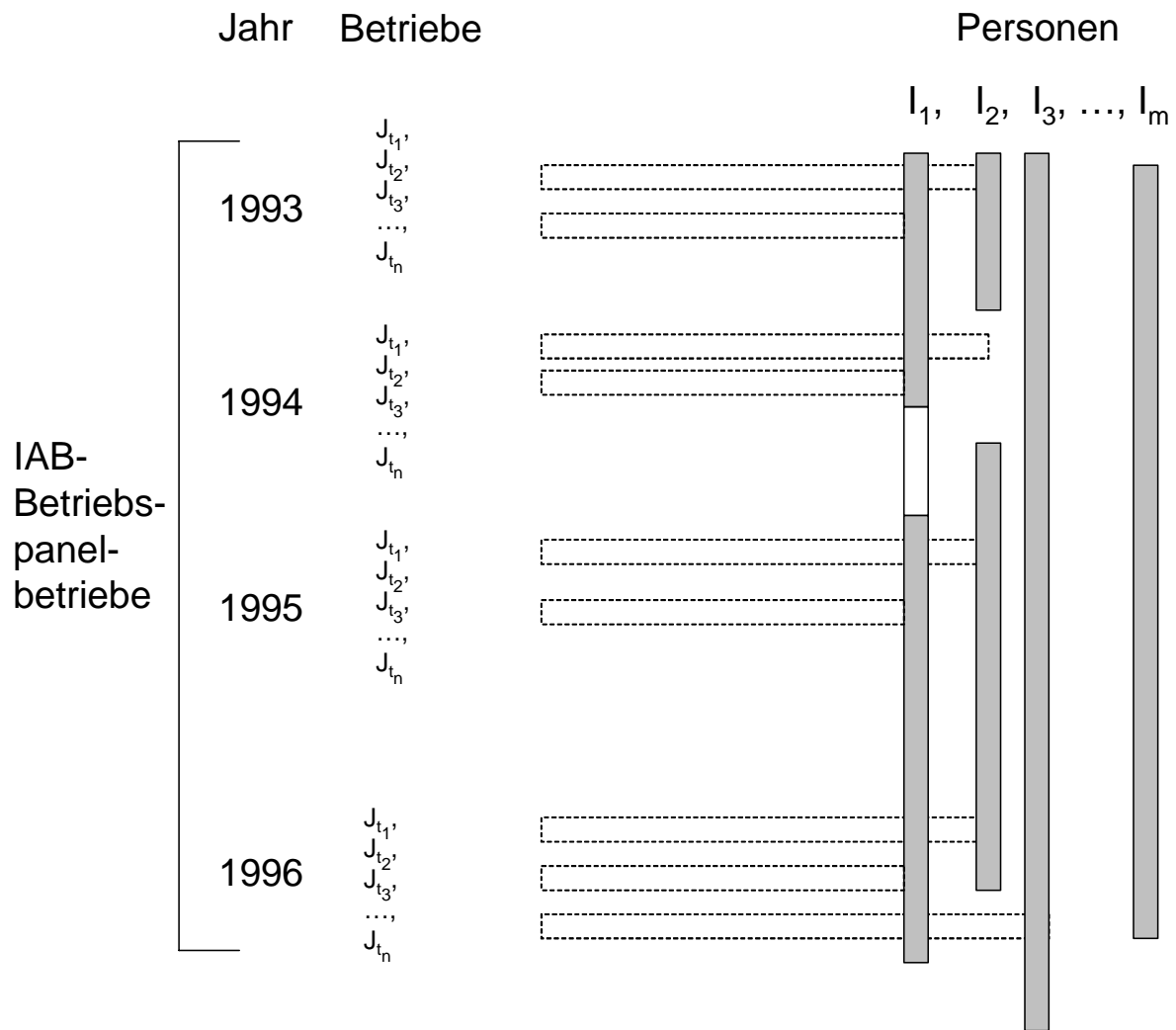
An den Ziehungswahrscheinlichkeiten und Rücklaufquoten ändert sich im Verlauf der Jahre relativ wenig⁸. Die Fallzahl des IAB-Betriebspanels beträgt im Jahr 2001 beispielsweise mehr als 15500 Betriebe (etwa ein Prozent aller Betriebe), in denen etwa zehn Prozent aller sozialversicherungspflichtig Beschäftigten in Deutschland arbeiten. Entsprechende Personendatenbanken erreichen also im Zeitverlauf beachtliche Größen. Zwei Datenmodelle werden im Folgenden näher beschrieben.

Zum einen können alle Beschäftigungs- und Leistungsmeldungen für einen bestimmten Zeitraum in Personendatenbanken abgelegt werden. Sie umfassen derzeit maximal das Jahr 1990 bis zum aktuellen Rand. Für jedes Individuum gibt es – je nach gemeldeten Vorgängen – eine beachtliche Anzahl an einzelnen Meldungen oder Spells. Zum anderen besteht die Möglichkeit, den jeweils zu einem bestimmten Zeitpunkt eines jeden Jahres gültigen Beschäftigungsspell aus der BLH zu extrahieren. Meistens handelt es sich um eine Beschäftigung am 30. Juni in einem der IAB-Betriebspanelbetriebe, weil dieses Datum der Referenzzeitpunkt der meisten Fragen des Fragebogens an die Betriebe ist. Variablen in der Personendatenbank, die Leis-

⁸ Ausnahme sind die größten Betriebe. Panelmortalität kann in den beiden obersten Größenklassen kaum ausgeglichen werden. Die Anzahl der verknüpfbaren Personendaten bzw. -konten nimmt daher bei einer steigenden Anzahl von Betrieben aufgrund großbetrieblicher Panelmortalität ab bzw. bleibt trotz diverser Aufstockungsstichproben auf der Betriebsseite nahezu unverändert.

tungsbezug durch die Bundesagentur für Arbeit abbilden, werden bei der letztgenannten Datenstruktur obsolet. Mit diesen so genannten *Jahresscheiben* wird die Anzahl der einzelnen (Personen-) Beobachtungen deutlich reduziert. In der Abbildung 1 werden die beiden angesprochenen Modelle graphisch verdeutlicht.

Abbildung 1: Betriebe und Beschäftigte in LIAB-Daten



Die horizontalen Linien stellen befragte IAB-Betriebspanelbetriebe in dem jeweiligen Jahr dar, die vertikalen Personen. Grau hinterlegte Balkenabschnitte kennzeichnen Zeiten der Beschäftigung und hell hinterlegte gemeldete Arbeitslosigkeit. Ein unterbrochener Balken bedeutet, dass die Person weder ein sozialversicherungspflichtiges Beschäftigungsverhältnis ausübt noch Leistungen der BA in Anspruch nimmt.

Eine Jahresscheibe ist der Kreuzungspunkt der horizontalen und vertikalen Balken, d.h. die dazu gehörige Beschäftigtenmeldung zum 30. Juni. Für das Jahr 1996 stellt der längere untere horizontale Balken einen ostdeutschen Betrieb dar, der in diesem Jahr erstmalig befragt wird. Individuum I_1 lässt sich somit in jeder Jahresscheibe 1993 bis 1996 beobachten, eventuell in verschiedenen Betrieben. Der Leistungsbezug ist auf der Datenebene nicht identifizierbar. Person I_2 hat eine unterbrochene sozialversicherungspflichtige Erwerbsbiographie und kann im Jahr 1994 nicht beobachtet werden. Die Person I_3 ist in einem ostdeutschen Betrieb beschäftigt und daher erstmalig im Jahr 1996 beobachtbar. Der Balken für I_m soll graphisch verdeutlichen, dass es durchaus auch Beschäftigte gibt, die gar nicht in einem der IAB-Betriebspanelbetriebe auftauchen. Leistungsmeldungen bzw. die entsprechenden Variablen werden in Personendatenbanken beim Jahresscheibenmodell obsolet, denn Leistungsbezug ist durch die Bedingung, am 30. Juni in einem IAB-Betriebspanelbetrieb beschäftigt zu sein, unmöglich.

Der Unterschied zu einer Personendatenbank mit allen Beschäftigten- und Leistungsmeldungen einer Person besteht darin, dass auch Nicht-Kreuzungspunkte und die entsprechenden Erwerbszustände bestimmt werden können, also beispielsweise gemeldete Arbeitslosigkeit mit Leistungsbezug. So ist für Person I_1 die Arbeitslosigkeitsphase identifizierbar und für Person I_2 bekannt, dass sie im Jahr 1994 eine Erwerbsunterbrechung hat. Das ist mit Jahresscheiben nicht beobachtbar. Die Kreuzungspunkte einer Person und eines Betriebes sind, wenn alle Beschäftigten- und Leistungsmeldungen in der Personendatenbank vorliegen, das Auswahlkriterien für die Aufnahme der entsprechenden Personenkarten. Es handelt sich nicht zwingend um den 30. Juni eines jeden Befragungsjahres, sondern oft ist die Bedingung für die Aufnahme einer Person und ihrer zugehörigen Meldungen eine mindestens eintägige Beschäftigung in einem der IAB-Betriebspanelbetriebe in einem zu spezifizierenden Zeitraum.

Wenn die Arbeit mit Personendatenbanken nicht möglich oder gewünscht ist, besteht eine weitere Analysemöglichkeit. Die Jahresscheibeninformationen der Beschäftigten können auf die Betriebsebene aggregiert werden, beispielsweise als Summen oder Anteile. Auf diese Weise werden Informationen auf der Betriebsebene dann auch ohne einen weiteren Rückgriff auf die Personendatenbanken verfügbar. Sie eignen sich zur Ergänzung der Daten des IAB-Betriebspanels. Im IAB-Betriebspanel gibt es bei-

spielsweise in den meisten Jahren keine Information über den Anteil der Hochschulabsolventen an der Gesamtbelegschaft. Auch über die Alterszusammensetzung der Belegschaft sind bis zum Jahr 2002 keine genauen Informationen verfügbar. Analysen auf der Betriebsebene, die solche Betriebscharakteristika nutzen, werden mit Aggregation der interessierenden Merkmale auf die Betriebsebene und der Hinzuspielung der IAB-Betriebspaneldaten möglich. Auf die Aggregation von Personendaten auf die Betriebsebene wird in Abschnitt 9.2.1. eingegangen.

Die anderen beiden Gruppen von Daten (Jahresscheiben und Verlaufsdaten von Personen) werden im Folgenden *LIAB-Querschnittsmodelle* und *LIAB-Längsschnittmodelle* genannt, wobei sich die Bezeichnungen auf die zugehörigen Personendatenbanken beziehen. LIAB-Querschnittmodelle sind also immer Personendatenbanken mit Beschäftigtenmeldungen zum 30. Juni eines jeden Befragungsjahres (in einem IAB-Betriebspanelbetrieb). LIAB-Längsschnittmodelle sind Personendatenbanken, in denen alle Beschäftigten- und Leistungsmeldungen für Personen in einem bestimmten Zeitraum vorliegen.

In den LIAB-Quer- und Längsschnittmodellen findet die Verknüpfung der Betriebs- und Personendaten in einem weiteren Arbeitsschritt statt, wo die (aufbereiteten) Betriebsinformationen zu den Personendaten dazugespielt werden. Die Tabelle 2 stellt allgemeine Eigenschaften beider Gruppen von Personendatenbanken zusammen. Die Aufstellung in Tabelle 2 wurde mit Rücksicht auf verfügbare Betriebsinformationen aus dem IAB-Betriebspanel (z.B. Alter des Betriebs) und der Informationen zu Personen aus der Beschäftigten-Leistungsempfänger-Historik-Datei vorgenommen und benennt die Quellen häufig verwendeter Betriebs- und Beschäftigtenmerkmale.

Tabelle 2: Charakteristika potenzieller Linked-Employer-Employee-Datensätze als Verknüpfung von IAB-Personendaten mit Informationen aus dem IAB-Betriebspanel

Nr.	Charakteristika	LIAB-Querschnittmodell	LIAB-Längsschnittmodell
1	Betriebs-ID im Personendatensatz	ja	
2	Beschäftigte und Betriebe sind in einem Datensatz vorhanden	nein	
3	Häufigkeit des Links	z.B. jährlich (30. 06.)	tagesgenaue Spells
4	Datenquelle	Beschäftigtenhistorik/ <i>Fragebogen</i>	

(Fortsetzung nächste Seite)

Nr.	Charakteristika	LIAB-Querschnittmodell	LIAB-Längsschnittmodell
5	Betriebs-ID kann überprüft werden	ja	
6	Beobachtungsebene	Betrieb/Beschäftigte gebunden an Betrieb*	Betrieb/ Person
7	Betriebsgröße muss gebildet werden bzw. wird berichtet	Bildung durch Aggregation der Beschäftigtendaten auf die Betriebsebene und Fragebogenangaben	
8	Quelle Betriebsgröße	Beschäftigtenhistorik/ <i>Fragebogen</i>	
9	Alter des Betriebs bekannt	<i>≅ 90 Prozent</i>	
10	Frequenz der Betriebsbeobachtungen	<i>jährlich</i>	
11	Prozent der beobachtbaren Betriebe	<i>(ungewichtet) etwa 1 Prozent</i>	
12	Quelle Geburtsdatum	Sozialversicherungsnummer	
13	Quelle Mann/ Frau	Sozialversicherungsnummer	
14	Quelle schulische Bildung	Meldung der Arbeitgeber an die Sozialversicherungen	
15	Quelle Beruf	Meldung der Arbeitgeber an die Sozialversicherungen	
16	Einheit der Beschäftigtenhistorie	Person gebunden an Betrieb* (Arbeitsplatzkonzept)	Person
17	Beginn des Beschäftigungsverhältnisses immer bekannt	nein	ja (ab 1990)
18	Stellung des Betriebs im Unternehmensverbund bekannt	<i>ja</i>	
19	Frequenz der Personenbeobachtungen	jährlich	tagesgenaue Spells
20	Prozent des beobachtbaren Beschäftigtenuniversums (gleich sozialversicherungspflichtige Beschäftigung)	gewichtet bis zu 90 Prozent, ungewichtet bis zu 10 Prozent	

Erläuterungen: * mit gültigem IAB-Betriebspanel-Interview, kursive Schrift kennzeichnet Informationen, die im IAB-Betriebspanel verfügbar sind

Die ersten beiden Zeilen der Tabelle 2 beziehen sich auf die systemfreien Betriebsidentifikatoren in den Daten⁹. Die Daten des IAB-Betriebspanels können mit dem systemfreien Betriebsidentifikator eindeutig den Personendaten zugeordnet werden. Zeile drei verdeutlicht, dass die Verknüpfung der Personen- und Betriebsdaten eine eigenständig vom Anwender zu erbringende Leistung ist. Die Zeilen 3 bis 8 sowie 16, 17 und 20 fassen das zu den Datenmodellen Gesagte zusammen. Die Zeilen 10 bis

⁹ Systemfreie Identifikatoren sind fortlaufende und *eindeutige* Personen- und Betriebsidentifikationsnummern, mit denen die originalen Sozialversicherungs- und Betriebsnummern ersetzt werden.

12 sowie 18 und 19 beziehen sich auf Merkmale, die das IAB-Betriebspanel bereitstellt bzw. auf Auswirkungen bei der Verknüpfung mit den Personendaten. Den disproportionalen Stichprobenansatz des IAB-Betriebspanels verdeutlicht Zeile 20.

Die LIAB-Quer- und Längsschnittmodelle haben je eigene charakteristische Vor- und Nachteile. Die Tabelle 2 zeigt die grundsätzlichen Eigenschaften solcher Modelle bzw. das maximal Mögliche auf. Es sind auch Mischmodelle denkbar¹⁰. Im nächsten Kapitel werden die grundlegenden Eigenschaften der beiden LIAB-Datenmodelle angesprochen.

¹⁰ Beispielsweise können aus Längsschnitt-Personendaten Informationen für Jahresscheibendaten gewonnen werden. Das wird die zweite Version des LIAB-Querschnittmodells sein.

2. Veränderungen des Datensatzes und der Datensatzbeschreibung im Vergleich zu früheren Versionen

Die Erstellung der LIAB-Daten und ihre Bereitstellung für externe Wissenschaftler ist ein Novum in der deutschen Dateninfrastruktur. Es liegen keine Vergleiche zu vorherigen Versionen oder auf der Datenebene in die Tiefe gehende Datensatzbeschreibungen vor. Das Kapitel beschreibt daher die derzeit verfügbaren und geplanten LIAB-Versionen. Zuerst wird auf die grundlegenden Implikationen und die technischen Restriktionen bei dem LIAB-Quer- und Längsschnittmodell eingegangen.

LIAB-Querschnittmodell: Jahresscheiben (zum 30. Juni, derzeit von 1993 bis 2001)

In Personendatenbanken sind die Daten der Beschäftigten in Betrieben abgelegt, die im jeweiligen Jahr an der Befragung des IAB-Betriebspanels teilgenommen haben. Es handelt sich dabei um die aktuelle Meldung der beschäftigten Personen am 30. Juni eines jeden Befragungsjahres. Darum enthalten die Personendatenbanken nur Daten aus Beschäftigungsmeldungen.

Da in den einzelnen Querschnitten ein eindeutiger Identifikator für Personen und Betriebe vorhanden ist, können die einzelnen Querschnitte miteinander verbunden werden. Somit ist es möglich, Verläufe von Personen in Panelbetrieben zu verfolgen. Ein Betriebswechsel der Person in einen nicht durch das Betriebspanel befragten Betrieb als auch die Nichtteilnahme des Betriebes an der Befragung im Folgejahr führen dazu, dass die Person nicht mehr in dieser Version des LIAB enthalten ist. Extensionen der LIAB-Querschnittsdaten entstehen durch Varianten der Jahresscheiben, beispielsweise durch die Bereitstellung von Erwerbsverlaufsinformationen.

Die Jahresscheiben sind ein Mittelweg in der Balance zwischen verfügbaren und generierbaren Informationen einerseits und der Größe des Datensatzes und der damit benötigten Rechenzeiten andererseits. Mit den Jahresscheiben ist nahezu jeder beliebig mögliche Aufbau von Paneldatensätzen unter den geschilderten Rahmenbedingungen möglich. Die Meldungen in den Jahresscheiben werden durch weitere Kennziffern der Beschäftigten (z.B. erstmaliges Eintrittsdatum in den Beschäftigungsbetrieb) ergänzt.

**LIAB-Längsschnittmodell: auf der Personenebene vollständige Beschäftigten-
Leistungsempfänger-Historikdatei-Daten**

LIAB-Längsschnittmodelle ermöglichen umfassende Analysen der Betriebs- und Beschäftigtenseite. Auf der Betriebsseite sind u.a. auch unterjährige betriebliche Personalanpassungsprozesse (theoretisch täglich) analysierbar. Auf der Beschäftigtenseite gibt es beispielsweise Informationen darüber, ob und wie lange eine Person arbeitslos oder nicht in den Arbeitsmarkt integriert war bzw. ist. Damit eine Verknüpfung mit den Betriebsdaten hergestellt werden kann, muss die Person wenigstens einen Tag in einem IAB-Betriebspanelbetrieb beschäftigt sein.

Die Personendatenbanken erreichen und überschreiten die Kapazitätsgrenze einiger gängiger Statistikprogramme – selbst bei der Bildung von Subsamples¹¹. Die größten theoretisch denkbaren Datenbanken wären nur mit wenigen Auswertungsprogrammen zu bearbeiten (SAS, TDA). Im Unterschied zum LIAB-Querschnittmodell sind in der Datenbank alle Beschäftigten- und Leistungsmeldungen einer Person in einem bestimmten Zeitraum abgelegt. Solche LIAB-Daten erlauben detaillierte Analysen mit individuellen Erwerbsverlaufsinformationen über einen längeren Zeitraum hinweg. Die Variablenaufbereitung kann anspruchsvoll sein und ist oft mit langen Rechen- und Datenaufbereitungszeiten verbunden.

¹¹ Für die Bildung von Subsamples kann sowohl auf der Betriebs- wie Personenebene angesetzt werden. Auf der Betriebsseite kann die Beschränkung auf eine bestimmte Anzahl erfolgen, bei Personendaten kann der Zeitraum 1991 bis 2001 verkürzt werden.

2.1. Die technische Umsetzung der LIAB-Daten im FDZ: von Modellen zu Versionen

Bisher wurde die Beschäftigten- und Betriebsseite der LIAB-Daten sowie ihre Eigenschaften beschrieben. Diese Datenmodelle sind keine real existierenden Datensätze, sondern beschreiben die Bereitstellungsmöglichkeiten und grundsätzlichen Eigenschaften von Personendatenbanken bei einer simultanen Bereitstellung der Informationen aus dem IAB-Betriebspanel. Auch aufgrund technischer Restriktionen (z.B. der Dateigröße) ist eine maximale Umsetzung der Datenmodelle insbesondere beim LIAB-Längsschnittmodell oft nicht möglich. Ebenfalls ist die Verknüpfungsqualität der Betriebs- und Personendaten diskussionswürdig. Nicht immer können IAB-Betriebspanelbetrieben auch Personendaten zugeordnet werden. Im FDZ-Methodenreport Nr. 1 wird das zusammen mit anderen Verknüpfungsproblemen ausführlich diskutiert. Im achten Kapitel dieses Datenhandbuchs werden die wichtigsten Aspekte dieses Methodenreportes wiedergegeben.

LIAB-Versionen sind Personendatenbanken, die unterschiedlichen Auswahlkriterien bezüglich des abgedeckten Zeitraums, des Variablenkanons und der ausgewählten Betriebe unterliegen. Die Sozialversicherungs- und Betriebsnummer wird in jedem Datensatz durch fortlaufende und *eindeutige* Personen- und Betriebsidentifikationsnummern ersetzt (systemfreie Identifikatoren). Weitere Einschränkungen der Variablen und ihrer Inhalte gibt es nicht. Mit der Einrichtung der Gastwissenschaftlerarbeitsplätze im FDZ werden die ersten Versionen der LIAB-Datenmodelle zur Verfügung stehen. Die einzelnen Variablen der Personendatenbanken und Informationen zum Abdeckungsgrad der jeweiligen Version in Relation zum maximalen Abdeckungsgrad über Betriebe und Beschäftigte (vgl. Tabelle 2) werden in eigenen Datenreporten beschrieben. Gastwissenschaftlern soll dieses LIAB-Datenhandbuch auch helfen, den Umgang mit den Gegebenheiten vor Ort einzuschätzen, beispielsweise wenn auf langwierige Rechenvorgänge und auf den effektiven und zeitsparenden Umgang mit den Daten eingegangen wird.

Derzeit stehen im FDZ die drei folgenden Versionen zur Verfügung:

LIAB-Querschnittmodell, Version 1: Bedingung für die Aufnahme der Beschäftigtenmeldung in die Personendatenbank ist die sozialversicherungspflichtige Arbeitstätigkeit in einem IAB-Betriebspanelbetrieb zum 30. Juni. Die Personendatenbank

enthält daher ausschließlich Variablen aus der Beschäftigtenhistorikdatei sowie Informationen zum ersten Eintrittsdatum der Beschäftigten in den Betrieb und in das Erwerbsleben. Austritte aus dem Betrieb führen auf der Personenebene und Austritte aus dem IAB-Betriebspanel auf der Betriebsebene zur Unbeobachtbarkeit der Betriebe und/oder Beschäftigten. Der Dateiumfang einzelner Jahresscheiben beträgt zwischen 140 und 205 MB¹². Inhalt und Aufbau der Daten in den Personendatenbanken werden im FDZ-Datenreport Nr. 2 beschrieben.

LIAB-Längsschnittmodell, Version 1: Für je etwa 2100 ost- und westdeutsche Betriebe liegen alle Beschäftigungs- und Leistungsmeldungen von 1990 bis 2001 vor. Die Bedingung für die Aufnahme des Personenkontos ist eine mindestens eintägige Beschäftigung zwischen 1996 und 2001 in einem Betrieb, der von 1999 bis 2001 in jedem Jahr an der Befragung des IAB-Betriebspanels teilgenommen hat. Der Dateiumfang beträgt für den westdeutschen Personendatensatz 1520 MB, für den ostdeutschen 1150 MB. Inhalt und Aufbau der Daten werden im FDZ-Datenreport Nr. 3 beschrieben.

LIAB-Längsschnittmodell, Version 2: Die Bedingung für die Aufnahme des Personenkontos ist eine mindestens eintägige Beschäftigung zwischen 1997 und 2002 in einem Betrieb, der entweder zur Version 1 des LIAB-Längsschnittmodells gehört und/oder im Zeitraum 2000 bis 2002 kontinuierlich an der Befragung des IAB-Betriebspanels teilgenommen hat. Die Personenkonten liegen für einen Zeitraum von 1993 bis 2002 vollständig vor. Insgesamt handelt es sich um die Beschäftigten in 5293 westdeutschen und 4360 ostdeutschen Betrieben. Die Personendatenbanken liegen für jeden Landesteil getrennt vor. Dank eines neuen Datenorganisationssystems und aufgrund der effizienten Speicherung der Variablen lassen sich diese Personendatenbanken an der Kapazitätsgrenze der IAB-Rechner auch mit STATA[®] verarbeiten. Die Personendatenbanken der Version 2 verlangen von den Anwendern eine gewisse Disziplin bei der Datenaufbereitung, denn es können nur wenige Variablen zu diesen Rohdatensätzen hinzugefügt werden. Es handelt sich um mehr als 28 Mio. westdeutsche und etwa 16 Mio. ostdeutsche Meldungen.

Im Verlauf der ersten Projektphase des Forschungsdatenzentrums werden *voraussichtlich* außerdem folgende Versionen bereitgestellt:

¹² Die Größenangabe bezieht sich wie alle nachfolgenden auf das Softwareanalysetool STATA 8.2. SE.

LIAB-Querschnittmodell, Version 2: Aufgrund der Diskussionen und Anregungen auf dem LIAB-Workshop im November 2004 in Nürnberg wird eine Querschnittsversion erstellt, die Erwerbsverlaufsinformationen der am 30. Juni beschäftigten Personen für vergangene Zeiträume sowie Anschlussereignisse bei einem Betriebsaustritt generiert. Genaueres zur Art und zum Inhalt dieser synthetischen und originalen Zusatzvariablen wird im Internet bekannt gegeben.

LIAB-Längsschnittmodell 2, Version 3: Die Datenbank enthält alle Personendaten für die Längsschnittbetriebe des IAB-Betriebspanels von 1993 bis zum Jahr 2001, also nur westdeutsche Betriebe. Bedingung für die Aufnahme des Personenkontos von 1990 bis 2001 ist eine mindestens eintägige Beschäftigung in einem dieser Längsschnittbetriebe im letztgenannten Zeitraum.

Die Datenreporte der ersten drei Versionen geben Auskunft über den Abdeckungsgrad der jeweiligen Version über Betriebe und Beschäftigte. Anschließend erfolgen Erläuterungen zum Personendatenmaterial (den verfügbaren Variablen). Der FDZ-Methodenreport Nr. 1 beschreibt ausführlich die Datenverknüpfungsqualität der Betriebs- und Personendaten. Die Lektüre dieses Reports wird bei der Absicht einer konkreten Untersuchung empfohlen. Einige Versionsbeschreibungen, insbesondere die Konstruktion der Personendatenbanken, nehmen explizit Bezug auf diesen Methodenreport Nr. 1. Das nachfolgende Kapitel fasst das Gesagte zu den für Auswertungen zur Verfügung stehenden Versionen in Kurzübersichten zusammen.

3. Kurzbeschreibung der LIAB-Versionen, die derzeit im FDZ für Auswertungen zur Verfügung stehen

< Linked-Employer-Employee-Daten des IAB, LIAB-Querschnittmodell, aktualisierte Version 1 >

<i>Inhaltliche Charakteristika</i>	
Themen/Merkmalsgruppen	Personendatenbank mit BeH-Merkmalen und zwei ausgesuchten Informationen zum Erwerbsverlauf von Personen (letztere nur bis 2001 gefüllt), Betriebe wie IAB-Betriebspanel
Untersuchungseinheit	Beschäftigte / Betriebe
Fallzahlen	Betriebe: zwischen 3933 (min) im Jahr 1995 und 14878 (max) im Jahr 2001 Beschäftigte: zwischen 2.026.658 (min) im Jahr 1999 und 2.652.277 (max) im Jahr 1993
Zeitraum	1993 bis 2002
Zeitbezug	jährlich zum 30. Juni
Regionale Gliederung	Repräsentative Auswertungen bis auf Bundeslandebene (ostdeutsche Länder ab 1996, westdeutsche ab 2000)
Gebietsstand	laufende Bereinigungen wie in BLH
<i>Methodische Charakteristika</i>	
Erhebungsdesign	Die Personendatenbank enthält alle Beschäftigtenmeldungen zum 30. Juni in Betrieben, die am IAB-Betriebspanel teilgenommen haben , sonst wie IAB-Betriebspanel (disproportional geschichtete Zufallsstichprobe)
An der Erhebung beteiligte Institutionen	IAB
Frequenz der Datensammlung	einmal jährlich
Dateiformat/-größe	STATA (153 MB bis zu 205 MB), SAS, SPSS, TDA
(Dateiorganisation)	Datensätze liegen nach Jahren getrennt vor

(Fortsetzung nächste Seite)

<i>Datenzugang</i>	Gastwissenschaftlermodell; nach der Bewilligung des Antrages auf Datenzugang gemäß § 75 SGB X und erfolgreichem Aufbau des Analyse-datensatzes im Rahmen eines Gastwissenschaftleraufenthalts auch Datenfernverarbeitung (Nachbereitung des Gastaufenthalts)
<i>Vorteile der Version gegen-über anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - Gesamtbestand der IAB-Betriebspanelbetriebe kann analysiert werden - schneller und relativ unkomplizierter Aufbau von Paneldatensätzen - leichte und unkomplizierte Aggregation der Personendaten auf die Betriebsebene
<i>Nachteile der Version gegen-über anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - Vor- und Nachgeschichte der Personen ist nicht nachvollziehbar - vollständiger Datenverlust auf der Betriebs- und Personenebene bei Panelmortalität - keine Informationen über Personen, die den Betrieb wechseln (Ausnahme: IAB-Turnover¹³)

< Linked-Employer-Employee-Daten des IAB, LIAB-Längsschnittmodell, aktualisierte Version 1 >

<i>Inhaltliche Charakteristika</i>	
Themen/Merkmalgruppen	Personendatenbank mit BLH-Merkmalen, etwa 2100 Betriebe des IAB-Betriebspanel je in Ost- und Westdeutschland
Untersuchungseinheit	Beschäftigte / Betriebe
Fallzahlen	Betriebe: etwa 2100 pro Landesteil Personen: etwa 25.100.000 Meldungen (entspricht 1,9 Mio. Personen)
Zeitraum	Betriebe: mindestens 1999 bis 2001 Beschäftigte: 1991 bis 2002
Zeitbezug	Betriebe jährlich zum 30. Juni Beschäftigte: mindestens eintägige Beschäftigung zwischen 1996 und 2001 in einem der je etwa 2100 Betriebe
Regionale Gliederung	Ost- und Westdeutschland
Gebietsstand	laufende Aktualisierung wie in BLH
<i>Methodische Charakteristika</i>	
Erhebungsdesign	Die Personendatenbanken enthalten alle Beschäftigungs- und Leistungsmeldungen von 1991 bis 2001, wenn die entsprechende Person mindestens einen Tag in einem der je 2100 Betriebe zwischen 1996 und 2001 gearbeitet hat. Diese IAB-Betriebspanelbetriebe haben mindestens zwischen 1999 und 2001 in jedem Jahr ein gültiges Interview.

(Fortsetzung nächste Seite)

¹³ Beim IAB-Turnover wechseln Beschäftigte von einem IAB-Betriebspanelbetrieb in einen anderen.

<i>Methodische Charakteristika</i>	
An der Erhebung beteiligte Institutionen	IAB
Frequenz der Datensammlung	einmalig mit jährlichen Aktualisierungen
Dateiformat/-größe	STATA (2650 MB), SAS (4200 MB)
(Dateiorganisation)	Die Personendatenbanken sind getrennt für die ost- und westdeutschen Betriebe abgelegt. (West: ca. 1,5 GB ; Ost: ca. 1,0 GB)
<i>Datenzugang</i>	Gastwissenschaftlermodell; nach der Bewilligung des Antrages auf Datenzugang gemäß § 75 SGB X und erfolgreichem Aufbau des Analyse-datensatzes im Rahmen eines Gastwissenschaftleraufenthalts auch Datenfernverarbeitung (Nachbereitung des Gastaufenthalts)
<i>Vorteile der Version gegenüber anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - Vor- und Nachgeschichte der Personen ist nachvollziehbar - bei Panelmortalität liegen die Daten für die Betriebe in der Stichprobe auf der Betriebs- und Personenebene für einen gewissen Zeitraum weiterhin vor - Informationen über Personen, die den Betrieb wechseln - keine Konsistenzprobleme der Betriebs- und Personendaten
<i>Nachteile der Version gegenüber anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - Gesamtbestand der IAB-Betriebspanelbetriebe kann nicht analysiert werden - insbesondere die Aufbereitung der Längsschnittdaten für Personen kann anspruchsvoll sein oder werden - lange Rechenlaufzeiten

< Linked-Employer-Employee-Daten des IAB, LIAB-Längsschnittmodell, Version 2 >

<i>Inhaltliche Charakteristika</i>	
Themen/Merkmalgruppen	Personendatenbank mit BLH-Merkmalen, etwa 5300 westdeutsche und 5300 ostdeutsche Betriebe des IAB-Betriebspanels
Untersuchungseinheit	Beschäftigte / Betriebe
Fallzahlen	Betriebe: etwa 5300 (West) bzw. 4300 (Ost) pro Landesteil Personen: <ul style="list-style-type: none"> - West: etwa 28,5 Mio. Meldungen (entspricht 2.358.451 Personen) - Ost: etwa 15,8 Mio. Meldungen (entspricht 1.100.955 Personen)
Zeitraum	Betriebe: mindestens 1999 bis 2001 und mindestens von 2000 bis 2002 Beschäftigte: 1993 bis 2002

(Fortsetzung nächste Seite)

<i>Inhaltliche Charakteristika</i>	
Zeitbezug	Betriebe jährlich zum 30. Juni Beschäftigte: mindestens eintägige Beschäftigung zwischen 1997 und 2002 in einem der ausgewählten Betriebe
Regionale Gliederung	Ost- und Westdeutschland
Gebietsstand	laufende Aktualisierung wie in BLH
<i>Methodische Charakteristika</i>	
Erhebungsdesign	Die Personendatenbanken enthalten alle Beschäftigungs- und Leistungsmeldungen von 1993 bis 2002, wenn die entsprechende Person mindestens einen Tag in einem der ausgewählten Betriebe zwischen 1997 und 2002 gearbeitet hat. Diese IAB-Betriebspanelbetriebe haben mindestens zwischen 1999 und 2001 bzw. von 2000 bis 2002 in jedem Jahr ein gültiges Interview.
An der Erhebung beteiligte Institutionen	IAB
Frequenz der Datensammlung	einmalig mit jährlichen Aktualisierungen
Dateiformat/-größe	STATA West: 2420 MB ; Ost: 1346 MB SAS/SPSS West: 3870 MB ; Ost: 2153 MB
(Dateiorganisation)	Die Personendatenbanken sind nach den ost- und westdeutschen Betrieben abgelegt und können nur getrennt aufbereitet werden.
<i>Datenzugang</i>	Gastwissenschaftlermodell; nach der Bewilligung des Antrages auf Datenzugang gemäß § 75 SGB X und erfolgreichem Aufbau des Analysedatensatzes im Rahmen eines Gastwissenschaftleraufenthalts auch Datenfernverarbeitung (Nachbereitung des Gastaufenthalts)
<i>Vorteile der Version gegenüber anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - Vor- und Nachgeschichte der Personen ist nachvollziehbar - bei Panelmortalität liegen die Daten für die Betriebe in der Stichprobe auf der Betriebs- und Personenebene für einen gewissen Zeitraum weiterhin vor - Informationen über Personen, die den Betrieb wechseln
<i>Nachteile der Version gegenüber anderen LIAB-Versionen</i>	<ul style="list-style-type: none"> - insbesondere die Aufbereitung der Längsschnittdaten für Personen kann anspruchsvoll sein oder werden - sehr lange Rechenlaufzeiten - nur wenige Variablen können in den beiden Personendatenbanken für die Landesteile zum Ursprungsdatensatz hinzugefügt werden (Kapazitätsprobleme durch sehr großen Datensatz)

4. Variablenbeschreibung

Die Variablen in den Personendatenbanken der verschiedenen LIAB-Versionen unterscheiden sich nicht in ihrer inhaltlichen Füllung und werden daher gemeinsam beschrieben. Es gibt dennoch einige Unterschiede zwischen den Quer- und Längsschnittversionen. Einerseits enthalten Versionen des LIAB-Querschnittmodells nur Meldungen aus der BeH, die Versionen des LIAB-Längsschnittmodells zusätzlich die Meldungen aus der LeH. Andererseits sind die Variablen und ihre Inhalte in den beiden Datenmodellen unterschiedlich bedeutsam. Im Querschnittsmodell wird beispielsweise die Variable `spell_nr` kaum Verwendung finden können, in den Versionen der LIAB-Längsschnittmodelle hingegen ist diese Variable wichtig, um die einzelnen Meldungen einer Person (chronologisch) zu ordnen. Auf solche und ähnliche Unterschiede wird bei den Beschreibungen der einzelnen Variablen in Abschnitt 4.2. detailliert eingegangen. Bei den Erläuterungen der einzelnen Variablen werden aufgrund der Versionenvielfalt keine Häufigkeitsauszählungen vorgenommen. Diese sind in einem Anlageband dieses LIAB-Datenhandbuchs dokumentiert, dem LIAB-Codebuch. Im nächsten Abschnitt werden die grundsätzlichen Beziehungen zwischen den einzelnen Variablen vorgestellt, bevor im übernächsten Abschnitt die Einzelbeschreibung aller Variablen der Personendatenbanken erfolgt. Für die datendokumentationen des IAB-Betriebspanels wird auf das entsprechende Internet-Angebot des FDZ verwiesen.

4.1. Allgemeiner Teil

Der Aufbau der Daten ist in allen Versionen des LIAB-Quer- und Längsschnittmodells gleich. Fehlende Werte sind in der Regel als missing gekodet, in Ausnahmefällen werden unbekannte oder fehlende Werte in einzelnen Variablen als eine eigene Ausprägung explizit ausgewiesen. Während „echte“ missings aufgrund der logischen Zusammenhänge zwischen Variablen entstehen, sind Fehlkodierungen oder die Ausprägung „unbekannt“ in der Regel auf Defizite der Beschäftigungs- oder Leistungsmeldungen zurückzuführen. Die beiden Arten von fehlenden Werten werden im Folgenden erläutert.

Echte missings aufgrund logischer Zusammenhänge sind vor allem im LIAB-Längsschnittmodell bedeutsam, weil dort Beschäftigungs- und Leistungsmeldungen enthalten. Bei Beschäftigungsmeldungen macht die Füllung von Merkmalen, die sich ausschließlich auf Leistungsbezug durch die Bundesagentur für Arbeit beziehen, manchmal keinen Sinn (beispielsweise die Leistungsart, da diese ja bei Beschäftigung nicht vorliegt). Umgekehrt liegt der Fall etwas anders: in Leistungsmeldungen wird beispielsweise der erlernte Beruf erfasst, eine Variable, die auch bei Beschäftigungsmeldungen gefüllt ist (mit dem Unterschied, dass sie in dem einen Fall drei-, im anderen vierstellig vorliegt). Aber auch bei Leistungsbezug gibt es Variablen, deren Füllung keinen Sinn ergeben würde (beispielsweise die Stellung im Beruf).

Demgegenüber sollte bei Fehlkodierungen oder der Ausprägung „unbekannt“ eigentlich ein Wert in der entsprechenden Variable vorliegen. Warum manche Arbeitgeber dennoch bestimmte Felder in den Meldungen ihrer Mitarbeiter an die Sozialversicherungen nicht ausfüllen oder Datenübertragungsfehler auftauchen, ist nicht immer im Einzelfall bekannt. Ein prominentes Beispiel ist das Feld „Ausbildung“. Die Anzahl der Personen mit der Ausprägung „Ausbildung unbekannt“ nimmt seit Jahren mit hohen Steigerungsraten zu (vgl. Statistische Jahrbücher). Das kann kein Effekt eines tatsächlich sinkenden Qualifikationsniveaus sein, denn aus anderen Untersuchungen ist bekannt, dass das Qualifikationsniveau der Arbeitnehmer auf dem deutschen Arbeitsmarkt eher gestiegen als gesunken ist.

In Leistungsmeldungen ist neben Datenübertragungs- und Verarbeitungsfehler eine weitere mögliche Fehlerquelle das Vorliegen mehrerer Kundennummern für ein und dieselbe Person, die eine eindeutige Zuordnung der Werte in Ausnahmefällen er-

schwert bzw. nicht mehr ermöglicht. Fehlkodierungen kommen in Beschäftigungsmeldungen weitaus häufiger vor als in Leistungsmeldungen. Die LIAB-Personendaten wurden nicht verändert, ergänzt oder bereinigt. Den Anwendern soll nach Möglichkeit das originale Datenmaterial überlassen werden. Mögliche Bereinigungs- und Ergänzungsschritte werden bei einzelnen Variablenbeschreibungen vorgeschlagen und manchmal ein Umsetzungsvorschlag in der Sprache des Software-Analysetools STATA[®] 8.2. gemacht. Anwender können sich bei Fragen hierzu gerne – auch bei der Übersetzung in andere Programmiersprachen - an das FDZ der BA im IAB wenden.

Zwei Variablen *müssen* aufgrund der Datenschutzbestimmungen verändert werden, die Betriebs- und Personenidentifikatoren. Sie werden in so genannte systemfreie Identifikatoren umgewandelt. Bei den IAB-Betriebspanelbetrieben bezeichnen sie zwischen den Versionen und Datenmodellen immer die gleiche Einheit (den Betrieb). Bei den Personenidentifikatoren bezeichnet eine Nummer nur in der jeweiligen spezifischen Version immer die gleiche Person, zwischen den Versionen und Datenmodellen hingegen kann sie variieren. Innerhalb einer Version sind die Identifikatoren eindeutig und gleich. Das ist beispielsweise wichtig, wenn die verschiedenen Jahresdatensätze des LIAB-Querschnittmodells zu einem Panel zusammengefügt werden. Die Anonymisierung wird im siebten Kapitel ausführlich beschrieben. Vor diesem Hintergrund werden nachfolgend die einzelnen Variablen der Personendatenbanken in der alphabetischen Reihenfolge nach dem Variablennamen aus den LIAB-Testdaten beschrieben.

4.2. Beschreibung der einzelnen Variablen in den LIAB-Personendatenbanken

In der Übersicht 1 werden die einzelnen nachfolgend beschriebenen Variablen für einen Kurzüberblick aufgelistet. Ein grau hinterlegtes Feld kennzeichnet logische fehlende Werte in den Datenquellen BeH und LEH gemäß den bisherigen Ausführungen.

Übersicht 1: Kurzüberblick über die Variablen in den LIAB-Personendatenbanken

Variablenname	Seite	Ifd. Abschnittsnummer	Bezeichnung	LIAB		
				QS*	LS**	
				BeH	BeH	LeH
abg_gr	23	4.2.1.	Abgabegrund der Meldung			
abm_gr	24	4.2.2.	Abmeldegrund aus Leistung			
alter	25	4.2.3.	Alter der Person in Jahren			
arb_aa	26	4.2.4.	Arbeitsamtbezirk des Arbeitsortes			
arb_kr	27	4.2.5.	Kreiskennziffer des Arbeitsortes			
ausbild	28	4.2.6.	Ausbildung in Vbg. mit der beruflichen Qualifikation			
beginn	30	4.2.7.	Beginndatum der Meldung			
berstell	31	4.2.8.	Stellung im Beruf			
beruf	33	4.2.9.	Berufskennziffer			
betr_st	34	4.2.10.	Status der Betriebsnummer			
bew_art	35	4.2.11.	Bewilligungsart bei Leistung			
dat_ber	36	4.2.12.	Kennzeichen Datumsbereinigung			
dauer	37	4.2.13.	Dauer der Meldung in Tagen			
ein_bet	37	4.2.14.	Erstes Eintrittsdatum in den Betrieb			
ein_erw	38	4.2.15.	Erstes Eintrittsdatum in das Erwerbsleben			
ende	38	4.2.16.	Enddatum der Meldung			
euro	39	4.2.17.	Währungskennzeichen			
fam	40	4.2.18.	Familienstand der Person			
geschl	40	4.2.19.	Geschlecht der Person			
id	41	4.2.20.	Personenidentifikationsnummer			

(Fortsetzung nächste Seite)

Variablenname	Seite	Ifd. Abschnittsnummer	Bezeichnung	LIAB		
				QS*	LS**	
				BeH	BeH	LeH
idnum	42	4.2.21.	Identifikationsnummer für Betriebe			
kind	43	4.2.22.	Kinder vorhanden?			
leistart	43	4.2.23.	Leistungsart gruppiert			
level1	44	4.2.24.	Spellzähler der Episode pro Quelle			
nation	45	4.2.25.	Staatsangehörigkeit			
pers_gr	46	4.2.26.	Personengruppenschlüssel			
quelle	47	4.2.27.	Quelle			
tag_entg	47	4.2.28.	Tagesentgelt			
ost_west	49	4.2.29.	Ost-West-Kennzeichen			
wohn_aa	49	4.2.30.	Arbeitsamtbezirk des Wohnortes			
wohn_kr	50	4.2.31.	Kreiskennziffer des Wohnortes			
wz73	50	4.2.32.	Wirtschaftszweig nach BA-Klassifizierung 1973			
wz93	50	4.2.33.	Wirtschaftszweig nach BA-Klassifizierung 1993			

* Querschnittmodell

** Längsschnittmodell

4.2.1. Abgabegrund

Variablenlabel	Abgabegrund der Meldung
Variablenname	abg_gr
Herkunft	BeH, LeH
Bezug	einzelne Meldung
Datentyp	String
Länge	2
Detailbeschreibung	<p>1) BeH-Meldungen</p> <p>In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung erstattet hat. Wann und was Arbeitgeber mit welchen Schlüsselzahlen melden, ist im Anhang beschrieben. Nicht alle theoretisch denkbaren Abmeldegründe tauchen in den Daten tatsächlich auf (vgl. die Häufigkeitsauszählungen in den Codeplänen). Die BeH erfasst die Abgabegründe nur bei Meldungen mit Entgeltangaben (also Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen wegen fehlender Entgeltinformationen nicht enthalten sind. Ein Informationsverlust ist damit jedoch nicht verbunden, da die Angaben aus einer Anmeldung mit der darauf folgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden.</p> <p>Die Abgabegründe für Beschäftigungsmeldungen sind nach den Regeln des seit 1.1.1999 geltenden Meldeverfahrens (gemäß DEÜV¹⁴) verschlüsselt. Im Vergleich zum alten Meldeverfahren (nach den DEVO-/DÜVO-Regeln)¹⁵ werden die Abgabegründe nach dem neuen Recht differenzierter erfasst. Alte Gründe sind auf die neuen umgeschlüsselt.</p> <p>2) LeH-Meldungen</p> <p>Für LeH-Meldungen zeigt das Merkmal „Abgabegrund“ den Grund für das <i>Ende</i> des Bezugs von Arbeitslosengeld, -hilfe oder Unterhaltsgeld an. Die Gründe werden mit Buchstaben erfasst, um sie von Beschäftigungsmeldungen zu unterscheiden. Die einzelnen Schlüsselbuchstaben sind in den LIAB-Codeplänen aufgelistet.</p>

(Fortsetzung nächste Seite)

¹⁴ DEÜV: Datenerfassungs- und -übermittlungsverordnung, in Kraft getreten am 1. Januar 1999.

¹⁵ DEVO: Datenerfassungsverordnung; DÜVO: Datenübermittlungsverordnung; die 2. DEVO und die 2. DÜVO wurden jeweils am 1. Januar 1999 durch die Datenerfassungs- und -übermittlungsverordnung abgelöst.

Besonderheiten [Qualitätshinweise]	<p>Abgabegründe nach dem alten Meldeverfahren werden in das neue umgeschlüsselt. Im LIAB-Querschnittsmodell wird nicht empfohlen, den Abgabegrund „Ende der Beschäftigung“ als Austritt aus dem Betrieb zu werten. Die Trefferquote ist keineswegs 100 Prozent. Manchmal erfolgt direkt im Anschluss an die Abmeldung eine Wiederanmeldung bei dem gleichen Arbeitgeber. Das lässt sich im LIAB-Längsschnittmodell nachweisen, im LIAB-Querschnittsmodell bleiben solche Prozesse unsichtbar. Die Abgabegründe sind nicht immer konsistent mit der Meldedauer. So gibt es beispielsweise Meldungen mit dem Abgabegrund „Jahresmeldung“, die weniger als 365 Tage abdecken. LEH-Meldungen: Für viele Analysezwecke wird bei Leistungsbezugsmeldungen die Verwendung der Variable <code>abm_gr</code> ausreichend sein.</p>
---	---

4.2.2. Abmeldegrund aus dem Leistungsbezug (Zusammenfassung)

Variablenlabel	Abmeldegrund aus dem Leistungsbezug
Variablenname	abm_gr
Herkunft	LeH
Bezug	einzelne Meldung (nur Leistung)
Datentyp	numerisch
Länge	1
Detailbeschreibung	Für LeH-Meldungen fasst die Variable „Abmeldegrund“ die oft spezifischen Einzelgründe der Variable abg_gr zusammen. In den meisten Fällen ist diese Variable für Analysezwecke ausreichend.
Besonderheiten [Qualitätshinweise]	wie bei der Variable abg_gr bei Leistungsmeldungen. Die Variable gibt es nur in Versionen des LIAB-Längsschnittmodells. In Beschäftigungsmeldungen ist sie dort nicht gefüllt.

Wert	Label
1	Arbeitsaufnahme
2	Anspruch auf andere Leistung
3	sonstige Gründe
4	steht dem Arbeitsmarkt nicht zur Verfügung
5	Ablauf der Maßnahme
6	Sperr-, Säumniszeiten
7	Anspruch erschöpft
8	Bewilligungsabschnitt abgelaufen
9	Fehlkodierung/fehlt

4.2.3. Alter der Person

Variablenlabel [Bezeichnung]	Alter der Person in Jahren
Variablenname [Attributname]	alter
Herkunft [Verwendung]	BeH, LeH
Bezug	einzelne Meldungen, personenbezogen
Datentyp	numerisch
Länge	3
Detailbeschreibung	<p>1) BeH-Meldungen</p> <p>In BeH-Sätzen wird das Alter jeweils aus dem Geburtsjahr (Quelle: Sozialversicherungsnummer) und dem Meldejahr berechnet. Die Variable gibt demnach das Alter des Beschäftigten zum 31.12. des betreffenden Jahres an. Da das Jahr in der Versicherungsnummer 2-stellig erfasst ist, kann ein negatives Alter entstehen; so ergibt z.B. die Differenz aus dem Meldejahr 1996 und dem Geburtsjahr 97 aus der Versicherungsnummer die Differenz "-1". In Fällen wie diesen, in denen der Beschäftigungsbeginn vor der Geburt liegt, wird für das Jahrhundert jeweils "18" angenommen, so dass für den Beispielfall das Alter 99 ausgewiesen wird. In Einzelfällen kommen Altersangaben über 100 vor; dann wird 100 subtrahiert, weshalb es (wenige) nulljährige Beschäftigte gibt.</p> <p>2) LeH-Meldungen</p> <p>In LeH-Sätzen gibt die Variable „Alter“ jeweils das Alter zum Beginn der betreffenden Leistungsbezugsmeldung an. Ansonsten wie BeH.</p>
Besonderheiten [Qualitätshinweise]	Es gibt in den LIAB-Daten nur selten die beschriebenen weniger hilfreichen Altersangaben. Meldefehler können nicht immer vollständig ausgeschlossen werden. In den meisten Analysen ist eine Beschränkung auf Altersangaben im Intervall [15;65] möglich und sinnvoll. Damit werden – je nach LIAB-Datenmodell und –Version – etwa 0,2 Prozent aller Meldungen bzw. Personen ausgeschlossen.

4.2.4. Arbeitsamtbezirk des Arbeitsortes

Variablenlabel [Bezeichnung]	Arbeitsamtbezirk des Arbeitsortes
Variablenname [Attributname]	arb_aa
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	Die Information zum Arbeitsamtsbezirk des Arbeitsortes gibt es nur in der BeH. Angegeben ist jeweils das Arbeitsamt, in dessen Bezirk der Betrieb des Beschäftigten angesiedelt ist. Da sich die Grenzen der Arbeitsamtsbezirke im Zeitverlauf ändern, würden ohne Gebietsstandsberreinigung Fälle auftreten, bei denen der Arbeitsamtsbezirk eines Betriebes wechselt, ohne dass dieser Betrieb seinen Standort verlagert hat. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, sind die Arbeitsamtsangaben daher derzeit auf den Gebietsstand vom 31.12.2001 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Arbeitsamtsbezirk nach den Grenzen, welche die Arbeitsamtsbezirke zum 31.12.2001 hatten. Ein Problemfall stellt hierbei Berlin dar: Die Berliner Arbeitsämter wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt wird. Nähere Informationen zur Gebietsstandsberreinigung befinden sich in der Dokumentation der IAB-Regionalstichprobe. Die dortigen Informationen können auf die LIAB-Personendaten übertragen werden.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Arbeitsamtsbezirksnummern) befinden sich in den LIAB-Codeplänen. Die Variable ist ab dem Jahr 1999 gefüllt.

4.2.5. Kreiskennziffer des Arbeitsortes

Variablenlabel [Bezeichnung]	Kreiskennziffer des Arbeitsortes
Variablenname [Attributname]	arb_kr
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	Die Kreiskennzifferinformation stammt aus der BeH. Angegeben ist jeweils der Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. Angegeben ist jeweils der Kreis, in dem Beschäftigungsbetrieb angesiedelt ist.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Kreiskennziffern) befinden sich in den LIAB-Codeplänen. Die Variable ist nur bei Beschäftigungsmeldungen ab dem Jahr 1999 gefüllt.

4.2.6. Ausbildung (B2-Schlüssel)

Variablenlabel [Bezeichnung]	Ausbildung in Verbindung mit der beruflichen Qualifikation
Variablenname [Attributname]	Ausbild
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	1

(Fortsetzung nächste Seite)

<p>Detailbeschreibung</p>	<p>1) BeH-Meldungen</p> <p>Die Schul- und Berufsausbildung des Beschäftigten weist der Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ aus. Dieser fünfstellige Tätigkeitsschlüssel umfasst die Schlüssel:</p> <ul style="list-style-type: none"> • „Ausgeübte Tätigkeit“ (erste bis dritte Stelle, 3-Steller gemäß der „Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen“, hrsg. von der Bundesanstalt für Arbeit, Nürnberg 1988) • „Stellung im Beruf“ (vierte Stelle, so genannter B1-Schlüssel) • „Ausbildung“ (fünfte Stelle, so genannter B2-Schlüssel) <p>„Bei dem Signierschlüssel „Ausbildung“ handelt es sich um einen kombinierten Schlüssel, durch den sowohl die erreichte Schulbildung als auch die abgeschlossene Berufsausbildung (jeweils höchster Abschluss) der/des Beschäftigten erhoben werden soll. Dabei ist grundsätzlich zunächst die Schulbildung festzustellen, danach die Berufsausbildung. Lediglich bei den Schlüsselpositionen 5 (Abschluss einer Fachhochschule) und 6 (Hochschul-/Universitätsabschluss) wird auf die Feststellung einer eventuell sonstigen Berufsausbildung verzichtet“ (Bundesagentur für Arbeit, 2004).</p> <p>Nicht als Ausbildung zählen berufliche Fortbildung (z.B. Kurse an Sprach- und Dolmetscherschulen) sowie Fachkenntnisse, die durch praktische Tätigkeiten erworben wurden.</p> <p>„Für bestimmte Teilgruppen gibt es einen hohen Anteil fehlender Angaben, weil das Merkmal versicherungsrechtlich keine besondere Bedeutung hat [das betrifft z.B. die geringfügig entlohnten Beschäftigten; H.A.] (...). Zusammen mit einem Betriebswechsel treten manchmal Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammengestellt werden. Wenn ein Beschäftigter z.B. durch berufsbegleitende Weiterbildung einen höheren Abschluss erlangt hat, dann wird dieser Statuswechsel manchmal erst von einem neuen Beschäftigungsbetrieb erfasst“ (Meinken/Koch, 2004).</p>
<p>Besonderheiten [Qualitätshinweise]</p>	<p>Der mangelnde Füllungsgrad der Variable ist insbesondere im LI-AB-Querschnittmodell manchmal ein Auswertungshindernis. Auffällig ist, dass die Ausprägung „Ausbildung unbekannt/fehlt“ zum aktuellen Rand hin deutlich zunimmt. Einen möglichen Umgang mit dem Problem inkonsistenter Füllungen der Bildungsvariable beschreiben Osikominu/Fitzenberger/Völter (2004).</p>

Wert	Label
0	ohne Ausbildung (ohne Schul- und beruflichen Abschluss)
1	ohne abgeschlossene Berufsausbildung bis mittlere Reife
2	mit abgeschlossener Berufsausbildung (abgeschlossene Lehr- oder Anlernausbildung, Abschluss einer Berufsfach-/Fachschule ¹⁶) bis mittlere Reife
3	ohne abgeschlossene Berufsausbildung mit (Fach-)Abitur
4	mit abgeschlossener Berufsausbildung (abgeschlossene Lehr- oder Anlernausbildung, Abschluss einer Berufsfach-/Fachschule) mit (Fach-)Abitur
5	Abschluss einer Fachhochschule (frühere Bezeichnung: höhere Fachschule) z.B. Ingenieurschulen, höhere Fachschulen für Sozialarbeit, höhere Wirtschaftsfachschulen, höhere Fachschulen für Jugendleiterinnen, Jugendleiterseminare, höhere Fachschulen für Sozialpädagogik, höhere Handelsschulen, soweit mit Fachhochschulreife abgeschlossen.
6	Hochschul-/Universitätsabschluss
7	Ausbildung unbekannt, Angabe nicht möglich
9	Keine Angabe

¹⁶ Zu den Berufsfach- oder Fachschulen zählen z.B. Technikerschulen, Krankenpflegeschulen, Frauenfachschulen, Handelsschulen, Wirtschaftsfachschulen, Fachschulen für Betriebswirtschaft, Verwaltungs- und Wirtschaftsakademien, Meisterschulen, höhere Handelsschulen, soweit mit „mittlerer Reife“ abgeschlossen.

4.2.7. Beginndatum

Variablenlabel [Bezeichnung]	Beginndatum der Meldung
Variablenname [Attributname]	beginn
Herkunft [Verwendung]	BeH, LeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	Die Variable gibt im SAS/STATA-Format die Tage an, die seit dem 01. Januar 1960 bis zum Beginndatum der Meldung vergangen sind. Zu beachten ist das Episodensplitting, das nach dem Zusammenführen der Meldungen aus BeH und LeH durchgeführt wird: Überschneiden sich Meldezeiträume im Konto einer Person, z.B. weil der betreffende Versicherte eine zweite Beschäftigung zusätzlich zu seinem bisherigen Beschäftigungsverhältnis aufgenommen hat, werden die sich zeitlich überschneidenden Spells so zerlegt, dass nur noch einzelne Meldungen und sich zeitlich exakt überdeckende Mehrfachspells übrig bleiben. Dazu wird die zeitlich früher beginnende Meldung jeweils dort zerschnitten, wo der Meldezeitraum der nachfolgenden Meldung begann; wenn die Meldungen nicht zeitgleich endeten, wird die länger laufende Meldung wiederum an derjenigen Stelle zerlegt, an der die früher endende Meldung aufhörte. Eine ausführliche Erläuterung und graphische Veranschaulichung des Episodensplittings befindet sich im Kapitel sechs.
Besonderheiten [Qualitätshinweise]	Im LIAB-Querschnittsmodell ist diese Variable oft weniger bedeutsam, weil sich aus dem Beginn- und Enddatum der Meldung nicht unbedingt ableiten lässt, dass Beschäftigte entweder kontinuierlich über das Jahr beschäftigt sind oder aus dem Betrieb ausscheiden. Im LIAB-Längsschnittmodell ist zu beachten, dass zeitgleiche Beschäftigungs- und Leistungsmeldungen in einem Personenkonto vorkommen können. Oft wird so verfahren, dass in diesem Fall Beschäftigung vor Leistung geht und die (exakt) zeitgleiche Meldung aus dem Datensatz entfernt wird. Das entscheiden Anwender selbstständig (vgl. auch 4.3.1.2.).

Nachfolgend ist jeweils der **01. Januar** im Zeitraum 1990 bis 2004 im SAS/STATA-Format als numerische Zahl dokumentiert.

Jahr	Datumswert	Jahr	Datumswert	Jahr	Datumswert
1990	10958	1995	12784	2000	14610
1991	11323	1996	13149	2001	14976
1992	11688	1997	13515	2002	15341
1993	12054	1998	13880	2003	15706
1994	12419	1999	14245	2004	16071

4.2.8. Stellung im Beruf (B 1 – Schlüssel)

Variablenlabel [Bezeichnung]	Stellung im Beruf
Variablenname [Attributname]	berstell
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	<p>1) BeH-Meldungen</p> <p>Die aktuelle berufliche Stellung des Beschäftigten wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt. Das Merkmal „Berufsstellung“ unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten; maßgeblich ist das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit. Nur für Vollzeitbeschäftigte liefert die Variable tatsächlich Angaben zur beruflichen Stellung, während für Teilzeitbeschäftigte nur qualitativ erfasst wird, ob sie mehr als die Hälfte der üblichen regelmäßigen Arbeitsstunden eines Vollzeitbeschäftigten geleistet haben oder nicht. Diese Grenze lag bis 1978 bei 20 Wochenstunden, zwischen 1979 und 1987 bei 15 Wochenstunden und seit 1988 bei 18 Wochenstunden.</p> <p>Die Unterscheidung zwischen vollzeitbeschäftigten Arbeitern (1,2) und Angestellten richtet sich ausschließlich nach der Art der Rentenversicherungspflicht: „Unterliegen Arbeitnehmer/innen der Versicherungspflicht in der bisherigen Arbeiterrentenversicherung, so sind sie nach der »Stellung im Beruf« als Arbeiter/innen (Schlüsselzahl 2) auszuweisen. Gehören die Arbeitnehmer/innen zum Personenkreis der bisherigen angestelltenversicherungspflichtigen Beschäftigten, so sind sie nach der »Stellung im Beruf« als Angestellte (Schlüsselzahl 4) zu kennzeichnen. Bei Beschäftigten, die von der Rentenversicherungspflicht befreit sind, ist zu entscheiden, zu welchem bisherigen Rentenversicherungszweig Versicherungspflicht bestünde, wenn ein Befreiungstatbestand nicht vorliegen würde“ (Bundesagentur für Arbeit, 2004).</p> <p>Falls mehrerer Schlüsselzahlen in Betracht kommen, ist der Arbeitgeber dazu angehalten, die Einstufung nach der überwiegend ausgeübten Tätigkeit vorzunehmen. Lässt sich diese nicht eindeutig feststellen, ist die Schlüsselzahl der höherwertigen Stellung im Beruf einzutragen (ebd.).</p>
Besonderheiten [Qualitätshinweise]	Die Ausprägungen 5 und 6 kennzeichnen fehlende/unbekannte Werte. In Abschnitt 4.3.1. und 4.3.2. wird außerdem auf die Struktur der unplausiblen Angaben im Zusammenhang mit anderen Meldeinformationen eingegangen.

Werte	Labels
0	in Ausbildung neben den Auszubildenden, Praktikanten und Volontären zählen auch Anlernlinge, Schüler an Schulen des Gesundheitswesens und Teilnehmer an geförderten Maßnahmen zur beruflichen Fortbildung, Umschulung und betrieblichen Einarbeitung zu der Kategorie Beschäftigte in beruflicher Ausbildung.
1	nicht formal qualifiziert
2	Facharbeiter
3	Meister, Poliere (gleichgültig ob Arbeiter oder Angestellte)
4	Angestellter (aber nicht Meister im Angestelltenverhältnis)
5	fehlt
6	unbekannt
7	Heimarbeiter, Hausgewerbetreibende
8	Teilzeit (bis zur Hälfte der Vollarbeitszeit, seit 1988: bis 18 Stunden)
9	Teilzeit (mehr als die Hälfte der Vollarbeitszeit, seit 1988: über 18 Stunden)

4.2.9. Berufskennziffer

Variablenlabel [Bezeichnung]	Berufsordnung
Variablenname [Attributname]	beruf
Herkunft [Verwendung]	BeH (ausgeübte Tätigkeit), LeH (erlernter Beruf)
Bezug	Meldung
Datentyp	numerisch
Länge	4

(Fortsetzung nächste Seite)

<p>Detailbeschreibung</p>	<p>1) BeH-Meldungen</p> <p>Die Berufsbezeichnung der aktuell vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Der Arbeitgeber verschlüsselt die Tätigkeit des Beschäftigten gemäß der „Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbezeichnungen“, in dem ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Berufe.</p> <p>Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die <i>überwiegend</i> ausgeübte Tätigkeit kennzeichnet.</p> <p>„Bestimmte Gruppen von Erwerbstätigen, wie z.B. Beamte, Selbständige, freiberuflich Tätige oder mithelfende Familienangehörige, sind nicht sozialversicherungspflichtig. Dadurch kann es bei Auswertungen zu einzelnen Berufsgruppen zu Ergebnissen kommen, die die Verhältnisse nicht so abbilden, wie sie sich unter Berücksichtigung aller Erwerbstätigen in der jeweiligen Berufsgruppe zweigen würden“ (Meinken/Koch, 2004).</p> <p>2) LeH-Meldungen</p> <p>In LeH-Meldungen sollte das Merkmal „Berufsordnung“ eigentlich den Ausbildungsberuf der Leistungsempfänger ausweisen; jedoch treten zahlreiche Werte auf, die keinem gültigen Berufscodex entsprechen. Wegen der offensichtlich schlechten Datenqualität ist die Auswertung des Merkmals in LeH-Meldungen nicht empfehlenswert.</p>
<p>Besonderheiten [Qualitätshinweise]</p>	<p>In BeH-Meldungen befindet sich hinter der dreistelligen Berufsordnung eine Null. Das Merkmal ist also sowohl bei Beschäftigungs- wie Leistungsmeldungen stets vierstellig. Es gibt einige wenige Berufskennziffern, die keine Entsprechung in der BA-Klassifizierung der Berufe haben. Die Fallzahlen dieser Fehlkodierungen sind sehr gering.</p> <p>Die einzelnen Ausprägungen sind in den LIAB-Codeplänen aufgelistet. Im Anhang A2 befindet sich ein Klassifizierungsvorschlag für die Berufe auf der Dreistellerebene in zehn Berufsgruppen von Blossfeld/Hamerle/Mayer (1985).</p>

4.2.10. Status der Betriebsnummer

Variablenlabel [Bezeichnung]	Status der Betriebsnummer
Variablenname [Attributname]	betr_st
Herkunft [Verwendung]	BeH, IAB-Betriebspanel
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	In der Variable wird ausgewiesen, ob es sich bei der abgegebenen Beschäftigtenmeldung um die aus einem Betrieb handelt, der für die spezifische Version des LIAB-Längsschnittmodells ausgewählt wurde, ob es sich um einen anderen IAB-Betriebspanelbetrieb handelt oder ob es ein Betrieb ohne eine Entsprechung im IAB-Betriebspanel ist.
Besonderheiten [Qualitätshinweise]	Die Variable gibt es nur in Versionen des LIAB-Längsschnittmodells, vgl. auch die Ausführungen zum systemfreien Betriebsidentifikator in Abschnitt 7.2..

Wert	Label
1	Betrieb der Version X Ost
2	Betrieb der Version X West
3	sonstiger IAB-Betriebspanelbetrieb
4	Betrieb ohne Entsprechung im Panel

4.2.11. Bewilligungsart

Variablenlabel [Bezeichnung]	Bewilligungsart bei Leistung
Variablenname [Attributname]	bew_art
Herkunft [Verwendung]	LeH
Bezug	Meldung
Datentyp	numerisch
Länge	2
Detailbeschreibung	In der Variable werden die Bewilligungsgründe für den Leistungsbezug zusammengefasst.
Besonderheiten [Qualitätshinweise]	Es gibt eine relativ hohe Anzahl der Ausprägung „sonstige Sätze“, aber die Variable ist zuverlässig gefüllt.

Wert	Label
1	Erst-/Neubewilligung ALG
2	Weiterbewilligung ALG ohne Unterbrechung
3	Weiterbewilligung ALG mit Unterbrechung
4	Neubewilligung ALG mit Aufstockung
5	Weiterbewilligung ALG mit Kuerzung
6	ALH direkt im Anschluss an ALG
7	ALH direkt im Anschluss an ALH
8	ALH mit Luecke zum vorherigen Satz
9	sonstige Saetze
10	Kappung auf maximalen Altersanspruch

Abkürzungen: ALG: Arbeitslosengeld

ALH: Arbeitslosenhilfe

4.2.12. Datumsbereinigung

Variablenlabel [Bezeichnung]	Kennzeichen Datumsbereinigung bei Überlappung BeH/LeH
Variablenname [Attributname]	dat_ber
Herkunft [Verwendung]	BeH und LeH
Bezug	Meldung
Datentyp	numerisch
Länge	1

(Fortsetzung nächste Seite)

<p>Detailbeschreibung</p>	<p>Das Merkmal gibt an, ob das Episodendatum einer Meldung bei der Bereinigung der Überschneidungen von BeH- und LeH-Meldungen korrigiert wurde.</p> <p>Bereinigung der Überlappungen von BeH-/LeH-Meldungen</p> <p>Zeitliche Überschneidungen zwischen sozialversicherungspflichtiger Beschäftigung und dem Bezug von Arbeitslosengeld, -hilfe oder Unterhaltsgeld werden bereinigt, da solche Überschneidungen rechtlich nicht zulässig sind. Hingegen schließen sich geringfügige Beschäftigung und Leistungsbezug nicht aus; entsprechend blieben diese Überschneidungen unangetastet.</p> <p>Bei der Korrektur von Überlappungen zwischen sozialversicherungspflichtiger Beschäftigung und Leistungsbezug wurde folgendermaßen vorgegangen:</p> <ul style="list-style-type: none"> • War das Entgelt im BeH-Satz 0 oder missing, wurde der BeH-Satz gelöscht und der LeH-Satz blieb erhalten. • Bei Entgelt größer 0 wurden Anfangs- und Enddatum der BeH- und LeH-Meldung miteinander verglichen: Der BeH-Satz wurde nur dann gelöscht, wenn er komplett in dem LeH-Satz enthalten war. In allen anderen Fällen wurde der LeH-Satz gelöscht und der BeH-Satz beibehalten. <p>Wurde ein Satz gelöscht, wurden alle weiteren Spells derselben Episode und derselben Quelle, also auch geringfügige Beschäftigungsmeldungen, ebenfalls gelöscht.</p> <p>Durchführen der Datumskorrekturen</p> <p>Wurden bei der Korrektur der Überlappungen Meldungen gelöscht, die Auslöser für ein Episodensplitting waren (zum Episodensplitting siehe Kapitel sechs), musste dieses Splitting nun wieder rückgängig gemacht werden. Dazu wurde der Episodenzeitraum eines der gesplitteten Spells entsprechend korrigiert und die übrigen gesplitteten Spells gelöscht.</p>
<p>Besonderheiten [Qualitätshinweise]</p>	<p>Sofern Anwender das Episodensplitting akzeptieren, ist das Merkmal von geringer Bedeutung. Im LIAB-Querschnittsmodell ist es praktisch wertlos, denn es kann nicht überprüft werden, wie die Datumsbereinigung in einem konkreten Fall stattgefunden hat. Im LIAB-Längsschnittmodell lässt sich das oft rekonstruieren und im Bedarfsfall korrigieren bzw. ändern. In der Mehrheit der Anwendungen wird das nicht notwendig sein.</p>

4.2.13. Dauer

Variablenlabel [Bezeichnung]	Dauer der Meldung in Tagen
Variablenname [Attributname]	dauer
Herkunft [Verwendung]	BeH, LeH
Bezug	Meldung
Datentyp	numerisch
Länge	4
Detailbeschreibung	Die Variable wird generiert durch die Differenz des Ende- minus dem Beginndatum der Meldung und gibt dementsprechend die Dauer der Meldung in Tagen an.
Besonderheiten [Qualitätshinweise]	Beschäftigungsmeldungen haben eine maximale Dauer von 366 Tagen, Leistungsmeldungen können unter Umständen auch länger andauern. Nicht alle Beschäftigungsmeldungen mit dem Abgabegrund „Jahresmeldungen“ dauern auch tatsächlich 365 (366) Tage an (vgl. 4.3.3.), was teilweise eine Konsequenz des Episodensplittings ist.

4.2.14. Eintrittsdatum Betrieb

Variablenlabel [Bezeichnung]	Erstes Eintrittsdatum in den Betrieb
Variablenname [Attributname]	ein_bet
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	In der Variable ist das erste Beginndatum einer Personenmeldung unter der meldenden Betriebsnummer abgelegt. Das kann auch eine Auszubildendenmeldung sein. Der Wert ist ebenso wie das Beginn- und Enddatum im SAS/STATA-Datumsformat verfügbar, wird also intern im Softwareanalysetool als Anzahl der Tage seit dem 01. Januar 1960 abgelegt.
Besonderheiten [Qualitätshinweise]	Die Variable existiert nur im LIAB-Querschnittsmodell derzeit bis zum Jahr 2001. Mit der Variable kann die Betriebszugehörigkeitsdauer zur aktuellen Meldung approximiert werden, sofern davon ausgegangen wird, dass die Person seit dem ersten Eintrittsdatum bis zur aktuellen Meldung den Betrieb nicht verlässt.

4.2.15. Eintrittsdatum in das Erwerbsleben

Variablenlabel [Bezeichnung]	Erstes Eintrittsdatum der Person in das Erwerbsleben
Variablenname [Attributname]	ein_erw
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	In der Variable ist das erste Beginndatum einer sozialversicherungspflichtigen Beschäftigung der Person (ohne Ausbildung ¹⁷) dokumentiert. Der Wert ist ebenso wie das Beginn- und Enddatum im SAS/STATA-Datumformat verfügbar, wird also intern im Softwareanalysetool als Anzahl der Tage seit dem 01. Januar 1960 abgelegt.
Besonderheiten [Qualitätshinweise]	Die Variable existiert nur im LIAB-Querschnittsmodell derzeit bis zum Jahr 2001. Mit der Variable kann die Berufserfahrung approximiert werden, sofern davon ausgegangen wird, dass die Person seit dem ersten Eintrittsdatum bis zur aktuellen Meldung kontinuierlich beschäftigt ist.

¹⁷ Daher kann der Eintritt in das Erwerbsleben in den meldenden Betrieb früher sein als der Eintritt in das Erwerbsleben.

4.2.16. Enddatum

Variablenlabel [Bezeichnung]	Enddatum der Meldung
Variablenname [Attributname]	ende
Herkunft [Verwendung]	BeH, LeH
Bezug	Meldung
Datentyp	numerisch
Länge	5
Detailbeschreibung	Die Variable gibt im SAS/STATA-Format die Tage an, die seit dem 01. Januar 1960 vergangen sind und das Enddatum der Meldung kennzeichnen. Zu beachten ist das Episodensplitting, das nach dem Zusammenführen der Meldungen aus BeH und LeH durchgeführt wird: Überschneiden sich Meldezeiträume im Konto einer Person, z.B. weil der betreffende Versicherte eine zweite Beschäftigung zusätzlich zu seinem bisherigen Beschäftigungsverhältnis aufgenommen hat, werden die sich zeitlich überschneidenden Spells so zerlegt, dass nur noch einzelne Meldungen und sich zeitlich exakt überdeckende Mehrfachspells übrig bleiben. Dazu wird die zeitlich früher beginnende Meldung jeweils dort zerschnitten, wo der Meldezeitraum der nachfolgenden Meldung begann; wenn die Meldungen nicht zeitgleich endeten, wird die länger laufende Meldung wiederum an derjenigen Stelle zerlegt, an der die früher endende Meldung aufhörte. Eine ausführliche Erläuterung und graphische Veranschaulichung des Episodensplittings befindet sich im sechsten Kapitel.
Besonderheiten [Qualitätshinweise]	Im LIAB-Querschnittsmodell ist diese Variable oft weniger bedeutsam, weil sich aus dem Beginn- und Enddatum der Meldung nicht unbedingt ableiten lässt, dass Beschäftigte entweder kontinuierlich über das Jahr beschäftigt sind oder aus dem Betrieb ausscheiden. Im LIAB-Längsschnittmodell ist zu beachten, dass zeitgleiche Beschäftigungs- und Leistungsmeldungen in einem Personenkonto vorkommen können. Oft wird so verfahren, dass in diesem Fall Beschäftigung vor Leistung geht und die (exakt) zeitgleiche Meldung aus dem Datensatz entfernt wird. Das entscheiden Anwender selbstständig (vgl. 4.3.1.2.).

4.2.17. Währungskennzeichen

Variablenlabel [Bezeichnung]	Währungskennzeichen
Variablenname [Attributname]	Euro
Herkunft [Verwendung]	BeH, LeH
Bezug	Meldung
Datentyp	Numerisch
Länge	1
Detailbeschreibung	<p>Das Währungskennzeichen gibt an, ob die einzelnen Entgelte jeweils auf D-Mark oder Euro bezogen sind. In den Jahren vor 1993 kommen nur DM-Beträge vor. Euro-Beträge kommen bereits ab 1993 vor, da Nachmeldungen bis einschließlich 1993 berücksichtigt werden (ITM, 2004).</p> <p>Zu beachten ist, dass es bei LeH-Sätzen keine Jahresgrenze für die Unterscheidung von DM und Euro gibt. Es gibt in allen Jahren beide Währungskennzeichen. Eine generelle Umrechnung für LeH in Eurobeträge ist ebenfalls nicht erfolgt, da dann ein Abgleich mit Leistungstabellen für die LeH-Datensätze nicht mehr möglich wäre (ITM, 2004).</p>
Besonderheiten [Qualitätshinweise]	In Auswertungen werden oft alle DM- in Euro-Beträge (oder umgekehrt) umgerechnet (1 Euro = 1.95883 DM). Meistens werden die Lohnangaben zusätzlich deflationiert.

Ausprägungen: 0 nein (gleich DM) 1 ja (gleich Euro)

4.2.18. Familienstand der Person

Variablenlabel [Bezeichnung]	verheiratet
Variablenname [Attributname]	fam
Herkunft [Verwendung]	LeH
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	Nur in LeH-Meldungen wird angegeben, ob die Person verheiratet ist oder nicht.
Besonderheiten [Qualitätshinweise]	Das Merkmal existiert nur in Versionen des LIAB-Längsschnittmodells und ist dort bei Beschäftigungsmeldungen nicht gefüllt. Qualitätsprüfungen können nicht durchgeführt werden, weil es keine logischen Zusammenhänge zu anderen Variablen im Datensatz gibt.

Ausprägungen: 0 nein (nicht verheiratet) 1 ja (verheiratet)

4.2.19. Geschlecht

Variablenlabel [Bezeichnung]	Geschlecht der Person
Variablenname [Attributname]	geschl
Herkunft [Verwendung]	Sozialversicherungsnummer; BeH, LeH
Bezug	Meldung, Person
Datentyp	numerisch
Länge	1
Detailbeschreibung	Die Angabe „Geschlecht“ wird aus der Sozialversicherungsnummer generiert.
Besonderheiten [Qualitätshinweise]	In einigen wenigen Fällen ist das Merkmal nicht für alle Meldungen des Personenkontos konstant. Dann kann es durch den gerundeten Mittelwert aller Meldungen ersetzt werden ¹ , was allerdings nur im LIAB-Längsschnittmodell möglich ist.

Ausprägungen: 1 Mann 2 Frau

4.1.20. Personenidentifikator

Variablenlabel [Bezeichnung]	Personenidentifikationsnummer
Variablenname [Attributname]	id
Herkunft [Verwendung]	Sozialversicherungsnummer
Bezug	Person
Datentyp	numerisch
Länge	7
Detailbeschreibung	Der systemfreie Personenidentifikator wird aus der originalen Sozialversicherungsnummer generiert. Die Nummern sind fortlaufend und identifizieren eindeutig Personen.
Besonderheiten [Qualitätshinweise]	Die ursprüngliche Sozialversicherungsnummer wird in den LIAB-Daten in <i>versionsspezifische</i> eindeutige systemfreie Personenidentifikatoren umgeschlüsselt. Versionsspezifisch bedeutet, dass Personen aus verschiedenen LIAB-Versionen nicht eindeutig zugeordnet sind. Die Identifikationsnummer 1 beispielsweise bezeichnet in der Version 1 des LIAB-Quer- und Längsschnittsmodells <i>verschiedene</i> Personen. In den einzelnen Jahresscheiben hingegen bezeichnet der Identifikator auf der Personenebene immer die gleiche Person (vgl. Abschnitt 7.2.1.).

4.1.21. Betriebsidentifikator

Variablenlabel [Bezeichnung]	Identifikationsnummer für Betriebe
Variablenname [Attributname]	idnum
Herkunft [Verwendung]	IAB-Betriebspanel, Betriebsnummer aus BLH
Bezug	Meldung
Datentyp	numerisch
Länge	7
Detailbeschreibung	<p>Der systemfreie Identifikator für Betriebe wird - sofern vorhanden - aus dem IAB-Betriebspanel verwendet. Im LIAB-Querschnittmodell ist das ein erschöpfendes Kriterium, weil es dort nur Meldungen in den Personendatenbanken aus IAB-Betriebspanelbetrieben gibt. Im Längsschnittmodell gibt es auch Meldungen aus Betrieben, die entweder gar nicht im IAB-Betriebspanel auftauchen oder aus Betriebspanelbetrieben stammen, die nicht für die spezifische Längsschnittversion ausgewählt sind. Letztere Art von Betrieben bekommt die Identifikationsnummern aus dem IAB-Betriebspanel zugewiesen. Erstere werden fortlaufend nummeriert. Der Ausgangspunkt ist die letzte vergebene Identifikationsnummer der IAB-Betriebspanelbetriebe plus 100.000. Ab dann werden die nicht befragten Betriebe fortlaufend nummeriert. In allen Versionen des LIAB-Quer- und Längsschnittmodells bezeichnen also die Identifikationsnummern der IAB-Betriebspanelbetriebe immer exakt den gleichen Betrieb, um die Verknüpfung der Betriebs- und Personendaten jederzeit zu gewährleisten. Nur bei nicht befragten Betrieben ändern sich die Identifikationsnummern zwischen den einzelnen (Längsschnitt-)Versionen und bezeichnen innerhalb einer Version immer den gleichen Betrieb.</p>
Besonderheiten [Qualitätshinweise]	<p>Aufgrund der o.g. Besonderheiten ist in alle Versionen des LIAB-Längsschnittmodells die Variable <code>betr_st</code> integriert, die angibt, welchen Status die Betriebsnummern im Rahmen des LIAB hat (vgl. 4.2.10). In Abschnitt 7.2. wird das Anonymisierungsverfahren für die Betriebsnummer ausführlich beschrieben.</p>

4.2.22. Kinder

Variablenlabel [Bezeichnung]	Kinder vorhanden (nur bei Leistung)
Variablenname [Attributname]	kind
Herkunft [Verwendung]	LeH
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	Die Variable enthält eine qualitative Information über Kinder der Person bei Leistungsbezug.
Besonderheiten [Qualitätshinweise]	Die Bedeutung des Wertes zwei in der Variable ist nicht genau bekannt. Es gibt nur sehr wenige Fälle mit dieser Ausprägung. Ähnlich wie beim Merkmal Familienstand sind Qualitätskontrollen kaum möglich, weil es keine logischen Zusammenhänge zu anderen Variablen im Datensatz gibt.

Ausprägungen: 0 nein (keine Kinder vorhanden) | 1 ja (Kinder vorhanden) | 2 (?)

4.2.23. Leistungsart

Variablenlabel [Bezeichnung]	Leistungsart gruppiert
Variablenname [Attributname]	leistart
Herkunft [Verwendung]	LeH
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	Die Variable fasst die vielfältigen Leistungsarten der Bundesagentur für Arbeit in vier Hauptkategorien zusammen.
Besonderheiten [Qualitätshinweise]	Die Variable existiert nur im LIAB-Längsschnittmodell. Die Ausprägung 5 sind beispielsweise Zuschüsse zur Rentenversicherung, d.h. dort gibt es relativ häufig Überschneidungen mit Beschäftigungsmeldungen.

Wert	Label
1	Arbeitslosengeld
2	Arbeitslosenhilfe
3	Unterhaltsgeld
5	Zuschuss nach §166b AFG od. §207a SGB III

4.2.24. Episodenbezogener Zähler

Variablenlabel [Bezeichnung]	Spellzähler der Episode pro Quelle
Variablenname [Attributname]	level1
Herkunft [Verwendung]	BeH, LeH
Bezug	Episode
Datentyp	numerisch
Länge	1
Detailbeschreibung	<p>Es handelt sich um einen Zähler, der parallele Zeiträume <i>innerhalb einer Quelle</i> aufaddiert.</p> <p>Beim Episodensplitting wird eine Level-Variable erzeugt, die anzeigt, ob innerhalb eines Versichertenkontos zeitliche Überschneidungen vorliegen. Die Variable level1 hingegen zeigt an, ob innerhalb einer Quelle, also BeH oder LeH, eine Überschneidung vorliegt. Die Level-Variable hat den Wert 0 für jeweils den ersten Datensatz, der für eine Episode pro Quelle vorliegt. Jeder weitere Datensatz für dieselbe Episode und gleiche Quelle (BeH oder LeH) wird dann aufaddiert. Um also z.B. alle Beschäftigungszeiträume (bzw. Leistungsempfangszeiten) jedoch ohne Mehrfachbeschäftigung zu erhalten, müsste der Filter level1 = 0 (und quelle = 1 für Beschäftigtendatensätze) gesetzt werden.</p> <p>Welcher Datensatz bei mehreren Datensätzen, die denselben Zeitraum abdecken, den Wert 0 erhält, hängt von der vor dem Episodensplitting durchgeführten Sortierung ab. Diese wird wie folgt vorgenommen:</p> <p>level1: (gleichzeitige Spells innerhalb einer Quelle)</p> <p>Beschäftigtendaten:</p> <ul style="list-style-type: none"> ➤ Versicherungsnummer (aufsteigend) ➤ Beginndatum (aufsteigend) ➤ Geringfügigkeitskennzeichen (aufsteigend) ➤ Tagesentgelt (absteigend) <p>Das Geringfügigkeitskennzeichen ist folgendermaßen definiert: Personengruppen 109, 110, 202, 209 und 210 (= geringfügig und kurzfristig Beschäftigte) erhalten den Wert 2, alle anderen Personengruppen erhalten den Wert 1.</p> <p>Diese Sortierung führt dazu, dass bei Mehrfachbeschäftigung immer die sozialversicherungspflichtige Beschäftigung – sofern eine solche vorliegt – den Wert 0 für Level1 erhält. Außerdem wird innerhalb dieser Sortierung immer die Beschäftigung mit dem höheren Entgelt mit dem Wert 0 gekennzeichnet.</p>

(Fortsetzung nächste Seite)

Detailbeschreibung	Leistungsempfangsdaten: Sortiert nach ➤ Versicherungsnummer (aufsteigend) ➤ Leistungsart Gruppe (aufsteigend) d.h. die Datensätze sind in der folgenden Reihenfolge sortiert: 1. ALG 2. ALHI 3. UHG 4. Beitrag zur Pflegeversicherung 5. Fehlkodierung
Besonderheiten [Qualitätshinweise]	Die Variable wird u.a. dazu genutzt, um die Haupt- und Nebenbeschäftigung zu identifizieren. Die Hauptbeschäftigung hat immer den Wert Null, alle anderen einen Wert größer Null.

4.2.25. Nationalität

Variablenlabel [Bezeichnung]	Staatsangehörigkeit
Variablenname [Attributname]	nation
Herkunft [Verwendung]	BeH, LeH
Bezug	Meldung
Datentyp	numerisch
Länge	3
Detailbeschreibung	Die Variable erfasst die Nationalität von Beschäftigten und Leistungsempfängern.
Besonderheiten [Qualitätshinweise]	Deutsche haben in der Variable den Wert Null. Die einzelnen Ausprägungen sind in den LIAB-Codeplänen dokumentiert.

4.2.26. Personengruppe

Variablenlabel [Bezeichnung]	Personengruppenschlüssel
Variablenname [Attributname]	pers_gr
Herkunft [Verwendung]	BeH
Bezug	Meldung
Datentyp	numerisch
Länge	3
Detailbeschreibung	<p>Das Personengruppenkennzeichen gibt an, welchem Personenkreis ein Beschäftigter zuzuordnen ist.</p> <p>BeH: Im Jahreszeitraummaterial (JZM) ab 1999 werden nur die Personengruppen 101, 102, 103, 105, 106, 112, 118, 119, 120, 140, 141, 142, 143, 201, 203, 205 geliefert sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen.</p> <p>Die BeH vor 1999 enthält alle Personengruppen, auch solche, die ab 1999 mit dem JZM nicht mehr geliefert werden. Bei der Betrachtung von Zeitreihen oder beim Vergleich von BeH mit BLH müssen gegebenenfalls entsprechende Personengruppen gefiltert werden.</p>
Besonderheiten [Qualitätshinweise]	<p>Die Variable kann u.a. dazu verwendet werden, um geringfügig Beschäftigte zu identifizieren. Eingetragen ist eine dreistellige numerische Zahl, die eine sehr detaillierte Unterscheidung von Beschäftigten erlaubt.</p> <p>Das Merkmal wird nur für BeH-Sätze gebildet. Es ist seit Einführung des neuen Meldeverfahrens 1999 in den Meldungen enthalten. Bei Nachmeldungen für die vorhergehenden Jahre, die erst ab 1999 eingegangen sind, ist er ebenfalls in den Meldungen enthalten. Für Meldungen, die vor 1999 eingegangen sind, wurde versucht anhand bestimmter Regeln eine Zuordnung zu den Personengruppen durchzuführen. In vielen Fällen ist das jedoch nicht eindeutig möglich. Die einzelnen Ausprägungen sind in den LIAB-Codeplänen aufgelistet.</p>

4.2.27. Quelle

Variablenlabel [Bezeichnung]	Quelle
Variablenname [Attributname]	quelle
Herkunft [Verwendung]	BLH
Bezug	Meldung
Datentyp	numerisch
Länge	1
Detailbeschreibung	Die Variable gibt an, ob es sich um eine Beschäftigungs- oder Leistungsmeldung handelt.
Besonderheiten [Qualitätshinweise]	keine

Ausprägungen: 1 Beschäftigungsmeldung 2 Leistungsmeldung

4.2.28. Tagesentgelt

Variablenlabel [Bezeichnung]	Tagesentgelt
Variablenname [Attributname]	tag_entg
Herkunft [Verwendung]	BLH
Datentyp	numerisch
Länge	9 (6.3)
Detailbeschreibung	<p>1) BeH-Meldungen</p> <p>In BeH-Sätzen weist das Merkmal jeweils das Brutto-Tagesentgelt des Beschäftigten aus. Es wird berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumtgelt und der Dauer des Zeitraums der ungesplitteten Originalmeldung¹⁸ in Kalendertagen und auf drei Nachkommastellen (Pfennig/Cent) gerundet. Das Tagesentgelt ist in der Regel bis Ende 1998 als DM-Betrag und ab 1999 als Euro-Betrag angegeben. Ob es sich bei den Entgeltangaben um DM- oder Euro-Beträge handelt, ist dem Merkmal euro (4.2.17.) zu entnehmen.</p>

(Fortsetzung nächste Seite)

¹⁸ Das Episodensplitting wird in Kapitel sechs erläutert.

<p>Detailbeschreibung</p>	<p>In den Jahren 1975-1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt, also Entgelt oberhalb der Geringfügigkeitsgrenze und bis zur Beitragsbemessungsgrenze in der gesetzlichen Rentenversicherung. Mit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren ab 1.1.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze ausgewiesen; die Beitragsbemessungsgrenze als obere Kappungsgrenze gilt weiterhin. Manchmal übersteigen die Entgeltangaben die Beitragsbemessungsgrenze dennoch. Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnet. Dabei ist es unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei Beschäftigungszeit oder Entgelthöhe handeln, was aber aufgrund der Versicherungsrelevanz dieser Informationen unwahrscheinlich ist.</p> <p>Die Geringfügigkeits- und Beitragsbemessungsgrenzen unterscheiden sich von Jahr zu Jahr und in West- und Ostdeutschland (maßgeblich ist jeweils der Standort des Betriebs).</p> <p>2) LeH-Meldungen</p> <p>Für LeH-Sätze gibt das Merkmal den täglichen Leistungssatz an. Ob es sich dabei um DM- oder Euro-Beträge handelt, ist dem Merkmal euro zu entnehmen.</p>
<p>Besonderheiten [Qualitätshinweise]</p>	<p>Im FDZ-Methodenreport Nr. 2 wird ein Verfahren vorgestellt, das Entgelte auf der Beitragsbemessungsgrenze in eine darüber liegende Normalverteilung überführt. Es wird empfohlen, die auf der nachfolgenden Seite abgedruckten Grenzen um weitere zwei bis drei DM (Euro) nach unten (obere Grenze) bzw. oben (untere Grenze) zu korrigieren, weil sich bereits kurz vor der Beitragsbemessungsgrenze in den IAB-Personendaten Häufungspunkte befinden, die nicht exakt einer log-normalen Verteilung folgen.</p>

Beitragsbemessungsgrenzen von 1990 bis 2004 in den LIAB-Personendaten:

	Renten- und Arbeitslosen-Versicherung <i>tägliches Entgelt</i>				Geringfügigkeitsgrenze <i>tägliches Entgelt</i>			
	DM	Euro	ln(DM)	ln(Euro)	DM	Euro	ln(DM)	ln(Euro)
Alte Bundesländer								
1.1. - 31.12.1990	207,07	105,71	5,3331	4,6607	15,50	7,914	2,7410	2,0686
1.1. - 31.12.1991	213,65	109,07	5,3643	4,6920	15,83	8,082	2,7620	2,0896
1.1. - 31.12.1992	222,90	113,79	5,4067	4,7344	16,44	8,395	2,7999	2,1276
1.1. - 31.12.1993	236,66	120,82	5,4666	4,7943	17,47	8,921	2,8608	2,1884
1.1. - 31.12.1994	249,81	127,53	5,5207	4,8484	18,46	9,424	2,9157	2,2433
1.1. - 31.12.1995	256,39	130,89	5,5467	4,8743	19,12	9,760	2,9507	2,2783
1.1. - 31.12.1996	262,25	133,88	5,5693	4,8969	19,39	9,901	2,9650	2,2926
1.1. - 31.12.1997	269,54	137,60	5,5967	4,9244	20,10	10,264	3,0010	2,3286
1.1. - 31.12.1998	276,11	140,96	5,6208	4,9485	20,43	10,432	3,0172	2,3448
1.1. - 31.12.1999	279,40	142,64	5,6327	4,9603	20,76	10,599	3,0331	2,3608
1.1. - 31.12.2000	281,92	143,92	5,6416	4,9693	20,71	10,570	3,0304	2,3581
1.1. - 31.12.2001	285,98	145,99	5,6559	4,9836	20,76	10,599	3,0331	2,3608
1.1. - 31.12.2002	289,75	147,92	5,6690	4,9967	20,98	10,710	3,0436	2,3712
1.1. - 31.12.2003	322,21	164,49	5,7752	5,1029	20,98	10,710	3,0436	2,3712
1.1. - 31.12.2004	325,37	166,10	5,7850	5,113	25,81	13,176	3,2508	2,5784
Neue Bundesländer								
1.7. - 31.12.1990	88,72	45,29	4,4855	3,8131	6,63	3,38	1,8909	1,2186
1.1. - 30.6.1991	98,58	50,33	4,5909	3,9185	7,28	3,72	1,9855	1,3132
1.7. - 31.12.1991	111,73	57,04	4,7161	4,0437	8,27	4,22	2,1125	1,4402
1.1. - 31.12.1992	157,33	80,32	5,0583	4,3860	9,89	5,05	2,2911	1,6188
1.1. - 31.12.1993	174,20	88,93	5,1602	4,4878	12,87	6,57	2,5550	1,8827
1.1. - 31.12.1994	193,92	99,00	5,2675	4,5951	14,52	7,41	2,6752	2,0029
1.1. - 31.12.1995	210,36	107,39	5,3488	4,6765	15,50	7,91	2,7410	2,0686
1.1. - 31.12.1996	222,90	113,79	5,4067	4,7344	16,44	8,39	2,7999	2,1276
1.1. - 31.12.1997	233,37	119,14	5,4526	4,7803	16,82	8,59	2,8224	2,1501
1.1. - 31.12.1998	230,09	117,46	5,4385	4,7661	17,15	8,75	2,8418	2,1694
1.1. - 31.3.1999	236,66	120,82	5,4666	4,7943	17,47	8,92	2,8608	2,1884
1.4. - 31.12.1999	236,66	120,82	5,4666	4,7943	20,76	10,60	3,0331	2,3608
1.1. - 31.12.2000	232,74	118,81	5,4499	4,7776	20,71	10,57	3,0304	2,3581
1.1. - 31.12.2001	239,95	122,50	5,4804	4,8081	20,76	10,60	3,0331	2,3608
1.1. - 31.12.2002	241,45	123,26	5,4867	4,8143	20,76	10,60	3,0331	2,3608
1.1. - 31.12.2003	268,50	137,07	5,5928	4,9205	20,76	10,60	3,0331	2,3608
1.1. - 31.12.2004	274,82	140,30	5,6161	4,9438	25,81	13,18	3,2508	2,5784

4.2.29. Ost-West-Kennzeichen

Variablenlabel [Bezeichnung]	Ost-West-Kennzeichen
Variablenname [Attributname]	ost_west
Herkunft [Verwendung]	BeH, LeH, IAB-Betriebspanel
Datentyp	numerisch
Länge	1
Detailbeschreibung	In der ersten Version des LIAB-Längsschnittmodells befindet sich eine Variable, mit der Personen identifiziert werden können, die in beiden Landesdatensätzen mit ihrem Personenkonto vertreten sind.
Besonderheiten [Qualitätshinweise]	Beim Zusammenspielen der Personendaten der Personendatenbanken aus Ost- und Westdeutschland muss einer der entsprechenden Personendatensätze mit der Ausprägung drei gelöscht werden, um Doppelmeldungen zu vermeiden.

Wert	Label
1	West
2	Ost
3	Ost und West

4.2.30. Arbeitsamt am Wohnort

Variablenlabel [Bezeichnung]	Arbeitsamtbezirk des Wohnortes
Variablenname [Attributname]	wohn_aa
Herkunft [Verwendung]	BeH, LeH
Datentyp	numerisch
Länge	3
Detailbeschreibung	In Beschäftigtendatensätzen ist das Wohnortarbeitsamt ab dem Jahr 1999 eingetragen. In LeH-Datensätzen ist das Wohnortarbeitsamt in allen Jahren verfügbar.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Arbeitsamtsbezirke) befinden sich in den LIAB-Codeplänen.

4.2.31. Kreiskennziffer des Wohnortes

Variablenlabel [Bezeichnung]	Kreiskennziffer des Wohnortes
Variablenname [Attributname]	wohn_kr
Herkunft [Verwendung]	BLH
Datentyp	numerisch
Länge	5
Detailbeschreibung	In Beschäftigungs- und Leistungsdaten ist die Kreiskennziffer ab dem Jahr 1999 enthalten.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Kreiskennziffern) befinden sich in den LIAB-Codeplänen.

4.2.32. Wirtschaftszweig nach BA-Klassifizierung 1973

Variablenlabel [Bezeichnung]	Wirtschaftszweig nach WZ 73
Variablenname [Attributname]	wz73
Herkunft [Verwendung]	BeH
Datentyp	numerisch
Länge	3
Detailbeschreibung	Die Klassifizierung der Wirtschaftszweige basiert auf der BA-Klassifikation von 1973 und ist bei Beschäftigungsmeldungen gefüllt.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Wirtschaftszweige) befinden sich in den LIAB-Codeplänen. Die Variable ist bis zum Jahr 2002 durchgängig gefüllt, enthält aber – gerade vor 1999 – einige Fehlkodierungen (etwa sechs Prozent).

4.2.33. Wirtschaftszweig nach BA-Klassifizierung 1993

Variablenlabel [Bezeichnung]	Wirtschaftszweig nach WZ93
Variablenname [Attributname]	wz93
Herkunft [Verwendung]	BeH
Datentyp	numerisch
Länge	5
Detailbeschreibung	Die Klassifizierung der Wirtschaftszweige basiert auf der BA-Klassifikation von 1993 und ist bei Beschäftigungsmeldungen ab dem Jahr 1999 gefüllt.
Besonderheiten [Qualitätshinweise]	Die einzelnen Ausprägungen (Wirtschaftszweige) befinden sich in den LIAB-Codeplänen. Die Variable gibt es ab dem Jahr 1999.

4.3. Logische Zusammenhänge zwischen Variablen

In diesem Abschnitt werden die bisher bekannten Inkonsistenzen und logischen Abhängigkeiten zwischen Variablen dokumentiert. Die Auflistung erhebt keinen Anspruch auf Vollständigkeit und wird laufend aktualisiert. Das Forschungszentrum ist dankbar für jeden weiteren Hinweis von Anwendern auf weitere Besonderheiten und Inkonsistenzen der LIAB-Daten. Die logischen Zusammenhänge zwischen den LIAB-Variablen werden in zwei Gruppen eingeteilt und in je eigenen Unterabschnitten behandelt. Erstens gibt es logische Zusammenhänge (und Inkonsistenzen) zwischen den einzelnen Variablen der Personendatenbanken. Zweitens gibt es logische Zusammenhänge (und Inkonsistenzen) zwischen Variablen, die sich auf den gleichen Sachverhalt beziehen und die Information wahlweise aus dem IAB-Betriebspanel oder den LIAB-Personendatenbanken generiert werden kann¹⁹. Logische Zusammenhänge nur zwischen Variablen des IAB-Betriebspanels hingegen werden in diesem Abschnitt nicht behandelt. Dafür greifen Anwender auf die Dokumentationen des IAB-Betriebspanels zurück.

¹⁹ Die Persondaten werden dafür auf die Betriebsebene aggregiert.

4.3.1. Logische Zusammenhänge in den LIAB-Personendatenbanken

4.3.1.1. Berufsstellung und Personengruppen

Zu beachten ist, dass die Berufsstellung nicht immer konsistent mit den Personengruppenschlüsseln ist. Exemplarisch ist nachfolgend eine Häufigkeitsauszählung aus dem LIAB-Längsschnittmodell für die Personengruppen bei der Angabe Berufsstellung gleich Auszubildende abgedruckt²⁰. Die absoluten Werte werden aus Datenschutzgründen bei weniger häufig besetzten Personengruppenschlüsseln nicht ausgewiesen.

Personengruppe	Häufigkeit	Personengruppe	Häufigkeit
101	12.679	112	7
102	937.999	118	7
103	25	120	/
105	16.706	141	6
106	2.003	999	/
109	517		

Die Personengruppenausprägungen 102 (Auszubildende), 105 (Praktikanten) und 141 (Auszubildende in der Seefahrt) sind konform mit den Auszubildendenmeldungen²⁰. Die Ausprägung 106 (Werkstudenten) ist ein Grenzfall. Dagegen handelt es sich bei den Ausprägungen 103 (Beschäftigte in Altersteilzeit), 109 (geringfügig Beschäftigte), 112 (mithelfende Familienangehörige) und 120 (Personen, bei denen eine Beschäftigung vermutet wird) wohl um Fehlkodierungen bzw. Anwender müssen mit Hilfe anderer Variablen (z.B. dem Tagesentgelt) entscheiden, welche der beiden Angaben (Personengruppenschlüssel oder Stellung im Beruf) nach Prüfung der anderen Variablen nicht stimmig ist.

²⁰ Die Ausprägung 101 wird für Beschäftigte in einer geförderten Fortbildungsmaßnahme eingetragen (vgl. 4.2.8.)

4.3.1.2. Variablen quelle, beginn und ende (nur LIAB-Längsschnittmodell)

Aufgrund des Episodensplittings (vgl. Kapitel 6) treten zeitliche Überschneidungsmeldungen zwischen Beschäftigungs- und Leistungszeiten einer Person nur auf, wenn es sich um ein gleiches Beginn- und Enddatum handelt. In diesem Fall ist es in den meisten Fällen einfacher für nachfolgende Rekodierungen oder für die Bildung neuer Variablen, die Leistungsmeldungen in einem solchen Fall zu löschen. Wenn Anwender beispielsweise die Betriebszugehörigkeitsdauer berechnen, würde die gleichzeitige Leistungsmeldung eine eventuell vorhandene kumulativ längere Betriebszugehörigkeitsdauer unterbrechen. Es gibt auch andere Variablenaufbereitungsroutinen, bei denen das gleichzeitige Vorliegen von Beschäftigungs- und Leistungsmeldungen störend ist. Die Empfehlung für Anwender ist die Löschung der Leistungsmeldungen in einem solchen Fall („Beschäftigung schlägt Leistung“)③. Eine andere Möglichkeit ist die Trennung von Beschäftigungs- und Leistungsdaten in je eigene Datensätze, die nach der Datenaufbereitung wieder zusammengefügt werden.

4.3.1.3. Vollzeitbeschäftigung und Tagesentgelt

Vollzeitbeschäftigte können über die Stellung im Beruf identifiziert werden. Bei Lohnanalysen werden häufig bestimmte Personengruppen aus der Analyse ausgeschlossen. Das unter 4.3.1. Gesagte wird hier berücksichtigt und Personengruppen wie Auszubildende u.a. aus den Analysen ausgeschlossen. Monatsverdienste sind oft besser und einfacher zu beurteilen als Tagesentgelte. Die Variable tag_entg wird mit 30,5 multipliziert und freigestellte Personen aus der Analyse ausgeschlossen. Damit lässt sich in der ersten Version des LIAB-Längsschnittmodells (West) der folgende Output produzieren²¹:

Percentiles		Smallest		
1%	177.219	.0833333		
5%	846.375	.0833333		
10%	1369.575	.0833333	Obs	2967946
25%	2056.369	.0833333	Sum of Wgt.	2967946
50%	2681.41		Mean	2738.5
		Largest	Std. Dev.	1255.494
75%	3467.167	244091.5		
90%	4357.907	270779	Variance	1576266
95%	4423.992	276086	Skewness	34.51609
99%	4514	406219.3	Kurtosis	7035.951

Das 1-Prozent-Perzentil der Einkommensverteilung vollzeitbeschäftigter Arbeitnehmer liegt bei etwa 177 Euro. Das ist für eine Vollzeitbeschäftigung ein sehr geringes Einkommen. Anwender überlegen bei der Verwendung der Lohnvariable oft, ob sie nicht den Wertebereich beschränken (beispielsweise als Untergrenze für eine Vollzeitbeschäftigung 500,- Euro, als Obergrenze bietet sich – sofern keine Lohnimputation erfolgt – die Beitragsbemessungsgrenze an²¹). Nicht auszuschließen ist, dass es sich bei den beispielhaft verdeutlichten Inkonsistenzen um Fehlkodierungen handelt (also beispielsweise Auszubildende/Praktikanten versehentlich mit den Berufsstellungen eins bis vier gemeldet werden). Im einfachsten Fall werden Anwender unplausible Entgelte auf missing setzen. Meist erfolgt das in einer eigens generierten Variable, damit die Originalwerte im Datensatz verfügbar bleiben.

²¹ Warum die Lohnvariable die obere Beitragsbemessungsgrenze überschreiten kann, beschreibt 4.2.28.

4.3.1.4. Abgabegrund und Dauer (nur Beschäftigtenmeldungen)

Wie bereits in 4.2.1 unter den Qualitätshinweisen angedeutet, haben Jahresmeldungen nicht immer eine Dauer von 365 (in Schaltjahren 366) Tagen, wie die beispielhafte Auswertung für das Jahr 2002 zeigt⁵.

dauer

	Percentiles	Smallest		
1%	204	1		
5%	245	1		
10%	306	1	Obs	1781776
25%	365	1	Sum of Wgt.	1781776
50%	365		Mean	351.9023
		Largest	Std. Dev.	39.64603
75%	365	365		
90%	365	365	Variance	1571.808
95%	365	365	Skewness	-3.446899
99%	365	365	Kurtosis	16.43204

Mehr als zehn und weniger als 25 Prozent aller Meldungen mit dem Abgabegrund „Jahresmeldung“ haben keine Dauer von 365 Tagen. Ein möglicher inhaltlicher Grund hierfür sind während eines Kalenderjahres erfolgte Änderungsmeldungen des Arbeitgebers, ohne dass Beschäftigte den Betrieb verlassen. Für die Beschäftigten wird eine Jahresmeldung abgegeben, ohne dass die Dauer tatsächlich 365 Tage beträgt. Eine andere, eher „technische“ Ursache ist das Episodensplitting (vgl. Kapitel sechs).

4.3.2. Logische Zusammenhänge zwischen Variablen des IAB-Betriebspanels und von auf die Betriebsebene aggregierten LIAB-Personendaten

In diesem Abschnitt wird auf Zusammenhänge zwischen Angaben aus dem IAB-Betriebspanel und den LIAB-Personendaten eingegangen. Grundsätzlich wird empfohlen, entweder nur die Angaben aus dem IAB-Betriebspanel oder nur die aus den LIAB-Personendatenbanken für einen Betrieb zu verwenden und die Angaben aus beiden Datenquellen in einer Analyse nicht zu mischen. Die Gründe dafür werden im Verlauf des Abschnitts erläutert. Bei den Personendateninformationen handelt es sich um auf die Betriebsebene aggregierte Kennzahlen. Wie solche Kennzahlen berechnet werden, beschreibt Abschnitt 9.2.1..

4.3.2.1. Ost/Westkennzeichen in Personendatenbanken und im IAB-Betriebspanel

Das Ost/West-Kennzeichen der Personendaten ist auf der Betriebsebene nicht konsistent. Die folgende beispielhafte Auswertung bezieht sich auf das Jahr 2002 im LI-AB-Querschnittsmodell, Version 1. Zuerst wird der Mittelwert für das Ost/West-Kennzeichen (Ost hat den Wert Null, West den Wert eins) auf der Betriebsebene gebildet und für jeden Betrieb eine Beobachtungen im Datensatz gehalten (vgl. 9.2.1.). Die Daten werden dann mit dem IAB-Betriebspanel verknüpft und die Variable help nach Ost und West ausgewertet⁵. Der Befehl nach erfolgreichem Mergen produziert den folgenden Output:

```
wo2002 = west
```

```
help
```

```
-----
      Percentiles      Smallest
  1%          .984375          0
  5%           1            0
 10%           1            0      Obs          9643
 25%           1            0      Sum of Wgt.    9643

 50%           1
                        Largest      Mean          .9983862
 75%           1            1      Std. Dev.     .0318808
 90%           1            1      Variance     .0010164
 95%           1            1      Skewness    -29.41672
 99%           1            1      Kurtosis    905.6522
```

```
-> wo2002 = ost
```

```
help
```

```
-----
      Percentiles      Smallest
  1%           0            0
  5%           0            0
 10%           0            0      Obs          4947
 25%           0            0      Sum of Wgt.    4947

 50%           0
                        Largest      Mean          .0122285
 75%           0            1      Std. Dev.     .075202
 90%          .0074074          1      Variance     .0056553
 95%          .0362694          1      Skewness     10.02226
 99%          .3125            1      Kurtosis     117.8804
```

Wenn das Ost/West-Kennzeichen in den Personendaten konsistent ist, würde im Westen der Wert eins, im Osten Null betragen. Es zeigt sich aber, dass es geringe Abweichungen von diesem Erwartungswert gibt. Daher wird empfohlen, dass Ost/West-Kennzeichen nach Möglichkeit aus dem IAB-Betriebspanel zu verwenden.

4.3.2.2. Lohnsumme im IAB-Betriebspanel und in den Personendaten²²

Die beiden Angaben lassen sich in der überwiegenden Mehrheit aller befragten Betriebe nicht direkt miteinander vergleichen. Zur Demonstration wird beispielhaft das Jahr 2002 im LIAB-Querschnittmodell, Version 1 miteinander verglichen. Nach der üblichen Variablenaufbereitung²³ erhält man den folgenden Output des Vergleichs beider Angaben²³.

----- differenz -----				
	Percentiles	Smallest		
1%	-416706.5	-9.89e+07		
5%	-72726.94	-2.81e+07		
10%	-28431.69	-1.23e+07	Obs	13002
25%	-3115.139	-6883016	Sum of Wgt.	13002
50%	691.0481		Mean	62649.75
		Largest	Std. Dev.	1436075
75%	12521.48	1.91e+07		
90%	87013.75	2.11e+07	Variance	2.06e+12
95%	288804.8	2.40e+07	Skewness	3.505959
99%	1685892	1.02e+08	Kurtosis	3727.716

Bei etwa 40 Prozent der Betriebe ist die im IAB-Betriebspanel angegebene Lohnsumme geringer als die aus den LIAB-Personendaten. Die Gründe hierfür können sehr vielschichtig sein, ohne dies im Einzelfall genauer bestimmen zu können. Möglich sind beispielsweise Erinnerungslücken bei den Befragten, aber auch Inkonsistenzen der Betriebs- und Personendaten (vgl. Kapitel acht). Eine weitere Quelle für die Abweichungen sind Sonderzahlungen, die in der entsprechenden Frage nach der Lohn- und Gehaltssumme des IAB-Betriebspanels nicht angegeben werden sollen, aber in den Entgeltzahlungen an die Beschäftigten versicherungsrelevant sind und daher in den Lohnsummen enthalten sind. In etwa 60 Prozent der Betriebe wird im IAB-Betriebspanel eine höhere Lohnsumme angegeben, als sie die LIAB-Personendatenbank errechnet. Das ist plausibel, weil die Lohnsummen der nicht sozialversicherungspflichtig Beschäftigten in den LIAB-Personendatenbanken nicht enthalten sind.

²² Der STATA-Programmcode für diesen Unterabschnitt im Anhang A1 beschreibt auch auf einfache und illustrative Weise das Zusammenspielen von Betriebs- und Personendaten.

²³ Die Variable „differenz“ ist die Lohnsumme laut IAB-Betriebspanel minus die Lohnsumme laut LIAB-Personendatenbanken.

Die zusammenfassende Statistik zeigt noch etwas anderes. Gültige Angaben zur Lohnsumme machen 13002 Betriebe, während 14590 Betriebe Entsprechungen in den LIAB-Personendatenbanken haben²⁴. Wenn die Lohnsumme mit Personendaten verknüpft wird – beispielsweise als durchschnittliche betriebliche Lohnsumme pro Beschäftigtem – ergeben sich – beispielsweise in Regressionsanalysen - leicht Ausfälle durch item-non-response in einer Größenordnung von etwa einer halben Millionen Beschäftigten, nämlich von allen Beschäftigten, die in einem Betrieb arbeiten, der keine Angaben zu seiner Lohn- und Gehaltssumme macht. Wie die Auszählung der Differenzen zeigt, ist eine Ersetzung der fehlenden Werte des Fragebogens durch die auf die Betriebsebene aggregierte Lohnsumme aus den Beschäftigtendaten nicht empfehlenswert. Jensen/Rässler (2005) beschreiben eine Möglichkeit, mit hohen Anteilen an fehlenden Werten in einzelnen IAB-Betriebspanelvariablen umzugehen.

²⁴ Nicht alle (Querschnitts-)Betriebe des IAB-Betriebspanels haben eine Entsprechung in den LIAB-Personendatenbanken (vgl. Kapitel 8).

4.3.2.3. Wirtschaftszweigklassifizierungen

Die Angaben beider Datenquellen sind miteinander vergleichbar, weil die Brancheneinteilung des IAB-Betriebspanels letztlich auf der Wirtschaftszweigsystematik der BLH-Daten basiert. Während das IAB-Betriebspanel nach 41 (vor 2000: 44) Branchen differenziert, ist über die Personendaten der vollständige WZ73- bzw. WZ93-Code verfügbar. Letzterer gestattet beispielsweise eine Rekodierung der Wirtschaftszweige nach NACE/STAN. Die WZ73 ist für alle Erhebungswellen des IAB-Betriebspanels in den Jahren 1993 bis derzeit 2002 verfügbar, sodass damit auch eine Lösung für den Bruch der Wirtschaftssystematik im IAB-Betriebspanel von 1999 auf 2000 existiert.

4.3.2.4. Anzahl sozialversicherungspflichtig Beschäftigter

Ausführungen hierzu befinden sich im Kapitel acht.

4.3.2.5. Personalstruktur

Für die meisten Beschäftigtengruppen ist ein direkter Vergleich der Fragebogenangaben und der Personendatenbanken möglich. Die betrieblichen Anteile von nicht sozialversicherungspflichtigen Beschäftigtengruppen sind obligatorisch nur über das IAB-Betriebspanel verfügbar. Manche Anteile, wie beispielsweise befristet Beschäftigte, haben zwar eine Entsprechung in den Personendatenbanken, die aber nicht an konkreten Personen eindeutig verankert werden kann (es sei denn, ein Betrieb beschäftigt nur oder gar nicht befristet).

Besser vergleichbar ist die Personalstrukturfrage im IAB-Betriebspanel und die auf die Betriebsebene aggregierten Personalstrukturinformationen. Beispielhaft werden die betrieblichen Auszubildendenanteile aus den beiden Datenquellen in ihrer Verteilung miteinander verglichen⁸. Die Variable `azubi_anteil` ist die Angabe aus der LIAB-Personendatenbank 2002 und `azubi_panel` die Angabe aus dem IAB-Betriebspanel im Jahr 2000.

azubi_anteil					
Percentiles		Smallest			
1%	0	0			
5%	0	0			
10%	0	0	Obs		14517
25%	0	0	Sum of Wgt.		14517
50%	.8403361		Mean		5.912798
		Largest	Std. Dev.		12.10069
75%	6.766917	100			
90%	16.66667	100	Variance		146.4268
95%	25	100	Skewness		4.223744
99%	65.32258	100	Kurtosis		26.3603

azubi_panel					
Percentiles		Smallest			
1%	0	0			
5%	0	0			
10%	0	0	Obs		14517
25%	0	0	Sum of Wgt.		14517
50%	.8474576		Mean		6.199469
		Largest	Std. Dev.		12.48414
75%	7.00637	100			
90%	17.30769	100	Variance		155.8537
95%	28.57143	100	Skewness		3.967041
99%	66.66666	100	Kurtosis		23.46883

Der Median und der Mittelwert sind sehr ähnlich, abweichende Verteilungen ergeben sich vor allem bei größeren Auszubildendenanteilen an der betrieblichen Gesamtbelegschaft. Hier ist zu vermuten, dass die Angaben aus den LIAB-Personendaten genauer sind, während bei den Befragungsdaten (IAB-Betriebspanel) insbesondere bei einer Vielzahl an Auszubildenden Ungenauigkeiten, Schätzwerte und Erinnerungslücken möglich sind. Auch hier gilt die allgemeine Empfehlung, nach Möglichkeit entweder die Angabe aus dem IAB-Betriebspanel *oder* die aus den LIAB-Personendatenbanken zu nutzen.

5. Datenquellen

Wie bereits einleitend beschrieben, handelt es sich bei den LIAB-Daten um eine Verknüpfung des IAB-Betriebspanels mit der Beschäftigten-Leistungsempfänger-Historik-Datei des IAB. Die Grundgesamtheit des IAB-Betriebspanels sind alle Betriebe in Deutschland mit mindestens einem sozialversicherungspflichtig Beschäftigten. Aus ihr werden ein Jahr vor der Feldphase die Betriebe der Stichprobe für das kommende Jahr nach einer Matrix von zehn Betriebsgrößenklassen (vgl. Tabelle 1) und 20 (bis 1999: 16) Wirtschaftszweigen gezogen. Da es zu dieser Zeit noch keine Beschäftigtendaten für das aktuelle Befragungsjahr gibt, basiert die Hochrechnungsmatrix für das Befragungsjahr auf der betrieblichen Vorjahresbeschäftigung²⁵.

Die Personendaten wiederum liegen erst mit einer Zeitverzögerung von etwa zwei Jahren vor. Früher liegen die Beschäftigtenmeldungen nicht *vollständig* im IAB vor. Nach etwa zwölf Monaten liegen nach dem 31. Dezember eines jeden Jahres etwa 90 Prozent aller Beschäftigtenmeldung vor. Für die Verknüpfung mit den Betriebsdaten ist das kein ausreichender Wert, weil auf der betrieblichen Ebene davon ausgegangen werden muss, dass Beschäftigtenmeldung nicht zufällig fehlen. So wird es nach zwölf Monaten immer einige Betriebe geben, die für alle ihre Beschäftigten die Meldungen abgegeben haben, andere haben einige ihrer Beschäftigten gemeldet und wieder andere noch gar keine. Von einer Vollständigkeit der Nachmeldungen ist nach den bisherigen Erfahrungen im IAB etwa 24 Monate nach dem 31. Dezember eines jeden Jahres auszugehen. Entsprechend können LIAB-Daten nur mit einer Zeitverzögerung von zwei Jahren zum aktuellen Rand zur Verfügung gestellt bzw. aktualisiert werden.

Die Verknüpfung mit den Leistungsempfangsdaten hingegen ist keine zeitliche Restriktion, weil diese direkt aus den Geschäftsstatistiken der BA generiert werden und im Gegensatz zu den Beschäftigtenmeldung nicht auf Umwegen über die Kranken- und Rentenversicherer zur BA gelangen. Wenn die BLH-Daten für ein neues Jahr vorliegen, werden bis zur Bereitstellung im FDZ Zeiten von etwa drei Monaten für die Stichprobenziehung, die Aufbereitung der Rohdaten und die Pseudoanonymisierung veranschlagt.

²⁵ Mit der Konsequenz, dass sich bei einer Hochrechnung der Wert der offiziellen Arbeitsmarktstatistiken zu sozialversicherungspflichtiger Beschäftigung im Vorjahr (relativ zum Befragungsjahr) ergibt.

6. Datenaufbereitung

Einleitend wurde darauf hingewiesen, dass eine Bereitstellung aller Meldungen für alle Betriebe des IAB-Betriebspanels und der Beschäftigten in ihnen über einen langen Zeitraum technisch nicht möglich ist und es auch kein besonders anwenderfreundliches Verfahren wäre, das zudem mit extrem langen Rechenzeiten verbunden ist. Diese Überlegungen münden in die LIAB-Datenmodelle und Versionen. Sie erschließen das Auswertungspotenzial der LIAB-Daten (je auf ihre modellhafte Weise), ohne dass dies zu übermäßigen technischen Handhabungsschwierigkeiten führt. Es wird darauf hingewiesen, dass sich je nach Untersuchungsgegenstand und -ansatz dennoch lange Rechenlaufzeiten ergeben können.

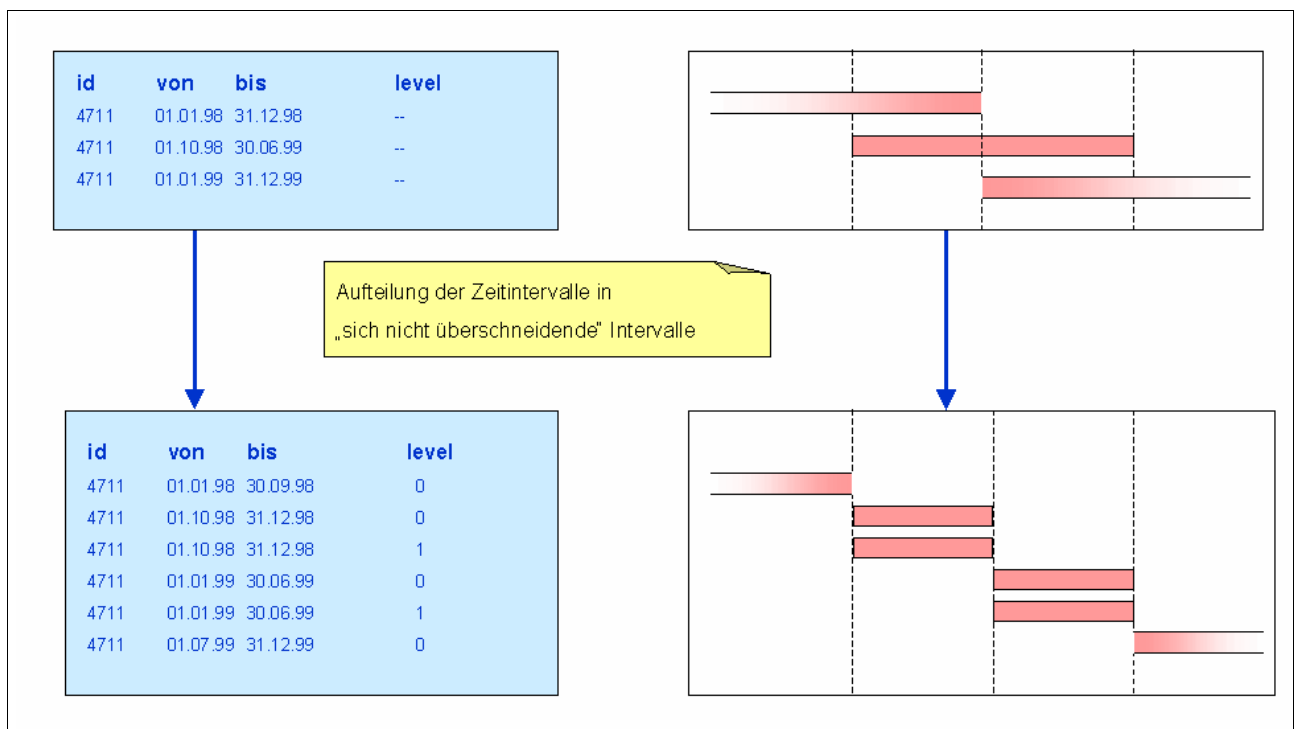
Anwendern wird das Personendatenmaterial möglichst unverändert zur Verfügung gestellt. Das hat Vor- und Nachteile. Die Vorteile bestehen sicherlich darin, dass durch Bereinigungsverfahren etc. die Anlegung einer gewissen Pfadabhängigkeit vermieden wird. Nachteile entstehen dadurch, dass Anwender sich relativ tief in die Meldeverfahren einarbeiten müssen, um wissenschaftlich verwertbare Daten und Ergebnisse zu generieren. Aus diesem Grund enthält dieses LIAB-Datenhandbuch Vorschläge, wie mit bestimmten Variablen umgegangen werden kann. Sofern diesen Vorschlägen gefolgt wird – was bei der überwiegenden Mehrheit der Anwendungen möglich sein sollte – entfallen diese Einarbeitungskosten zu einem erheblichen Teil.

LIAB-Personendatenbanken beinhalten einige kleine Veränderungen gegenüber dem Originaldatenmaterial, die für Anwender nicht sichtbar werden. Zum einen gilt das für die datenschutzrechtliche Vorgabe, dass die Variablennamen in den LIAB-Personendaten gegenüber den Originalvariablennamen verändert sind. Das hat auf die Daten selber keinen Einfluss. Es gibt aber auch einige wenige kleinere Datenänderungen, die sich im Hintergrund vollziehen. Ein wichtiger Bereinigungsverfahren der Personendaten ist das Episodensplitting.

6.1. Das Episodensplitting

Bei der Erstellung der BLH-Gesamtdatei wird ein Episodensplitting durchgeführt. Dabei werden bei Überschneidungen von Zeiträumen innerhalb eines Kontos diese Spells so geschnitten und dupliziert, dass vollständig parallele Zeiträume und Zeiträume ohne Überschneidung entstehen. Die Anzahl der Meldungen erhöht sich dadurch.

Abbildung 1: Grafische Darstellung des Episodensplitting



Durch das Episodensplitting werden Meldungen verdoppelt und der Gültigkeitszeitraum ändert sich. Die ursprünglichen Datumsvariablen für Beginn und Ende des Gültigkeitszeitraums sind auf Anfrage verfügbar.

6.1.1. Kontenbeispiel für das Episodensplitting²⁶

In dem nachfolgenden Kontenbeispiel wird gezeigt, welche Werte die kontobezogenen Merkmale in den gesplitteten Spells annehmen, wenn eine Überschneidung innerhalb einer Quelle vorliegt. Oben sind die ungesplitteten Sätze grafisch dargestellt, darunter die gesplitteten sowie die Werte der kontobezogenen Merkmale:

Das so genannte Episodensplitting wird durchgeführt, wenn es zeitliche Überschneidungen zum Beispiel von Beschäftigungsverhältnissen (Mehrfachbeschäftigung) gibt. Damit auch diese Informationen für Analysen verwertbar sind, werden mehrere Meldungen so aufgeteilt, dass Anfangs- und Enddatum übereinstimmen.

Meldefolge vor Episodensplitting:

```
|-----|                Hauptbeschäftigung
                |-----|                geringfügige Beschäftigung
```

Meldefolge nach Episodensplitting:

```
|-----|                Hauptbeschäftigung
                |-----|                Hauptbeschäftigung
                |-----|                geringfügige Beschäftigung
                |-----|                geringfügige Beschäftigung
```

Beim Episodensplitting werden Variablen erzeugt, die für den Umgang mit den Daten hilfreich sind. Gesplittete Spells, die ursprünglich in einem Spell zusammengefasst waren, haben dieselbe Satznummer (satznr). Die Satznummer ermöglicht einen Rückbezug auf die Beschäftigtensätze (BeH) und Leistungsbezugssätze (LeH). Sie ist eindeutig für ungesplittete Sätze, bei gesplitteten Sätzen kommt dieselbe Satznummer mehrfach vor. Bei BeH-Sätzen enthält die Satznummer in den ersten 4 Stellen das Jahr und anschließend einen neunstelligen Satzzähler. Bei LeH-Sätzen steht statt dem Jahr der Wert '9999' in den ersten 4 Stellen der Satznummer.

²⁶ Auszug aus: Silke Hamann: Die IAB-Regionalstichprobe 1975-2001 (originale Version), Dezember

Beim Episodensplitting werden außerdem Level-Variablen erzeugt, die die Zahl der Überschneidungen anzeigen. Die Variable „level2“ gibt wie bisher für die jeweilige Episode an, um den wievielten zeitlich parallel liegenden Satz (innerhalb einer gegebenen Sortierfolge) es sich handelt²⁷. Die Level2-Variable hat den Wert 0 für jeweils den ersten Spell, der für eine Episode vorliegt. Jeder weitere Spell für dieselbe Episode wird dann hochgezählt. Um also z.B. alle Beschäftigungszeiträume (bzw. Leistungsempfangszeiten) jedoch ohne Mehrfachbeschäftigung zu erhalten, müsste der Filter level2 gleich Null und quelle gleich eins. gesetzt werden. Welcher Spell bei mehreren Spells, die denselben Zeitraum abdecken, in der Variablen level2 den Wert 0 erhält, hängt von der vor dem Episodensplitting durchgeführten Sortierung ab. Diese sieht folgendermaßen aus:

Versicherungsnummer (aufsteigend)

Episoden-Beginndatum (aufsteigend)

Status (aufsteigend)

Für Sätze aus der Beschäftigten-Historik des IAB (BeH) ist das Tagesentgelt (absteigend) maßgeblich, bei Sätzen aus der Leistungsempfänger-Historik des IAB (LeH) ist die Reihenfolge für den Status zufällig. Die Sortierung führt dazu, dass erstens bei zeitlichen Überschneidungen Beschäftigungsmeldungen vor LeH-Meldungen stehen, dass zweitens innerhalb von gleichzeitigen Beschäftigungen geringfügige Beschäftigungen nach hinten sortiert werden und dass schließlich innerhalb dieser Reihenfolge zuerst die Meldungen mit dem höheren Entgelt kommen. Zeitliche Überschneidungen, die nach dem Splitting nur einen Tag betragen, wurden gelöscht, da davon ausgegangen wurde, dass es sich hierbei um Falschmeldungen der Kassen handelt. Darüber hinaus werden doppelte Meldungen gelöscht. Als Identitätskriterium dienten dabei die Versicherungsnummer, das Beginn- und Enddatum, der Abgabegrund, die Betriebsnummer und das Entgelt. Wenn es geringe Unterschiede in den Angaben gibt, beispielsweise in den Nachkommastellen des Entgelts, bleiben beide Sätze erhalten.

2004, S. 4-6.

²⁷ Demgegenüber gibt die Variable „level1“ an, um den wievielten zeitlich parallel liegenden Satz aus derselben Quelle (also LeH oder BeH) es sich handelt.

6.2. Bereinigung von Versicherungsnummern und Konsequenzen für die Aktualisierung der LIAB-Versionen

Jedes Jahr melden die Rentenversicherungsträger im Rahmen der Ergebnisse ihrer Versichertenkontenklärung stillgelegte oder falsch registrierte Versicherungsnummern an die BA. Ersteres ist ein recht seltenes Ereignis (etwa 70 von zwei Millionen Sozialversicherungsnummern werden in einem Zehnjahreszeitraum als nicht gültig oder doppelt identifiziert), letzteres ist geringfügig bedeutsamer (etwa 1500 von zwei Millionen Versicherungsnummern enthalten eine falsche Zahl oder einen falschen Buchstaben). Diese Korrekturen, die die Datenqualität erhöhen, werden in der den LIAB-Personendatenbanken zugrunde liegenden Datenschicht, der BLH vorgenommen.

Bei Aktualisierungen der LIAB-Versionen am aktuellen Rand werden die Personendatenbanken in der Regel komplett neu aufgebaut, d.h. auch bereits vorliegende Jahre werden bei einem Aktualisierungslauf neu gezogen. Aufgrund der Bereinigungen bei den Versicherungsnummern ergeben sich daher für ein und dasselbe Jahr geringfügige Abweichungen der LIAB-Personendaten je nach Aktualisierungsstand. Fast ausschließlich ändert sich dann im Einzelfall nur die Versicherungsnummer, während sich Meldeinhalte in der überwiegenden Mehrheit nicht betroffen sind. Analyseergebnisse für nicht aktuelle Zeiträume ändern sich also in der Regel nicht, aber es kann zu Umsortierung bei der Anonymisierung der Versicherungsnummer kommen, d.h. das der gleiche Personenidentifikator (id) in verschiedenen aktualisierten Versionen nicht unbedingt die gleiche Person kennzeichnet. Innerhalb einer aktualisierten Version tritt dieses Problem natürlich nicht auf, wie das nächste Kapitel zeigen wird. Derzeit ist kein Fall bekannt, wo die beschriebene Veränderung des Grunddatenmaterials von Bedeutung ist.

7. Datenanonymisierung

7.1. Datenschutzbestimmungen²⁸

Bei den Daten, die durch das FDZ Forschern zur Verfügung gestellt werden und deren Nutzung Beratungsgegenstand ist, handelt es sich in der Regel um Sozialdaten gemäß § 67 SGB X bzw. um Daten, die Sozialdaten gleichgestellt sind. Sozialdaten sind Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlichen Person, die von der BA im Hinblick auf ihre Aufgaben nach dem SGB III erhoben und verarbeitet und dann im IAB für die Forschung aufbereitet werden. Sozialdaten sind damit besonders sensible Daten, die nach § 35 Abs. 1 SGB I als Sozialgeheimnis zu schützen sind.

Dieser besondere Schutz spiegelt sich in den Regelungen zur Datenweitergabe für die Forschung und Planung wider. Der Zugang zu *nicht anonymisierten, schwach anonymisierten oder pseudonymisierten Sozialdaten*²⁹ durch die Forschung und Planung wird in §75 SGB X geregelt.

Als Bedingungen für eine solche Nutzung sind festgelegt:

- Die Daten müssen für ein Projekt aus dem Sozialleistungsbereich verwendet werden,
- das öffentliche Interesse muss das Geheimhaltungsinteresse des Betroffenen erheblich überwiegen,
- das Einholen der Einwilligung des Betroffenen muss unzumutbar sein und
- der Zweck des Vorhabens darf nicht auf andere Weise erreichbar sein (z.B. durch die Analyse von anderen Daten, zu denen ein einfacherer Zugang möglich ist).

²⁸ Der Abschnitt ist aus: Allmendinger/Kohlmann (2005).

²⁹ Unter ‚schwacher‘ Anonymisierung wird die systemfreie Umschlüsselung von Identifikatoren im Datensatz bei Erhalt des vollständigen Merkmalsumfangs verstanden. Bei der Pseudonymisierung werden Identifikatoren systematisch umgeschlüsselt bei Erhalt des vollständigen Merkmalsumfangs.

Die Übermittlung von Sozialdaten³⁰ muss beim BMWA beantragt und von ihm genehmigt werden. Es werden umfangreiche Maßnahmen zur Gewährleistung von Datenschutz und Datensicherheit bei den Datenempfängern vorausgesetzt. Grundsätzlich ist eine Nutzung der Daten für kommerzielle Interessen ausgeschlossen. Insofern Sozialdaten dahingehend verändert werden, dass Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft reidentifiziert werden können (§16 Abs. 6 Bundesstatistikgesetz, §67 Abs. 8 SGB X), gelten die Daten als faktisch anonymisiert. In diesem Fall handelt es sich nicht mehr um Sozialdaten. Faktisch anonymisierte Daten der BA und des IAB können an wissenschaftliche Einrichtungen auf Antrag übermittelt werden, wenn es sich bei dem Forschungsvorhaben um ein Projekt aus der Arbeitsmarkt- und Berufsforschung handelt (§282 Abs. 7 SGB III), kommerzielle Forschungsinteressen sind auch hier ausgeschlossen. Neben der Anonymisierung von Personendaten ist das IAB auch verpflichtet, Betriebsdaten faktisch zu anonymisieren (vgl. § 67 Absatz 1 Satz 2 SGB X). Die faktische Anonymität ist dann gegeben, wenn ein Datenangreifer *unverhältnismäßig viel Zeit, Kosten und Arbeitskraft* aufbringen muss, um einen Datensatz zu deanonymisieren.

³⁰ Die externe Übermittlung von Sozialdaten der BA und des IAB nach § 75 SGB X kann beim Datenzentrum der Statistik der BA (e-mail: Service-Haus.Statistik-Datenzentrum@arbeitsagentur.de) beantragt werden. Die dabei entstehenden Kosten werden dem Antragsteller in Rechnung gestellt. Das Datenzentrum der Statistik erstellt auch kostenpflichtige Sonderauswertungen aus den amtlichen Statistiken der BA zur Lage und Entwicklung auf dem Arbeits- und Ausbildungsmarkt, zur Beschäftigung, zur Arbeitsförderung einschließlich der Eingliederungsbilanz sowie zu den Lohnersatz- und sonstigen Leistungen.

7.2. Die schwache Anonymisierung der LIAB-Originaldaten

Aus 7.1. ergeben sich für die LIAB-Daten die Konsequenz, die Sozialversicherungsnummer auf Seiten der Personen und die Betriebsnummer auf Seiten der Betriebe durch geeignete systemfreie Identifikatoren zu ersetzen (eindeutige Identifikatoren).

7.2.1. Anonymisierung der Sozialversicherungsnummer

Die Sozialversicherungsnummer wird in eigenständigen Datensätzen jeweils systemfrei ersetzt. Nur in miteinander vernetzten Datensätzen wie den einzelnen Jahresscheiben muss eine Datensatzübergreifende Anonymisierung vorgenommen werden. Das bedeutet, dass in den LIAB-Längsschnittversionen die Anonymisierung durch die Aneinanderreihung der folgenden Arbeitsschritte erfolgt:

- (i) Sortieren nach Versicherungsnummer
- (ii) Ersetzen des Buchstabens durch eine numerische Zahl
- (iii) Zusammensetzung einer numerischen Versicherungsnummer
- (iv) Sortieren nach dieser numerischen Nummer
- (v) Eine Hilfsvariable wird auf eins in der ersten Meldung der Person gesetzt
- (vi) Die Hilfsvariable wird über den gesamten Datensatz aufaddiert. Das ist der systemfreie Personenidentifikator
- (vii) Die originale Sozialversicherungsnummer wird in einer den Forschern nicht zugänglichen Schlüsseldatei hinterlegt und dann aus dem Datensatz gelöscht

Im LIAB-Querschnittsmodell werden zunächst die Versicherungsnummern aus allen Wellen zusammengespielt und jede Versicherungsnummer einmal behalten. Dann erfolgen die Schritte (i) bis (vii). Die so erzeugte Datei bestehend aus dem systemfreien Personenidentifikator und der Sozialversicherungsnummer wird wieder mit den einzelnen Jahresscheiben verknüpft und die in dem jeweiligen Jahr nicht vorhandenen Sozialversicherungsnummern gelöscht. Abschließend wird die Sozialversicherungsnummer aus den Jahresdatensätzen entfernt. Im Ergebnis erhält man einen

eindeutigen systemfreien Identifikator, wo jede Person in jedem Jahr eindeutig und über die Jahre konsistent identifiziert werden kann.

7.2.2. Anonymisierung der Betriebsnummer

In Versionen des LIAB-Querschnittmodells geschieht die Anonymisierung der Betriebsnummer durch die Verknüpfung der variable `idnum` aus dem IAB-Betriebspanel mit der Betriebsnummer in den Personendaten. Es findet ein einfacher Austausch statt und die Betriebsnummer wird gelöscht. Die Daten des IAB-Betriebspanels können konsistent mit den Personendaten verknüpft werden.

Im LIAB-Längsschnittmodell muss ein mehrstufiges Konzept Anwendung finden. Zu unterscheiden sind drei Typen von Betrieben: (1) Betriebe, die für die LIAB-Längsschnittversion ausgewählt werden (2) sonstige IAB-Betriebspanelbetriebe (3) alle anderen Betriebe. Zuerst werden die Längsschnittpersonendaten mit den IAB-Betriebspanelnummern (`idnum`) verknüpft. Dabei wird in einem vorgeschalteten Schritt eine Variable erzeugt, die angibt, ob der betreffende Betrieb aus dem IAB-Betriebspanel zu den ausgewählten Betrieben der spezifischen Längsschnittversion gehört. Die IAB-Betriebspanelnummer ersetzt dann wie im LIAB-Querschnittmodell die originale Betriebsnummer. Es verbleiben die Betriebsnummern der Betriebe, die nicht an der Befragung teilgenommen haben. In diesen Betrieben arbeiten Personen vor oder nach ihrer Beschäftigung in einem der IAB-Betriebspanelbetriebe. Damit es zu keinen Verknüpfungsproblemen mit den Betriebspaneldaten kommt, werden die systemfreien Betriebsidentifikatoren für nicht befragte Betriebe zunächst wie die systemfreien Personenidentifikatoren gebildet. Das Resultat ist eine natürliche Zahl zwischen 1 und N. Da diese Werte bis etwa 100 Tsd. bereits für die IAB-Betriebspanelbetriebe vergeben sind, wird zu der letzten im IAB-Betriebspanel vorkommenden Nummer 100.000 dazu addiert. Zu der erzeugten fortlaufenden Zahl der Nicht-IAB-Betriebspanelbetriebe wird dieser Wert dazuaddiert. So liegen die Identifikationsnummern der am IAB-Betriebspanel teilnehmenden und nicht teilnehmenden Betriebe weit genug auseinander. Es gibt keine Überschneidungsmenge zwischen systemfreien Identifikationsnummern.

Für Anwender ergeben sich aus dem geschilderten Vorgehen keine Konsequenzen. Der Status der Betriebe (1) bis (3) ist in der Variable `betr_st` dokumentiert (vgl. 4.2).

10). Wenn die Ost/West-Datensätze in Längsschnittversionen zusammengespielt werden, ist zu bedenken, dass bei den Nicht-IAB-Betriebspanelbetrieben im Extremfall Betriebsidentifikatoren gleich sein können, obwohl sie unterschiedliche Ausgangsbetriebe darstellen. Praktisch ist das aber nahezu bedeutungslos, weil solche Betriebsnummern und die dazugehörigen Personenmeldungen nicht mit IAB-Betriebspanelinformationen angereichert werden können. Ergebnisse, die für Personen auf die Betriebsebene aggregiert werden, müssen auf die Betriebe in der Auswahl für die jeweilige Längsschnittversion beschränkt bleiben, denn für die Betriebe, die nicht zu der Auswahl für die jeweilige Version gehören, können keine vollständigen (bezüglich der Beschäftigten) Aggregationen auf die Betriebsebene erfolgen, weil nur Beschäftigte erfasst sind, die aus einem IAB-Betriebspanelbetrieb „irgendwann einmal“ in diesen Betrieb wechseln.

8. Datenqualität und Probleme

Bei der Datenverknüpfungen des IAB-Betriebspanels mit der Beschäftigten-Leistungsempfänger-Historik können Probleme auftreten. Einige Betriebe beispielsweise können den Personendaten nicht zugeordnet werden. Bei anderen tauchen Konsistenzprobleme auf. Auf die Gründe der Inkonsistenzen kann an dieser Stelle nicht ausführlich eingegangen werden. Im FDZ-Methodenreport Nr. 1 wird das ausführlicher als in diesem Kapitel diskutiert.

Ein geeignetes Maß und damit Prüfgröße für die Übereinstimmung der Daten des IAB-Betriebspanels und den LIAB-Personendatenbanken ist die Gesamtanzahl an sozial-versicherungspflichtig Beschäftigten. Die Angabe liegt als Befragungsdatum und als Aggregat der Personendaten auf die Betriebsebene vor. Für die Klassifizierung der Verknüpfungsqualität wird aus beiden Datenquellen (Befragungsdaten, prozessproduzierte Daten) die Angabe des aktuellen Befragungsjahres verwendet. Bei Längsschnittbetrieben hingegen werden die Information der Vorjahresbeschäftigung aus dem Organisationsfile des IAB-Betriebspanels und die der entsprechenden Vorjahresscheibe miteinander verglichen. Nach diesen Vergleichsprozeden wurden zusätzlich Längsschnittbetriebe als innerhalb der Toleranzgrenzen liegend klassifiziert, wenn sie aufgrund des Vorjahresvergleichs zwar (knapp) außerhalb, beim Vergleich der aktuellen Anzahl der sozialversicherungspflichtig Beschäftigten aus dem Fragebogen und der auf die Betriebsebene aggregierten Anzahl der Beobachtungen in der aktuellen Jahresscheibe innerhalb der definierten Toleranzgrenzen liegen. Aufgrund von Erinnerungslücken und daraus resultierenden Ungenauigkeiten bei der Angabe für einen vergangenen Stichtag werden (kleinere) Abweichungen bei den Angaben aus beiden Datenquellen zugelassen³¹ und als prozentuale Abweichungen bestimmt. Bei einem solchen Vorgehen sollten je nach Betriebsgröße unterschiedliche Toleranzintervalle gelten, weil ein Ein- oder Austritt von Beschäftigten bezogen auf die Gesamtbelegschaft in kleineren Betrieben eine höhere prozentuale Veränderung hervorruft als in größeren.

Es werden folgende Toleranzwerte für die Differenz der betrieblichen Fragebogenangaben von dem entsprechenden Ergebnis der Aggregation der Personendaten auf

³¹ Zwar bezieht sich die Abfrage immer auf den 30. Juni, allerdings können natürlich nicht alle Interviews zeitnah und damit quasi gleichzeitig durchgeführt werden.

die Betriebsebene bei der späteren Berechnung einer Qualitätskontrollvariablen zugelassen:

- (1) Betriebe bis fünf Beschäftigte: 40 Prozent
- (2) Betriebe mit fünf bis 19 Beschäftigten : 30 Prozent
- (3) Betriebe mit 20 bis 100 Beschäftigten: 20 Prozent
- (4) Betriebe mit mehr als 100 Beschäftigten: 10 Prozent

Mit den Toleranzgrenzen werden betriebliche Personalschwankungen zu mehr als neunzig Prozent eingeschlossen. Solche Schwankungen der Mitarbeiterzahl können demnach bei der vorgenommenen Definition als Quelle für nicht deckungsgleiche Angaben aus dem IAB-Betriebspanel und der Beschäftigten-Leistungsempfänger-Historik ausgeschlossen werden. Wenn Betriebe nicht im Toleranzintervall liegen, kann die gleichzeitige Verwendung von Personen- und IAB-Betriebspaneldaten problematisch sein³². Die prozentuale Abweichung ist in einer zusätzlichen Variable dokumentiert, die auf Wunsch externen Forschern bei ihrem Gastaufenthalt im IAB zur Verfügung gestellt wird³³.

Bei der nachfolgenden Qualitätsklassifikation in Tabelle 2 wird diese Information über die prozentuale Abweichung beispielsweise genutzt, um Betriebe mit Abweichungen von mehr als 100 Prozent gesondert auszuweisen. Die Ausweisung der Ergebnisse erfolgt für die gültigen Querschnittsfälle der jeweiligen Jahre.

³² Wenn ein Betrieb beispielsweise Fragebogenangaben für 1000 Beschäftigte macht (z.B. Investitionssumme), kann die simultane Nutzung von und Verknüpfung mit nur 100 vorliegenden Personendatensätzen Irritationen auslösen.

³³ Die Variable eignet sich zur Filterführung.

Tabelle 2: Klassifizierung der Verknüpfungsqualität von IAB-Betriebspanel- und den entsprechenden Daten aus der Beschäftigten-Leistungshistorikdatei des IAB (alle Befragungsjahre 1993 bis 2001)

a) Westdeutschland

absolute Anzahl westdeutsche Betriebe	1993	1994	1995	1996	1997	1998	1999	2000	2001
im Toleranzbereich	-	2665	2433	2369	2333	2345	2493	2342	4079
nach Fragebogenangaben im Toleranzbereich	2769	357	547	556	567	846	775	2968	2503
keine Verknüpfung mit Individualdaten möglich	76	55	117	147	222	178	155	292	290
nicht im Toleranzbereich	1270	982	879	812	866	817	842	2192	2172
nicht im Toleranzbereich, Kleinstbetrieb*	111	50	59	113	51	67	74	223	404
Abweichung mehr als 100 Prozent	39	13	61	32	63	36	88	399	400
insgesamt	4265	4166	4096	4029	4102	4289	4427	8416	9848
anteilig westdeutsche Betriebe	1993	1994	1995	1996	1997	1998	1999	2000	2001
im Toleranzbereich	-	64%	59%	59%	57%	55%	56%	28%	41%
nach Fragebogenangaben im Toleranzbereich	65%	9%	13%	14%	14%	20%	18%	35%	25%
keine Verknüpfung mit Individualdaten möglich	2%	1%	3%	4%	5%	4%	4%	3%	3%
nicht im Toleranzbereich	30%	24%	21%	20%	21%	19%	19%	26%	22%
nicht im Toleranzbereich, Kleinstbetrieb*	3%	1%	1%	3%	1%	2%	2%	3%	4%
Abweichung mehr als 100 Prozent	1%	0%	1%	1%	2%	1%	2%	5%	4%
insgesamt	100%	100%	100%	100%	100%	100%	100%	100%	100%

* mit weniger als fünf Beschäftigten

b) Ostdeutschland

absolute Anzahl ostdeutsche Betriebe	1996	1997	1998	1999	2000	2001
im Toleranzbereich	-	2882	2999	3153	3494	3397
nach Fragebogenangaben im Toleranzbereich	2740	883	922	1124	817	1036
keine Verknüpfung mit Individualdaten möglich	103	151	128	161	167	193
nicht im Toleranzbereich	1260	705	709	733	765	759
nicht im Toleranzbereich, Kleinstbetrieb*	147	30	56	71	171	191
Abweichung mehr als 100 Prozent	63	97	90	93	101	112
insgesamt	4313	4748	4904	5335	5515	5688
anteilig ostdeutsche Betriebe	1996	1997	1998	1999	2000	2001
im Toleranzbereich	-	61%	61%	59%	63%	60%
nach Fragebogenangaben im Toleranzbereich	64%	19%	19%	21%	15%	18%
keine Verknüpfung mit Individualdaten möglich	2%	3%	3%	3%	3%	3%
nicht im Toleranzbereich	29%	15%	14%	14%	14%	13%
nicht im Toleranzbereich, Kleinstbetrieb*	3%	1%	1%	1%	3%	3%
Abweichung mehr als 100 Prozent	1%	2%	2%	2%	2%	2%
insgesamt	100%	100%	100%	100%	100%	100%

* mit weniger als fünf Beschäftigten

Quelle: Personendaten LIAB-Querschnittsmodell, Version 1 und IAB-Betriebspanel 1993 bis 2001, Organisationsdatenfile 2002

Auch die Betriebe mit bis zu fünf Beschäftigten werden gesondert ausgewiesen, weil bei ihnen nur ein Beschäftigungsein- oder -austritt zu einer wenigstens 20-prozentigen Beschäftigungsveränderung führt. Damit wird ein Verlassen des Toleranzbereichs wahrscheinlicher und die hier beschriebenen Einordnungskriterien könnten bei den kleinsten Betrieben problematisch sein. Wie die Tabelle 2 zeigt, sind

diese Betriebe aber für das Gesamtergebnis nicht sonderlich bedeutsam. Mit jährlichen Schwankungen liegen durchschnittlich etwa drei von vier IAB-Betriebspanelbetrieben in dem definierten Toleranzintervall, sofern dazu die ersten beiden ausgewiesenen Kategorien gezählt werden. Die fehlende Zuordnung zu den Personendaten ist ein eher marginales Problem und, wie weiterführende Analysen zeigen, fast ausschließlich auf kleinste Betriebe zurückzuführen. Die letzten drei Kategorien beschreiben Betriebe, die nicht in den definierten Toleranzintervallen liegen.

Eine numerische Variable mit den in Tabelle 2 ausgewiesenen Informationen und eine weitere, die bei Betrieben außerhalb des Toleranzbereichs über die festgestellte prozentuale Abweichung informiert, ist auf Anfrage im FDZ verfügbar. Der entsprechende Datensatz enthält die systemfreien Betriebsidentifikatoren der Querschnittsfälle 1993 bis 2001 aus dem IAB-Betriebspanel, eine Jahresinformation sowie die beiden interessierenden Variablen der Klassifizierungsqualität und der prozentualen Abweichung. Die hier benannten Informationen sind also kompatibel abrufbar für alle Versionen der LIAB-Datenmodelle. Der Datensatz umfasst insgesamt mehr als 78 Tsd. Beobachtungsjahre der Betriebe. Vertiefende Informationen zum Themenkomplex enthält der FDZ-Methodenreport Nr. 1.

9. Testdaten und Hinweise für Datenauswertungen

9.1. Hinweise und Anregungen für die Datenauswertungen

Die nachfolgenden Ausführungen sind als eine mögliche Herangehensweise an die LIAB-Daten zu verstehen. Selbstverständlich sind auch andere Vorgehensweisen denkbar. Die Umsetzung der nachfolgenden Vorschläge mag im Einzelfall nicht auf alle inhaltlichen Fragestellungen anwendbar sein. Das Forschungsdatenzentrum übernimmt keine Gewähr für die Richtigkeit und Vollständigkeit der nachfolgenden Ausführungen. Die Sammlung wird laufend aktualisiert und erweitert.

9.1.1. Grundsätzliches zur Verknüpfung der Betriebs- und Personendaten

Die Verknüpfung der Betriebs- und Personendaten ist neben allen inhaltlichen Datensatzaufbereitungen ein in jeder LIAB-Auswertung durchzuführender Arbeitsschritt. Da die verknüpften Datensätze meist beachtliche Größen erreichen, wird die getrennte Aufbereitung der Betriebs- und Personendaten empfohlen. Nicht benötigte Variablen sollten in einer Anwendung zur Abkürzung der Rechenzeiten möglichst frühzeitig gelöscht werden. Ebenfalls ist die getrennte Aufbereitung der Betriebs- und Personendaten zu empfehlen, beispielsweise um Kodierungsprobleme und -fehler möglichst eindeutig und schnell zu identifizieren. Für die Variablen des IAB-Betriebspanels kommt hinzu, dass die Variablenaufbereitung wesentlich schneller geht, wenn sie noch nicht mit den Personendaten verknüpft sind. Bei dem Aufbau von Paneldatensätzen wird insbesondere für das LIAB-Querschnittmodell empfohlen, die Daten jahresweise aufzubereiten und danach die Betriebs- und Personendaten für jedes Jahr getrennt zu verknüpfen. Erst danach wird aus den einzelnen Jahresdatensätzen die Panelstruktur aufgebaut. Anwender müssen dann daran denken, jede Variable, die sich auf den gleichen Sachverhalt bezieht, in jedem Jahr gleich zu benennen. Bei dem Aufbau eines Paneldatensätzen ist zudem die Generierung einer Jahresvariable erforderlich.

Die Aufbereitung der IAB-Betriebspaneldaten kann bereits vor dem Gastaufenthalt über das Schalterstellenmodell erfolgen (Datenfernverarbeitung). Das erspart nicht nur Zeit vor Ort, sondern ermöglicht bereits im Vorfeld Überlegungen zu möglichen Schwierigkeiten (z.B. geringe Fallzahlen, Ausfälle aufgrund von item-non-response) und ihrer Behebung.

9.1.2. Planung und Umsetzung der Forschungsinhalte

Aufgrund der Komplexität der LIAB-Daten ist eine zielorientierte Fokussierung auf die Erkenntnisinteressen notwendig. Die Erstellung einer Auswertungsstrategie und die Zerlegung in einzelne Arbeitsschritte vor der eigentlichen Programmierarbeit sind oftmals hilfreich. Einzelne Planungsschritte münden meist in je einzelne Auswertungsprogramme. Die Arbeitsergebnisse eines jeden solchen Auswertungsprogramms werden in einem Zwischendatensatz abgesichert. Das erleichtert die spätere Kontrolle jedes einzelnen Arbeitsschrittes. Datenkorrekturen, die in den Auswertungsprogrammen dokumentiert sind und nicht „per Hand“ vorgenommen werden, erleichtern die spätere Nachvollziehbarkeit und das Aufspüren etwaiger Ungenauigkeiten oder Fehler.

9.1.3. Umgang mit den Personendatenbanken

Vor den eigentlichen Datenaufbereitungsarbeiten ist bei vielen Anwendungen fraglich, ob alle Personen im Datensatz benötigt werden. Die nachfolgende Aufzählung ist ein Auszug häufig vorgenommener Selektionen von Personen-/Beschäftigtengruppen.

- Auszubildende ja/nein
- Teilzeitbeschäftigte ja/nein
- Berücksichtigung von Personen mit Freistellungsphasen ja/nein (Entgelt ist gleich Null in entsprechenden Meldungen)
- Berücksichtigung von geringfügig Beschäftigten, die seit 1999 gesondert erfasst werden (der Ausschluss erfolgt über den Personengruppenschlüssel)
- Berücksichtigung bzw. Umgang mit Personen, die mehr als ein Beschäftigungsverhältnis gleichzeitig ausüben (beispielsweise kann nur die Hauptbeschäftigung im Datensatz gehalten werden; der Ausschluss erfolgt dann über den Spellzähler *pro Episode*)

Insbesondere im LIAB-Längsschnittmodell wird empfohlen, den Umgang mit Meldungen, die eigentlich nicht benötigt werden, sorgfältig zu prüfen. Ein Beispiel ist die Betriebszugehörigkeitsdauer. Wenn Meldungen – welche auch immer – vor der Bildung der Betriebszugehörigkeitsdauervariable gelöscht wird, entstehen Meldelücken, die zu irritierenden Ergebnissen führen können. Grenzfälle sind beispielsweise geringfügige Beschäftigungszeiten zwischen zwei Vollzeitmeldungen, die in einem Betrieb ohne zeitliche Meldelücken erfolgen. Die eine Frage bezieht sich darauf, ob es sich um ein oder zwei Beschäftigungsverhältnisse handelt, die andere, ob die geringfügige Beschäftigungszeit mit zur Betriebszugehörigkeitsdauer gerechnet werden wird.

9.1.4. Varianz der Betriebsmerkmale nach der Verknüpfung mit den Personendaten und Gewichtung

Die Varianz von Betriebsmerkmalen weicht bei einer IAB-Betriebspanelauswertung in der Regel von der in einer LIAB-Auswertung mit Personendaten ab. Verantwortlich hierfür ist das disproportionale Stichprobendesign des IAB-Betriebspanels und den entsprechenden Konsequenzen bei dem Aufbau der LIAB-Personendatenbanken (vgl. FDZ-Datenreport Nr. 1). Wenn ein Betriebsmerkmal insbesondere in kleineren Betrieben variiert, in größeren jedoch weniger, dann ist in nach der Verknüpfung mit der Personendatenbank eine sehr geringe Varianz des Merkmals zu beobachten. Ein möglicher Ausgleich ist die Gewichtung der Analyseergebnisse. Dann müssen Anwender – neben inhaltlichen Überlegungen zur Gewichtung der Analyseergebnisse - daran denken, den Gewichtungsfaktor aus den Betriebsdaten mit den Personendaten zu verknüpfen.

9.1.5. Hinweise zu den Variablen in den LIAB-Personendatenbanken

Sensible Variablen entstehen durch Zusatzinformationen bei der Verknüpfung von Betriebs- und Personendaten, wenn dadurch die Reidentifikation von Betrieben oder Personen erleichtert wird. In der nachfolgenden Übersicht 2 sind diese Variablen aufgelistet.

Übersicht 2: Sensible Variablen in den LIAB-Daten

Sensible Merkmale in LIAB-Personendatenbanken	
AA-Bezirk Arbeitsort	AA-Bezirk Wohnort
Kreiskennziffer Arbeitsort	Kreiskennziffer Wohnort
Wirtschaftszweig nach Klassifikation WZ 73	Wirtschaftszweig nach Klassifikation WZ 93

Die Sensibilität entsteht durch die relativ tiefe Untergliederung, die im Einzelfall eine eindeutige Identifikation von Betrieben ermöglicht, wenn die Betriebs- und Personendaten miteinander verknüpft werden. Die Merkmale stehen im Gastaufenthalt für Auswertungen nur zur Verfügung, wenn im Antrag die Nutzung der sensiblen Merkmale gesondert begründet wird.

Im LIAB-Querschnittmodell stehen einige Personenmerkmale zur Verfügung, die im Regelfall für die Auswertungen nicht benötigt werden. In der Übersicht 3 sind diese Variablen aufgelistet.

Übersicht 3: selten genutzte Personenmerkmale im LIAB-Querschnittmodell

oftmals entbehrliche Personenmerkmale im LIAB-Querschnittsmodell	
Spellzähler personenbezogen	Beginn- und Enddatum
Abgabegrund d. Meldung	Dauer d. Meldung in Tagen
Kennz. Datumsbereinigung	Ost/West-Kennzeichen

Der versichertenbezogene Spellzähler kennzeichnet, um die wievielte Meldung es sich im Rahmen aller Meldungen der Person seit dem 01. Januar 1991 (West) bzw. 1992 (Ost) handelt. Diese Information ist in einem Querschnittsmodell, das nur die Meldung zum 30. Juni eines jeden Beobachtungsjahres von einer Person in einem IAB-Betriebspanelbetrieb beinhaltet, nicht besonders aussagekräftig. Vom Beginn- und Enddatum einer solchen Meldung (und damit der Dauer der Meldung in Tagen) kann nicht eindeutig auf den Verbleib oder das Ausscheiden einer Person aus dem entsprechenden Betrieb geschlossen werden. Solche Rückschlüsse sollten selbst dann nur mit Vorbehalt bei dem Einbezug des Abgabegrundes der Meldung gezogen werden, beispielsweise weil eine Person aus dem Betrieb ausscheiden kann und kurze Zeit später wieder eingestellt wird (im Extremfall einen Tag später). Das Ost/West-Kennzeichen wird nicht benötigt, weil die entsprechende Information über das IAB-Betriebspanel zur Verfügung steht.

9.2. Beispieldaten und -programme

9.2.1. Aggregation der Personendaten auf die Betriebsebene

Mit den LIAB-Personendaten des Querschnittmodells lassen sich interessante Informationen auf der Betriebsebene gewinnen. Beispielhaft wird nachfolgend die Altersstruktur in Betrieben (als Prozentanteile von Altersgruppen) und der betriebliche Durchschnittslohn für unterschiedliche Bildungsniveaus im Jahr 2002 berechnet. Sofern es der Wunsch von LIAB-Nutzern ist, werden auf den Internet-Seiten des FDZ in der LIAB-Rubrik weitere Beispielprogramme dokumentiert.

Zuerst wird der Datensatz eingelesen und die Beschäftigten in Altersgruppen kategorisiert.

```
use "X:\Nutzer\personendaten_liabquer2002.dta", clear

drop if alter<15

*** Kommentar: 247 Beobachtungen werden gelöscht

generate altersgruppe=1 if alter>=15 & alter<20
replace altersgruppe=2 if alter>=20 & alter<30
replace altersgruppe=3 if alter>=30 & alter<40
replace altersgruppe=4 if alter>=40 & alter<50
replace altersgruppe=5 if alter>=50 & alter<60
replace altersgruppe=6 if alter>=60 & alter<65
replace altersgruppe=7 if alter>=65 & alter~=.

#delimit ;
label define alt
1 "15 bis 19 Jahre"
2 "20 bis 29 Jahre"
3 "30 bis 39 Jahre"
4 "40 bis 50 Jahre"
5 "50 bis 60 Jahre"
6 "60 bis 65 Jahre"
7 "über 65 Jahre" ;

#delimit cr
```

Jetzt wird eine (0,1)-Variable für jede Person generiert.

```
tab altersgruppe, gen(alt_)
```

Im Ergebnis hat jede Person in einer der sieben (0,1)-Variablen eine eins und in den anderen eine Null. Der Vorteil dieser Vorgehensweise besteht darin, dass bei einer Aggregation auf die Betriebsebene der Mittelwert dieser (0,1)-Variablen dem be-

trieblichen Anteil der jeweiligen Gruppe an allen Beschäftigte entspricht. Da die Variablen fortlaufend nummeriert sind, kann eine Schleife eingesetzt werden.

```
local y=1

while `y'<8 {
  bysort idnum: egen alter_`y'=mean(alt_`y')
  replace alter_`y'=alter_`y'*100
  local y=`y'+1
}
```

In den Variablen alter_1 bis alter_7 befinden sich nun die angestrebten betrieblichen Anteile der gebildeten Altersgruppen. Jetzt werden die Durchschnittslöhne für verschiedene Beschäftigtengruppen am Beispiel der Stellung im Beruf berechnet.

```
cap drop help
gen help=tag_entg if berstell==0
bysort idnum: egen azubilohn=mean(help)

drop help
gen help=tag_entg if berstell==1
bysort idnum: egen angelerntenlohn=mean(help)

drop help
gen help=tag_entg if berstell==2
bysort idnum: egen facharbeiterlohn=mean(help)

drop help
gen help=tag_entg if berstell==3
bysort idnum: egen meisterlohn=mean(help)

drop help
gen help=tag_entg if berstell==4
bysort idnum: egen angestelltenlohn=mean(help)
drop help
```

Die Variablenaufbereitung ist damit abgeschlossen. Abschließend werden die gewünschten Variablen und für jeden Betrieb eine Beobachtung gehalten.

```
keep alter_* idnum *lohn
bysort idnum: keep if _n==1
save daten2002.dta, replace
```

Die Bildung von betrieblichen Strukturinformationen anhand der LIAB-Personendaten ist hervorragend für SPSS-Anwendungen geeignet, weil über den AGGREGATE-Befehl die Informationen in gleicher Weise wie in STATA aufbereitet werden können. Der Einfachheit halber werden nur die betrieblichen Alterstrukturen in dem nachfol-

genden SPSS-Auswertungsprogramm generiert. Die Durchschnittslöhne lassen sich analog wie im STATA-Beispiel berechnen.

```
GET
  FILE='X:\Nutzer\ personendaten_liabquer2002.sav'.

SELECT IF (alter ge 15).
EXEC.

COMPUTE alter1=0.
COMPUTE alter2=0.
COMPUTE alter3=0.
COMPUTE alter4=0.
COMPUTE alter5=0.
COMPUTE alter6=0.
COMPUTE alter7=0.
EXEC.

If (alter ge 15 AND alter lt 20) alter1=1.
If (alter ge 20 AND alter lt 30) alter2=1.
If (alter ge 30 AND alter lt 40) alter3=1.
If (alter ge 40 AND alter lt 50) alter4=1.
If (alter ge 50 AND alter lt 60) alter5=1.
If (alter ge 60 AND alter lt 65) alter6=1.
If (alter gt 65) alter7=1.
EXEC.

AGGREGATE
  /OUTFILE='X:\Nutzer\aggr.sav'
  /BREAK=idnum
  /alter1 = MEAN(alter1) /alter2 = MEAN(alter2) /alter3 =
MEAN(alter3) /alter4 = MEAN(alter4) /alter5 = MEAN(alter5) /alter6
  = MEAN(alter6) /alter7 = MEAN(alter7).
```

Die Leistung entspricht exakt dem STATA-Programm. Die gewünschten Daten befinden sich in der Datei aggr.sav. Die Werte für alter_1 bis alter_7 müssten mit 100 multipliziert werden, wenn die Auswertung von Prozentanteilen erfolgen soll. Fast immer wird anschließend der erhaltene Datensatz mit dem entsprechenden IAB-Betriebspaneljahresdatensatz verknüpft.

9.2.2. Bestimmung von Ein- und Austrittsraten

In diesem Abschnitt wird eine beispielhafte Anwendung mit dem LIAB-Querschnittsmodell in der Programmiersprache STATA vorgenommen, die Bestimmung von Ein- und Austritten von Personen aus Betrieben. Programmzeilen sind an der kursiven Schrift erkennbar. Im LIAB-Längsschnittmodell ist die Frage fast trivial, denn Ein- und Austritte können anhand der Betriebsnummer im Erwerbsverlauf mit zwei Befehlen identifiziert werden³⁴. Im LIAB-Querschnittsmodell hingegen verrät die Fragestellung einiges über die Arbeit mit den LIAB-Daten, insbesondere zum Verhältnis der Betriebs- und Personendaten. Die ausformulierte Fragestellung für die beispielhafte Auswertung lautet:

„Wie ist die Altersstruktur der vollzeitbeschäftigten betrieblichen Ein- und Austritte (ohne Auszubildende) in bzw. aus Beschäftigung in ost- und westdeutschen Betrieben im Jahr 2002, gewichtet und ungewichtet?“

Im LIAB-Querschnittsmodell bedeutet das zu allererst, dass die Beobachtungswellen 2001 und 2002 zusammengespielt werden müssen, denn ob Personen zum Jahr 2002 aus einem Betrieb aus- oder im Jahr 2002 in ihn eintreten, lässt sich nur mit einem Referenzzeitpunkt angeben, dem Jahr 2001. Zuerst werden also die Betriebe auf Seiten des IAB-Betriebspanels identifiziert, die an beiden Jahren am IAB-Betriebspanel teilgenommen haben. Dazu wird der Paneldatensatz des IAB-Betriebspanels (vgl. auch 9.3.3.) verwendet.

```
clear
set mem 200m
set more off
cd "X:\Nutzer"
capture log close
log using mobilität.log, replace
use paneldaten.dta, clear
keep if jahr==2001 | jahr==2002
bysort idnum: gen obs=_N
keep if obs==2
keep if jahr==2002
keep idnum wo2002 hochquer
sort idnum
save basis2002.dta, replace
```

³⁴ In der ersten Version des Längsschnittmodells West beispielsweise: (1) sort id spell_nr (2) by id: gen exit=1 if idnum~=idnum[_n-1] & betr_st==2. Mit diesen beiden Befehlen wird die Austrittsinformation in den ersten darauf folgenden Spell geschrieben. Wenn er in der letzten Beschäftigungsmeldung aus dem ausscheidenden Betrieb stehen soll, wird [_n-1] durch [_n+1] ersetzt. Wahlweise kann auch zwischen einem Austritt in Arbeitslosigkeit (betr_st==.) oder Beschäftigung (betr_st==1 | betr_st==2 | betr_st==3) unterschieden werden. Letzteres ist im Querschnittsmodell nicht möglich.

EXKURS zur Gewichtung

Mit einer solchen Vorgehensweise gehen die Panelausfälle für die Analysen verloren, also alle Betriebe, die 2002 erstmals an der Befragung teilnehmen. Werden die verbleibenden Betriebe mit dem richtigerweise anzuwendenden Querschnittshochrechnungsfaktor gewichtet, ergibt sich aufgrund der Ausfälle nicht der Wert in der Grundgesamtheit. Eine Behebungsmöglichkeit besteht darin, auf Basis der Hochrechnungszellen die ausfallenden Gewichtungsfaktoren auf die Panelbetriebe zu verteilen, die in beiden Jahren an der Befragung teilgenommen haben. Dazu muss das obige Programm leicht abgewandelt werden.

```
use paneldaten.dta, clear
keep if jahr==2001 | jahr==2002
gen obs=1 if jahr==2001
replace obs==2 if jahr==2002
bysort idnum: egen obs_kum=sum(obs)

*** Ergebnis: Betriebe nur in 2001 haben eine 1, nur in 2002 eine 2
und in beiden Jahren eine 3
*** Jetzt werden die Betriebe identifiziert, die nur in 2002 teilge-
nommen haben und ihr
*** Zellenhochrechnungswert gleichmäßig auf die Betriebe verteilt,
die in beiden Jahren
*** teilgenommen haben

cap drop help*
gen help=hochquer if obs_kum==2
bysort grkl_10 bra20fb02: egen help1=sum(help)
gen help2=1 if obs_kum==3 and jahr==2002
bysort grkl10 bra20fb02: egen help3=sum(help2)
gen zusatzgewicht=help1/help3
drop help*

*** Der Wert wird zu dem ursprünglichen Hochrechnungsfaktor hinzuad-
diert

keep if obs_kum==3
keep if jahr==2002
gen hochquer_korr=hochquer+zusatzgewicht
keep idnum wo2002 hochquer*
sort idnum
save basis2002.dta, replace
```

Das Verfahren führt insbesondere in größeren Betriebsgrößenklassen zu einer vertretbaren Korrektur. Bei den kleinsten Betrieben sind die Hochrechnungsfaktoren aus der Stichprobenanlage des IAB-Betriebspanels heraus bereits sehr groß (bis etwa zu

3000). Sie werden durch Zusatzgewichte stärker erhöht. Es wird ebenfalls der ursprüngliche Gewichtungsfaktor im Datensatz gehalten, um in den Analysen gegebenenfalls auf diesen ausweichen zu können.

 ENDE EXKURS

Dieser Basisdatensatz wird mit den Personendaten verbunden. Letztere werden zuerst vorbereitet, um die Rechenzeiten und den benötigten Speicherplatz zu minimieren. Bei nur zwei Jahren ist eine Schleife nicht sonderlich elegant, aber zur Demonstration findet sie hier Anwendung.

```
local y=2001

while `y'<2003 {
  use personendaten_liab_quer_`y'.dta, replace
  keep idnum alter berstell pers_gr
  gen vollzeit==1 if berstell>0 and berstell<5
  replace vollzeit==. if pers_gr>=102 & pers_gr<=106
  replace vollzeit==. if pers_gr==109 | pers_gr==209
  drop pers_gr berstell
  compress
  sort idnum
  save personendaten`y'.dta, replace
  local y=`y'+1
}
```

Wenn Anwender keine Schleifen verwenden möchten, ersetzen sie bitte das `y' durch konkrete Jahreszahlen und lassen die beiden ersten und letzten Zeilen weg. Resultat der Prozedur: zwei Datensätze (personendaten2001.dta und personendaten2002.dta) mit den Variablen idnum alter und vollzeit.

Der Einfachheit halber braucht nach der obigen Selektion nicht mehr darauf geachtet werden, dass Beschäftigte auch von Vollzeit in Teilzeit wechseln können, ohne den Betrieb zu verlassen. Sie werden hier als Ein- und Austritte (in Vollzeitbeschäftigung, sei es nun vom externen Arbeitsmarkt oder aus dem eigenen Betrieb) gewertet. Falls noch vorhanden, werden Doppelbeschäftigungen ausgeschlossen.

```
local y=2001

while `y'<2003 {
  use personendaten`y'.dta, clear
  drop if level1>0 & level1~=.
  sort idnum
  merge idnum using basis2002.dta
```

```

tab _merge
keep if _merge==3
drop _merge
gen year==`y'
save personendaten`y'_nurpanelbetriebe.dta, replace
local y=`y'+1
}
append using personendaten2001_nurpanelbetriebe.dta
compress
sort id
save auswertungsdatensatz.dta, replace

```

In diesem Datensatz befinden sich nun alle vollzeitbeschäftigten Personen in den Panelbetrieben der Jahre 2001 und 2002. Es wurde die Variable `year` generiert, damit bekannt ist, aus welchem Jahr die Personenjahrbeobachtung stammt. Dann wird analog zu der Bestimmung der Panelbetriebe verfahren.

```

cap drop obs*
gen obs=1 if year==2001
replace obs=2 if year==2002
bysort id idnum: egen obs_kum=sum(obs)
gen exit=1 if obs_kum==1
gen entry=1 if obs_kum==2

```

Jetzt wird die Forschungsfrage beantwortet.

```

cap drop help*
gen help=alter if exit==1
bysort wo2002: egen exit_meanage=mean(help)
gen help1=alter if entry==1
bysort wo2002: egen entry_meanage=mean(help1)
drop help*
bysort wo2002: sum *_meanage if _n==1
bysort wo2002: sum *_meanage [aweight=hochquer] if _n==1

```

Mit dem Datensatz kann jetzt auch sehr einfach die betriebliche Ein- und Austrittsrate bestimmt werden. Sie wird hier nach der Formel $2 \cdot E_t / (N_t + N_{t-1})$ gebildet, mit E als den Ein- oder Austritten und N als Gesamtbeschäftigung zum Zeitpunkt t bzw. $t-1$, also hier 2001 und 2002.

```

cap drop help*
gen help=1 if year==2002
bysort idnum: egen total2002=sum(help)
drop help
gen help=1 if year==2001
bysort idnum: egen total2001=sum(help)
drop help
bysort idnum: egen exit_no=sum(exit)
bysort idnum: egen entry_no=sum(entry)
gen exitrate=2*exit_no/(total2001+total2002)

```

```
*** jeder Betrieb wird nur einmal behalten ***
```

```
bysort idnum: keep if _n==1
```

```
bysort wo2002: sum exitrate entryrate, d
```

```
bysort wo2002: sum exitrate entryrate [aweight=hochquer], d
```

```
bysort wo2002: sum exitrate entryrate [aweight=hochquer_korr], d
```

Die betriebliche Wachstumsrate der Vollzeitarbeitsplätze erhalten Anwender mit

```
gen growth=2*(total2002-total2001)/(total2002+total2001)
```

```
log close
```

```
clear
```

```
exit
```

9.3. Beschreibung und Umgang mit den Testdaten sowie weitere Arbeitshilfen

9.3.1. Beschreibung der Testdaten des LIAB-Querschnittmodells

0.) Die Testdaten bestehen aus zwei Einzelteilen, den Betriebsdaten des IAB-Betriebspanels und Personendatenbanken (Jahresscheiben). Die Verknüpfung beider Datensätze erfolgt über den systemfreien Betriebsidentifikator (idnum).

1.) Die Testdaten sind so konzipiert, dass sie sowohl die Struktur der Personendaten wie auch die Verknüpfungsqualität repräsentieren. Aufgrund der Komplexität der Originaldaten kann das nur ansatzweise und stilisiert geschehen.

2.) Ähnlich wie bei den Testdaten zum IAB-Betriebspanel (und auch in den Originaldaten) gibt es über die Jahre Ausfälle auf der Betriebs- und Personenebene.

2a) Auf der Betriebsseite scheiden jedes Jahr sieben Betriebe aus dem Panel aus. Die größten Betriebe mit mindestens 250 Beschäftigten sind in allen Jahren 1993 bis 2001 vertreten. Das ist in den Originaldaten leider nicht immer so.

2b) Auf der Personenseite wird eine zufällig verteilte Panelmortalität erzeugt, die sich quer durch die Betriebe zieht. Insgesamt scheiden etwa 10 Prozent aller Beschäftigten jährlich aus dem Panel aus und etwa 10 Prozent kommen neu hinzu. In der Regel bleibt die Beschäftigtenzahl in den Betrieben unverändert. Letzteres ist in den Originaldaten eher selten der Fall.

3.) Angestrebt werden Fallzahlen in der Personendatenbank von jährlich etwa 5000.

4.) Auf der *Betriebseite* wurde für 2.) folgender Größenmix verwendet:

- 1 Betrieb mit 1000 Beschäftigten
- 2 Betriebe mit 500 Beschäftigten
- 4 Betriebe mit 250 Beschäftigten
- 5 Betriebe mit 100 Beschäftigten
- 10 Betriebe mit 50 Beschäftigten
- 20 Betriebe mit 25 Beschäftigten
- 40 Betriebe mit 10 Beschäftigten
- 20 Betriebe mit 5 Beschäftigten

Mit Beschäftigten sind immer Sozialversicherungspflichtige gemeint. Außerdem gibt es wenige Betriebe, die keine Entsprechung in der jeweiligen Personendatenbank haben.

5.) In den Betrieben bis 50 Beschäftigte gibt es einige wenige Betriebe, in denen die Zahl der sozialversicherungspflichtig Beschäftigten nicht mit der Anzahl der Personmeldungen übereinstimmt (etwa 5 Prozent). Das ist in den Originaldaten häufiger der Fall. Nur bei der Betriebsgröße stimmen mit dieser eingebauten Ungenauigkeit die Angaben in den Testdaten des IAB-Betriebspanels (z.B. a01svb93) und in der Personendatenbank für das jeweilige Jahr überein. Alle anderen Merkmale sind zufällig verteilt. Es gibt nahezu keine (sinnvollen) Entsprechungen anderer Merkmale auf der Betriebseite mit Kennzahlen aus der jeweiligen Personendatenbank (Jahrescheibe), die auf die Betriebsebene aggregiert werden. In den Originaldaten ist natürlich von einer wesentlich besseren Qualität auszugehen. **Mit den LIAB-Testdaten sind keine inhaltlichen Auswertungen möglich.**

6.) Die Größenstruktur der Betriebe ist nicht die gleiche wie in den Originaldaten (dem IAB-Betriebspanel). Dort sind Großbetriebe weitaus häufiger vertreten. Die Abbildung dieser Struktur ist nicht möglich, weil die Fallzahlen in Testdaten definitionsgemäß möglichst gering sein sollten und nur der Erstellung fehlerfrei laufender Auswertungsprogramme dienen.

7.) Die einzelnen Variablen sind vollständig belabelt, bei kategorialen Variablen wurden auch die gültigen Wertelabels vergeben, sofern es sich nicht um mehr als zehn Ausprägungen handelt. Bei mehr Ausprägungen werden manchmal für eine Auswahl an einzelnen Werten Labels vergeben.

8.) Weitere Informationen zu den Daten, den Variablen und ihrer inhaltlichen Bedeutung befinden sich im Internet (<http://fdz.iab.de/pageText.asp?PageID=18>)

9.3.2. Beschreibung der Testdaten des LIAB-Längsschnittmodell

0.) Die Testdaten bestehen aus zwei Einzelteilen, den Betriebsdaten des IAB-Betriebspanels und Personendatenbanken (Jahresscheiben). Die Verknüpfung beider Datensätze erfolgt über den systemfreien Betriebsidentifikator (idnum).

1.) Die Testdaten sind so konzipiert, dass sie sowohl die Struktur der Personendaten wie auch die Verknüpfungsqualität repräsentieren. Aufgrund der Komplexität der Originaldaten kann das nur ansatzweise und stilisiert geschehen.

2) Auf der Betriebsseite scheiden erst ab 1995 Betriebe aus dem Panel aus. Definitionsgemäß kommt es in dieser Version 1 des LIAB-Längsschnittmodells auch in den Originaldaten in den Jahren 1999 bis 2001 zu keinerlei Ausfällen auf der Betriebsseite (vgl. FDZ-Datenreport Nr. 3). Die größten Betriebe mit mindestens 250 Beschäftigten sind in allen Jahren 1993 bis 2001 vertreten.

2a) Auf der Personenseite sind weitaus häufiger als in den Originaldaten so genannte Recalls identifizierbar, also Wiederbeschäftigungen einer Person in dem gleichen Betrieb nach einer Unterbrechung. Es gibt Beschäftigte, die nach einer gewissen Zeit von einem IAB-Betriebspanelbetrieb in einen Betrieb wechseln, der nicht am IAB-Betriebspanel teilgenommen hat. Das ist auch in den Testdaten der Fall. Genauso ist der Wechsel in einen anderen IAB-Betriebspanelbetrieb möglich.

2b) Es gibt Personenkonten, die gar nicht mit den Betriebsdaten verknüpft werden können. In den Originaldaten sind das Beschäftigte, die zwar zwischen 1996 und 1998 in einem der für die Version 1 ausgewählten Betriebe gearbeitet haben, bei denen für die entsprechenden Betriebe in diesem Zeitraum aber keine gültigen Interviews vorliegen.

3.) Angestrebt werden Fallzahlen in der Personendatenbank von etwa 70.000 Meldungen. Aus Vereinfachungsgründen wurden nur Westkonten generiert.

4.) Auf der *Betriebseite* wurde für 2.) und 3.) folgender Größenmix verwendet:

- 1 Betrieb mit 1000 Beschäftigten
- 4 Betriebe mit 250 Beschäftigten
- 10 Betriebe mit 100 Beschäftigten
- 10 Betriebe mit 50 Beschäftigten
- 10 Betriebe mit 25 Beschäftigten
- 10 Betriebe mit 10 Beschäftigten
- 10 Betriebe mit 5 Beschäftigten

Mit Beschäftigten sind immer Sozialversicherungspflichtige gemeint. Zu den Stichtagen 30. Juni eines jeden Befragungsjahres des IAB-Betriebspanels wird diese Sollgröße auch in den Testdaten erreicht.

5.) Es gibt definitionsgemäß für diese Version 1 des LIAB-Längsschnittmodells maximal geringe Abweichungen (immer weniger als fünf Prozent) bei der Zahl der Beschäftigten laut IAB-Betriebspanel (z.B. a01svb93) und der Anzahl der sozialversicherungspflichtig Beschäftigten durch die Aufaggregation der Personendaten zum Stichtag 30. Juni auf die Betriebsebene. Nur bei der Betriebsgröße stimmen mit diesen kleinen eingebauten Ungenauigkeiten die Angaben in den Testdaten des IAB-Betriebspanels (z.B. a01svb93) und in der Personendatenbank für das jeweilige Jahr überein. Alle anderen Merkmale sind zufällig verteilt. Es gibt nahezu keine (sinnvollen) Entsprechungen anderer Merkmale auf der Betriebseite mit Kennzahlen aus der jeweiligen Personendatenbank (Jahresscheibe), die auf die Betriebsebene aggregiert werden. In den Originaldaten ist natürlich von einer wesentlich besseren Qualität auszugehen. **Mit den LIAB-Testdaten sind keine inhaltlichen Auswertungen möglich.**

5a.) Die Verlaufsinformationen für die Personen ergeben in der Regel einen sinnvollen Erwerbsverlauf, zumindest was Beginn- und Enddatum der Meldungen angeht. Die Meldungsinhalte allerdings sind wiederum mehr oder minder zufällig gefüllt.

6.) Die Größenstruktur der Betriebe ist nicht die gleiche wie in den Originaldaten (dem IAB-Betriebspanel). Dort sind Großbetriebe weitaus häufiger vertreten. Die Abbildung dieser Struktur ist nicht möglich, weil die Fallzahlen in Testdaten definitions-

gemäß möglichst gering sein sollten und nur der Erstellung fehlerfrei laufender Auswertungsprogramme dienen.

7.) Die einzelnen Variablen der Personendatenbank sind vollständig belabelt, bei kategorialen Variablen wurden auch die gültigen Wertelabels vergeben, sofern es sich nicht um mehr als zehn Ausprägungen handelt. Bei mehr Ausprägungen werden manchmal für eine Auswahl an einzelnen Werten Labels vergeben.

8.) Weitere Informationen zu den Daten, den Variablen und ihrer inhaltlichen Bedeutung befinden sich im Internet (<http://fdz.iab.de/pageText.asp?PageID=18>)

9.3.3. Der Paneldatensatz des IAB-Betriebspanels

Abschließend wird auf den Paneldatensatz des IAB-Betriebspanels hingewiesen. Er wird als Arbeitshilfe explizit auch für die LIAB-Daten zur Verfügung gestellt. Mit ihm ist die Aufbereitung der IAB-Betriebspaneldaten vergleichsweise einfach. Der Variablenkanon kann ohne großen Aufwand erweitert werden. Auch für diesen Datensatz gibt es Testdaten, die allerdings nicht ohne Weiteres mit den Testdaten der LIAB-Personendaten verknüpft werden können. Es wird die Verwendung des ebenfalls frei verfügbaren STATA-Do-Files zur Generierung des Paneldatensatzes empfohlen, der dann an den LIAB-Testdaten für das IAB-Betriebspanel zur Anwendung kommt. Dann haben die Identifikationsnummern für die einzelnen Betriebe des Paneldatensatzes auch garantiert eine Entsprechung in den Testdaten für die LIAB-Personendaten.

10. Abkürzungsverzeichnis

ALG	Arbeitslosengeld
ALH	Arbeitslosenhilfe
BA	Bundesagentur für Arbeit
FDZ	Forschungsdatenzentrum
BeH	Beschäftigten-Historik-Datei
BLH	Beschäftigten-Leistungsempfänger-Historik-Datei
DEVO	Datenerfassungsverordnung;
DEÜV	Datenerfassungs- und -übermittlungsverordnung
IAB	Institut für Arbeitsmarkt- und Berufsforschung
DÜVO	Datenübermittlungsverordnung
JZM	Jahreszeitraummaterial
LeH	Leistungsempfänger-Historik-Datei
LIAB	Linked-Employer-Employee-Daten des IAB
SGB	Sozialgesetzbuch
WZ	Wirtschaftszweigklassifikation

11. Literaturverzeichnis für das LIAB-Datenhandbuch

- Abowd, J./Kramarz, F. (1999): *The analysis of labor markets using matched employer-employee data*. In: Ashenfelter, O./Card, D. (Hrsg.): *Handbook of Labor Economics*, Vol. 3B, Elsevier, Amsterdam, chapter 40, pp. 2567 - 2627
- Alda, H. (2005): FDZ-Methodenreport Nr. 1: *Konsistenz der Linked-Employer-Employee-Daten des IAB, LIAB*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
- Alda, H. (2005a): FDZ-Datenreport Nr. 1: *Betriebe und Beschäftigte in den Linked-Employer-Employee-Daten des IAB, LIAB*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
- Alda, H. (2005b): FDZ-Datenreport Nr. 2: *Datenbeschreibung der ersten Version des LIAB-Querschnittmodells*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
- Alda, H. (2005c): FDZ-Datenreport Nr. 3: *Datenbeschreibung der ersten Version des LIAB-Längsschnittmodells*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
- Alda, H./Bender, St./Gartner, H. (2005): *Der Linked-Employer-Employee-Datensatz aus IAB-Betriebspanel und den prozessproduzierten Daten des IAB (LIAB)*. IAB-Discussion-Paper No.6/2005, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.
- Allmendinger/Kohlmann (2005): *Datenverfügbarkeit und Datenzugang am FDZ*, Allgemeines Statistisches Archiv, Ausgabe 2005 (im Erscheinen)
- Bellmann, L. (2002): *Das IAB-Betriebspanel - Konzeption und Anwendungsbereiche*, Allgemeines Statistisches Archiv 86, 177 – 188
- Blossfeld, H.P./Hamerle, A./Mayer, K.U. (1985): *Ereignisanalyse - statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*, Campus-Verlag, Frankfurt u.a.

Bundesagentur für Arbeit: *Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung*. Ausgabe Januar 2005, S. IX.

Osikominu, A./Fitzenberger, B./Völter, R. (2005): *Imputation Rules to Improve the Education variable in the IAB-Employment-Subsample*. Beitrag zur 1. Nutzerkonferenz zu den Daten der BA und des IAB 15. – 16. Juli 2005, Nürnberg

Hamann, S. (2004): *Die IAB-Regionalstichprobe 1975-2001 (originale Version)*, S. 4-6

Jensen, U./ Rässler, S. (2005): *Where have all the data gone? Stochastic production frontiers*, In: Institut für Arbeitsmarkt- und Berufsforschung (Hrsg.): 1. Nutzerkonferenz zu den Daten der BA und des IAB 15. – 16. Juli 2005, Sammlung der präsentierten Papers, Nürnberg

Kölling, A. (2000): *European Datawatch: The IAB-Establishment-Panel*, Schmollers Jahrbuch, 120 Jg., 291 - 300

Meinken, Holger; Koch, Iris: *BA-Beschäftigtenpanel 1998-2002. Codebuch*, Nürnberg, 2004

Servicebereich IT und Informationsmanagement (Autorengemeinschaft) (2004): *Handbuch zur Beschäftigten-Leistungsempfängerhistorikdatei des IAB*, Nürnberg

12. Literaturliste der LIAB-Anwendungen

- Achatz, J.; Gartner, H. & Glück, T. (2004): *Bonus oder Bias? Mechanismen geschlechtsspezifischer Entlohnung*, IAB-Discussion Paper No. 2
<http://doku.iab.de/discussionpapers/2004/dp0204.pdf>
- Achatz, J.; Allmendinger, J. & Hinz, T. (2001): *Sex Segregation in Organizations. A Comparison of Germany and the US*. Überarbeiteter Beitrag für die ASA-Tagung in Washington D.C., USA, August 2000.
- Andrews, M.; Schank, T. & Upward, R. (2004): *Practical Estimation Methods for Linked Employer-Employee Data*, IAB-Discussion Paper No. 3
<http://doku.iab.de/discussionpapers/2004/dp0304.pdf>
<http://doku.iab.de/discussionpapers/2004/Stata-files.zip>
- Bauer, T. (2003): *Flexible Workplace Practices and Labor Productivity*, IZA-Discussion Paper No. 700
<ftp://ftp.iza.org/dps/dp700.pdf>
- Bauer, T.K. & Bender, S. (2001): *Flexible Work Systems and the Structure of Wages: Evidence from Matched Employer-Employee Data*, IZA-Discussion Paper No. 353
<ftp://repec.iza.org/RePEc/Discussionpaper/dp353.pdf>
- Bauer, T.K. & Bender, S. (2002): *Technological Change, Organizational Change, and Job Turnover*, Labour Economics, 11: 265-291
<ftp://ftp.iza.org/dps/dp570.pdf>
- Bauer, T.K.; Bender, S. & Bonin, H. (2004): *Dismissal Protection and Worker Flows in Small Establishments*, IAB-Discussion Paper No. 1
<http://doku.iab.de/externe/2004/k040501f15.pdf>
<http://doku.iab.de/discussionpapers/2004/dp0104.pdf>
- Beblo, M.; Bender, S. & Wolf, E. (2004): *The Wage Effects of Entering Motherhood: A withinfirm matching approach*, mimeo Mannheim, Nürnberg.
- Bellmann, L.; Bender, S. & Kölling, A. (2002): *Der Linked Employer-Employee-Datensatz aus IAB-Betriebspanel und Beschäftigtenstatistik der Bundesanstalt für Arbeit (LIAB)*. In: G. Kleinhenz (Hrsg.) (2002): IAB-Kompendium Arbeits-

- markt- und Berufsforschung, Beiträge aus der Arbeitsmarktforschung 250, Nürnberg, 21-29, http://doku.iab.de/beitrab/2002/beitr250_103.pdf
- Bellmann, L. & Schank, T. (2000): *Innovations, Wages and Demand for Heterogeneous Labour: New evidence from a Matched Employer-Employee Data-Set*, IZA-Discussion Paper No. 112, <ftp://repec.iza.org/RePEc/Discussionpaper/dp112.pdf>
- Bellmann, L.; Bender, S. & Schank, T. (1999): *Flexibilität der Qualifikationsstruktur aus betrieblicher Sicht: Substitutionalität oder Komplementarität*. Jahrbücher für Nationalökonomie und Statistik 219: 109-126
- Bellmann, L.; Caliendo, M.; Hujer, R. & Radic, D. (2002): *Beschäftigungswirkungen technisch-organisatorischen Wandels - eine mikroökonometrische Analyse mit dem Linked IAB-Panel*. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 35: 506-522
- Bender, S.; Konietzka, D. & Sopp, P. (2000): *Diskontinuität im Erwerbsverlauf und betrieblicher Kontext*, Kölner Zeitschrift für Soziologie und Sozialpsychologie, 52: 475-499
- Bender, S.; Konietzka, D. & Sopp, P. (1999): *Spaltung und Exklusion auf dem Arbeitsmarkt? Die Entstehung von Arbeitsmarktsegmenten in Ostdeutschland und individuelle Erwerbschancen*. In: Schwengel, H. & Höpken, B. (Hrsg.): *Grenzenlose Gesellschaft? Band 2/1: Sektion, Forschungskomitees, Arbeitsgruppen*, Pfaffenweiler: Centaurus-Verlagsges., 493-396
- Brixy, U.; Kohaut, S. & Schnabel, C. (2004): *Do Newly Founded Firms Pay Lower Wages? First Evidence from Germany*. Erlangen u.a.: IAB-Discussion Paper No. 4 <http://doku.iab.de/discussionpapers/2004/dp0404.pdf>
- Bender, S.; Preißler, J. & Wübbecke, C. (2000): *Betriebliche Determinanten des Generationenaustausches in westdeutschen Betrieben. Eine Untersuchung auf der Basis des IAB-Employer-Employee-Datensatzes für die Jahre 1994 und 1995*. In: George, R. & Struck, O. (Hrsg.): *Generationenaustausch im Unternehmen*, München: Hampp, 209-225

- Gartner, H. & Stephan G. (2004): *How Collective Contracts and Works Councils Reduce the Gender Wage Gap*, eingereicht in: Economic Letters.
- Grotheer, M.; Struck, O.; Bellmann, L. & Gewiese, T. (2004): *Determinanten der Beschäftigungsstabilität. Chancen und Risiken von ‚Entrants‘ im ost-westdeutschen Vergleich. In: Beschäftigungsstabilität im Wandel?: empirische Befunde und theoretische Erklärungen für West- und Ostdeutschland*, München u.a.: Hampp, 125-156
- Hujer, R.; Caliendo, M. & Radic, D. (2002): *Skill Biased Technological and Organizational Change. Estimating a mixed simultaneous equation model using the IAB establishment panel*. IZA-Discussion Paper No. 566
<ftp://ftp.iza.org/dps/dp566.pdf>
- Kölling, A. (2002): *Fachkräftebedarf als betriebliches Matching-Problem*, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 35: 566-579
- Kölling, A. & Schank, T. (2002): *Skill Biased Technological Change, International Trade and the Wage Structure*, Diskussionspapiere Nr. 14
<http://www.arbeitsmarkt.wiso.uni-erlangen.de/pdf/Diskussionspapiere/dp14.pdf>
- Kölling, A. & Schank, T. (2003): *Skill Based Technological Change, International Trade and the Wage Structure. New Evidence on the Determinants of the Employment Structure from Linked Employer-Employee Panel Data for Germany*, Beiträge zur Arbeitsmarkt- und Berufsforschung No. 277, Nürnberg, 113-134
- Kölling, A.; Schnabel, C. & Wagner, J. (2002): *Establishment Age and Wages. Evidence from German Linked Employer-Employee*, IZA-Discussion Paper No. 679
<ftp://ftp.iza.org/dps/dp679.pdf>
- Kölling, A.; Schnabel, C. & Wagner, J. (2002): *Establishment Age and Wages. Evidence from German Linked Employer-Employee*, Diskussionspapiere Nr. 13, Universität Erlangen, Nürnberg, Lehrstuhl für Arbeitsmarkt- und Regionalpolitik, Erlangen u.a.
<http://www.arbeitsmarkt.wiso.uni-erlangen.de/pdf/Diskussionspapiere/dp13.pdf>
- Schank, T. (2003): *Die Beschäftigung von Un- und Angelernten. Eine Analyse mit dem Linked Employer-Employee Datensatz des IAB*, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 36: 257-270

Schank, T.; Schnabel, C. & Wagner, J. (2004): *Exporting Firms do not Higher Wages, Ceteris Paribus. First Evidence from Linked Employer-Employee Data.* IZA-Discussion Paper No. 1185,

<http://doku.iab.de/externe/2004/k040702f39.pdf>

<http://www.arbeitsmarkt.wiso.uni-erlangen.de/pdf/Diskussionspapiere/dp27.pdf>

Schwerdt, W. & Bender, S. (2003): *Was tun Lehrlinge nach ihrer Ausbildung? Eine Analyse mit dem Linked Employee-Datensatz des IAB.* Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 36: 46-59

Anhang

Das LIAB-Codebuch mit den Auszählungen der Variablen der LIAB-Personendatenbanken ist aufgrund seines Umfangs in einem eigenem Anlageband dokumentiert, ebenso wie die LIAB-Codepläne, in der alle Variablen und ihre Ausprägungen aufgelistet sind. Der nachfolgende Anhang A1 in diesem Datenhandbuch stellt die Programmcodes aus den einzelnen Kapiteln und Abschnitten zusammen. Im Anhang A2 befindet sich ein Klassifizierungsvorschlag für die Berufskennziffern.

Anhang A 1: Programmcodes³⁵

Ziffer	STATA-Code
①	1.) bysort id: egen help=mean(geschl) 2.) gen help1=int(help) 3.) replace geschl=help1 if help~=geschl 4.) drop help*
②	1.) tab pers_gr if berstell==0
③	1.) local y=1 2.) while `y' < 4 { 3.) sort id spell_nr 4.) cap drop help 5.) gen help=1 if quelle= =2 & beginn[_n-1]= =beginn & ende[_n-1]= =ende & quelle[_n-1]= = 1 6.) drop if help= =1 7.) sort id spell_nr 8.) drop spell_nr 9.) by id: gen spell_nr=_n 10.) local y=`y'+1 11.) } Die Schleife läuft insgesamt dreimal, damit ganz sicher auch drei(vier)fache Doppelmeldungen gelöscht werden. Anschließend werden die durch die Löschung von Meldungen entstehenden Lücken in der Variable spell_nr korrigiert. Da die Beschäftigungsmeldung auch hinter der Leistungsmeldung stehen kann, muss die Schleife auch in dieser Richtung durchgeführt werden.

(Fortsetzung nächste Seite)

³⁵ Anmerkung: Die Nummerierungszeichen 1.), 2.) usw. werden in realen Programmcodes weggelassen, sonst laufen die Programme nicht fehlerfrei. Sie dienen hier zur Absetzung und Identifizierung einzelner Programmzeilen.

Ziffer	STATA-Code
③	<pre> local y=1 while `y' < 5 { sort id spell_nr cap drop help cap drop help1 gen help=1 if quelle= =2 & beginn[_n+1]= =beginn & ende[_n+1]= =ende & quelle[_n+1]= = 1 drop if help= =1 sort id spell_nr drop spell_nr by id: gen spell_nr=_n local y=`y'+1 } </pre>
④	<ol style="list-style-type: none"> 1.) gen vollzeit=1 if berstell= =1 berstell= =2 berstell= =3 berstell= =4 2.) replace vollzeit=. if pers_gr>=102 & pers_gr<=106 3.) replace vollzeit=. if pers_gr==109 pers_gr==209 4.) gen monatslohn=tag_entg*30.5 5.) replace monatslohn=. if monatslohn= = 0 6.) sum monatslohn if vollzeit= =1 & beginn>=14245³⁶, detail
⑤	<ol style="list-style-type: none"> 1.) use "X:\Nutzer\personendaten_liabquer2002.dta", clear 2.) cap drop dauer 3.) gen dauer=ende-beginn+1 4.) sum dauer if abg_gr==50,d
⑥	<ol style="list-style-type: none"> 1.) bysort idnum: egen help=mean(west) 2.) bysort idnum: keep if _n= =1 3.) bysort wo2002: sum help, d
⑦	<ol style="list-style-type: none"> 1.) cd "X:\Nutzer\" 2.) use test2002.dta 3.) keep idnum wo2002 4.) sort idnum 5.) save ost_west2002.dta, replace 6.) use personendaten_liabquer2002.dta", clear 7.) sort idnum 8.) merge idnum using ost_west2002.dta 9.) tab _merge 10.) keep if _merge==3 11.) replace tag_entg=147.95 if wo2002==1 & tag_entg>147.95 & tag_entg~=. 12.) replace tag_entg=123.29 if wo2002==2 & tag_entg>123.29 & tag_entg~=. 13.) drop wo2002 _merge 14.) cap drop help* 15.) gen help=tag_entg*30.5

(Fortsetzung nächste Seite)

³⁶ Das Beginndatum größer gleich 14245 (entspricht dem ersten Januar 1999) beschränkt die Lohnanalyse auf Euro-Beträge.

Ziffer	STATA-Code
7	16.) bysort idnum: egen lohnsumme_liab=sum(help) 17.) drop help 18.) keep idnum lohnsumme_liab 19.) bysort idnum: keep if _n==1 20.) sort idnum 21.) merge idnum using "X:\Nutzer\test2002.dta" 22.) tab _merge 23.) keep if _merge==3 24.) drop _merge 25.) mvdecode j34, mv(-8,-9) 26.) gen differenz=j34-lohnsumme_liab 27.) sum differenz, detail
8	1.) use "X:\Nutzer\personendaten_liabquer2002.dta", clear 2.) cap drop help* 3.) gen help=1 if berstell==0 4.) bysort idnum: egen anzahl_azubis=sum(help) 5.) gen help2=1 6.) bysort idnum: egen insgesamt=sum(help2) 7.) gen azubi_anteil=anzahl_azubis/insgesamt*100 8.) bysort idnum: keep if _n==1 9.) keep azubi_anteil idnum 10.) sort idnum 11.) merge idnum using "X:\Nutzer\test2002.dta" 12.) mvdecode j41aus j01svb02, mv(-8,-9) 13.) drop if _merge==2 14.) drop _merge 15.) gen azubi_panel=j41aus/j01svb02*100 16.) keep if azubi_panel~=. & azubi_anteil~=. 17.) sum azubi*,d

Anhang A 2: Berufsgruppen nach einem Vorschlag von Blofeld/Hamerle/Mayer (1985)

Bezeichnung der Berufsgruppe	Beschreibung der Berufsgruppe	Berufskennziffern nach der deutschen Klassifizierung der Berufe (1970)
Produktion		
Agrarberufe (AGR)	Dominant landwirtschaftlich orientierte Berufe	011-022,041-051,053-062
Einfache manuelle Berufe (EMB)	Alle manuellen Berufe, die 1970 mindestens einen 60 % en Anteil an Ungelernten aufwiesen	071-133,135-141,143,151-162,164,176-193,203-213,222-244,252,263,301,313,321-323,332-346,352-371,373,375-377,402,403,412,423-433,442,452-463,465-472,482,486,504,512-531,543-549,971,981-983
Qualifizierte manuelle Berufe (QMB)	Alle manuellen Berufe, die 1970 höchstens einen 40 % en Anteil von Ungelernten aufwiesen	134,142,144,163,171-175,201,202,221,251,261,262,270-291,302,305-312,314,315,331,351,372,374,378-401,411,421,422,441,451,464,481,483-485,491-503,511,541,542
Techniker (TEC) Ingenieure (ING)	Alle technischen Fachkräfte	303,304,621-635,721,722,733,857,
	Hochqualifizierte Fachkräfte zur Lösung naturwissenschaftlicher und technischer Probleme	032,052,601-612,726,883
Dienstleistung		
Einfache Dienste	Alle einfachen persönlichen Dienste	685,686,688,706,713-716,723-725,741-744,791-794,805,838,911-913,923-937
Qualifizierte Dienste	Im wesentlichen Ordnungs- und Sicherheitsberufe, sowie qualifizierte Dienstleistungsberufe	684,704,705,711,712,801-804,812,814,851,852,854-857,892-902, 921-922
Semiprofessionen (SEMI)	Dienstleistungsberufe, die sich durch eine Verwissenschaftlichung der Berufspositionen auszeichnen	821-823,853,861-864,873-877
Professionen (PROF)	Freie Berufe und hochqualifizierte Dienstleistungsberufe	811,813,831-837,841-844,871,872,881,882,891
Einfache kaufmännische und Verwaltungsberufe (EVB)	Relativ unqualifizierte Büro- und Handelsberufe	682,687,731,732,734,773,782-784
Qualifizierte kaufmännische und Verwaltungsberufe (QVB)	Berufe mit mittleren und höheren verwaltenden und distributiven Funktionen	031,681,683,691-703,771,772,774-781
Manager (MAN)	Berufe, die die Kontrolle und Entscheidungen über den Einsatz von Produktionsfaktoren besitzen, sowie Funktionäre in Organisationen	751-763

Imprint

FDZ *Datenreport*

No. 7/2005

Publisher

The Research Data Centre (FDZ)
of the Federal Employment Service
in the Institute for Employment Research
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff

Stefan Bender, Dagmar Herrlinger

Technical production

Dagmar Herrlinger

Copyright

Reproduction – also in parts – only with permission of
the FDZ

Download

http://doku.iab.de/fdz/reporte/2005/DR_7.pdf

Internet

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

Corresponding author

Dr. Holger Alda, Tel.: +49 (0) 911/179-3138

E-Mail: holger.alda@iab.de