

Zensus 2011: Berufskodierung und Auswertung der Berufsstrukturen

Im Rahmen des Zensus 2011 – der europaweiten Volks-, Gebäude- und Wohnungszählung 2011 – wurden als Teil der stichprobenartigen Haushaltebefragung¹⁾ die Auskunftspflichtigen der Haushaltsstichprobe auch zu ihrer am 09. Mai 2011 ausgeübten beruflichen Tätigkeit befragt.

Innerhalb des großen Auswertungskomplexes der Zensusdaten handelt es sich bei dem Teilgebiet der Berufskodierung explizit um die Erfassung der im Fragebogen der Haushaltebefragungen²⁾ als Klartext abgefragten Berufsangaben in einem webbasierten, menügesteuerten Programm. Ziel war die Erfassung des Merkmals Beruf durch Kodierung der Merkmalsausprägung in einen 5-stelligen Code anhand der im Programm hinterlegten nationalen Klassifikationstabelle der Berufe 2010 (KldB 2010)³⁾, welche eine effektive Auswertung der Arbeitsmarktstruktur in Deutschland ermöglichte. Im Programm wurde der 5-stellige KldB-Code automatisiert auch in den internationalen 1-stelligen ISCO 08-Kode⁴⁾ umgeschlüsselt, welcher wiederum die Erfüllung der EU-Pflichtangabe darstellt.

Die Vielfalt der im Erhebungsbogen frei formulierbaren Berufsangaben erschwerte deren Auswertung. Dieses Problem ließ sich aber durch die Klassifizierung der Angaben nach Tätigkeiten und Anforderungsniveaus lösen.

Die Kodierung der Berufe erfolgte sowohl maschinell als auch manuell. Im Vorfeld der manuellen Kodierung wurden bei IT.NRW bereits diejenigen Fragebogenangaben zur Merkmalsausprägung „Beruf“ maschinell kodiert, die einen eindeutigen oder sehr hohen Ähnlichkeitsabgleich (ab 92 %) mit dem hinterlegten KldB-Code aufwiesen. Konnte der Beruf jedoch im maschinellen Kodierungsverfahren nicht eindeutig zugeordnet werden, wurden diese Fälle in das manuelle Kodierverfahren überführt und durch Mitarbeiterinnen und Mitarbeiter des LSKN im zugehörigen IT-Fachverfahren bearbeitet.

Im Folgenden wird zunächst der Ablauf der Erhebung dargestellt; es folgen erste Ergebnisse.

1) Vgl. Loth, A., Lange, J., Rohrschneider, L.: Zensus 2011 – Ziehung der Stichprobe und Durchführung der Haushaltebefragungen, in: Statistische Monatshefte Niedersachsen, Heft 2 (2011), S. 86-89.

2) Haushaltsstichprobe (Frage 45 und 46) und Erhebung an Sonderanschriften mit Zusatzmerkmalen (Frage 49 und 50).

3) Klassifikation der Berufe 2010 (KldB 2010). Die Klassifikation der Berufe 2010 kann bei der BA heruntergeladen werden.

4) Die Berufssystematik International Standard Classification of Occupations ist ein von der Internationalen Arbeitsorganisation (ILO) zusammengestelltes, international gültiges Klassifikationsschema für Gruppen von Berufen.

Durchführung der Berufskodierung

Von August bis Dezember 2012 arbeitete im LSKN ein Team aus Kodierkräften und Administratoren zur Qualitätssicherung an der manuellen Kodierung der Berufe⁵⁾. Die Administratoren führten nach Abschluss der manuellen Kodierung die Qualitätssicherung fort und schlossen den Aufgabenbereich der Berufskodierung im Dezember 2012 mit der finalen Freigabe der Datensätze termingerecht ab.

Die für die manuelle Berufskodierung insgesamt eingespielten rund 320 000 Datensätze verteilten sich aufgrund einer vorgegebenen Einteilung im IT-Fachverfahren auf insgesamt 10 Berufsbereiche (siehe Abb. 1), welche ihrerseits auf 25 Wirtschaftszweige aufgeteilt waren.

Vor Beginn der Berufskodierung und auch noch parallel zu ihr fand eine Plausibilisierung aller Datensätze der Haushaltebefragungen statt, hier wurden jedoch die Fragen nach der beruflichen Tätigkeit noch ausgelassen. Durch dieses Plausibilisierungsmodell lagen der Berufskodierung zu Beginn ausschließlich unplausibilisierte Datensätze zugrunde.

Alle Datensätze mit unbefüllten Angaben zum Beruf wurden zunächst in einer Korrekturdatenbank abgelegt. Nach Abschluss der Berufskodierung durchliefen alle kodierten Datensätze einen Plausibilisierungslauf, bei dem geprüft wurde, ob diese Datensätze überhaupt eine Berufsangabe enthalten durften. Beispielsweise mussten unter 15-Jährige keine Angabe zum Beruf beantworten. Die Imputation fehlender Berufsangaben bildete nach dem Plausibilisierungsverfahren den Abschluss der Bearbeitung im Aufbereitungsverfahren der Berufskodierung. Sowohl Plausibilisierung als auch Imputation wurden rein maschinell beim Statistischen Landesamt Nordrhein-Westfalen durchgeführt.

Häufigkeitsverteilung der kodierten Berufe auf die Berufsbereiche⁶⁾

Bei der Verteilung der in Niedersachsen kodierten Berufsangaben auf die 10 Berufsbereiche ist eine deutliche Zweiteilung erkennbar, bei der sich die Mehrheit der Berufe auf 5 große Berufsbereiche konzentriert (siehe Abb. 1).

5) Ein Anteil von 9,6 % der getätigten Berufsangaben wurde maschinell kodiert.

6) Alle in diesem Bericht verwendeten Werte basieren auf Auswertungen des Statistischen Bundesamtes von Anfang Februar 2013.

Im Bereich der *Rohstoffgewinnung, Produktion und Fertigung* sind mit 19,8 % die meisten der kodierten Berufsangaben angesiedelt, dicht gefolgt von den Bereichen *Unternehmensorganisation, Buchhaltung, Recht und Verwaltung* (17,9 %) und *Gesundheit, Soziales, Lehre und Erziehung* (17,1 %). Zahlenmäßig ebenfalls stark vertreten sind die Bereiche *Verkehr, Logistik, Schutz und Sicherheit* (15,7 %) und *Kaufmännische Dienstleistungen, Warenhandel, Vertrieb, Hotel und Tourismus* (14,4 %).

Zu den vergleichsweise gering vertretenen Berufsbereichen gehören die Bereiche *Bau, Architektur, Vermessung und Gebäudetechnik* (6,4 %), *Land-, Forst- und Tierwirtschaft und Gartenbau* (3,4 %), *Naturwissenschaft, Geografie und Informatik* (2,4 %) und *Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung* (2,1 %). Die wenigsten Berufsangaben wurden im Bereich *Militär* kodiert (0,8 %).

Häufigkeitsverteilung der kodierten Berufe auf die Anforderungsniveaus

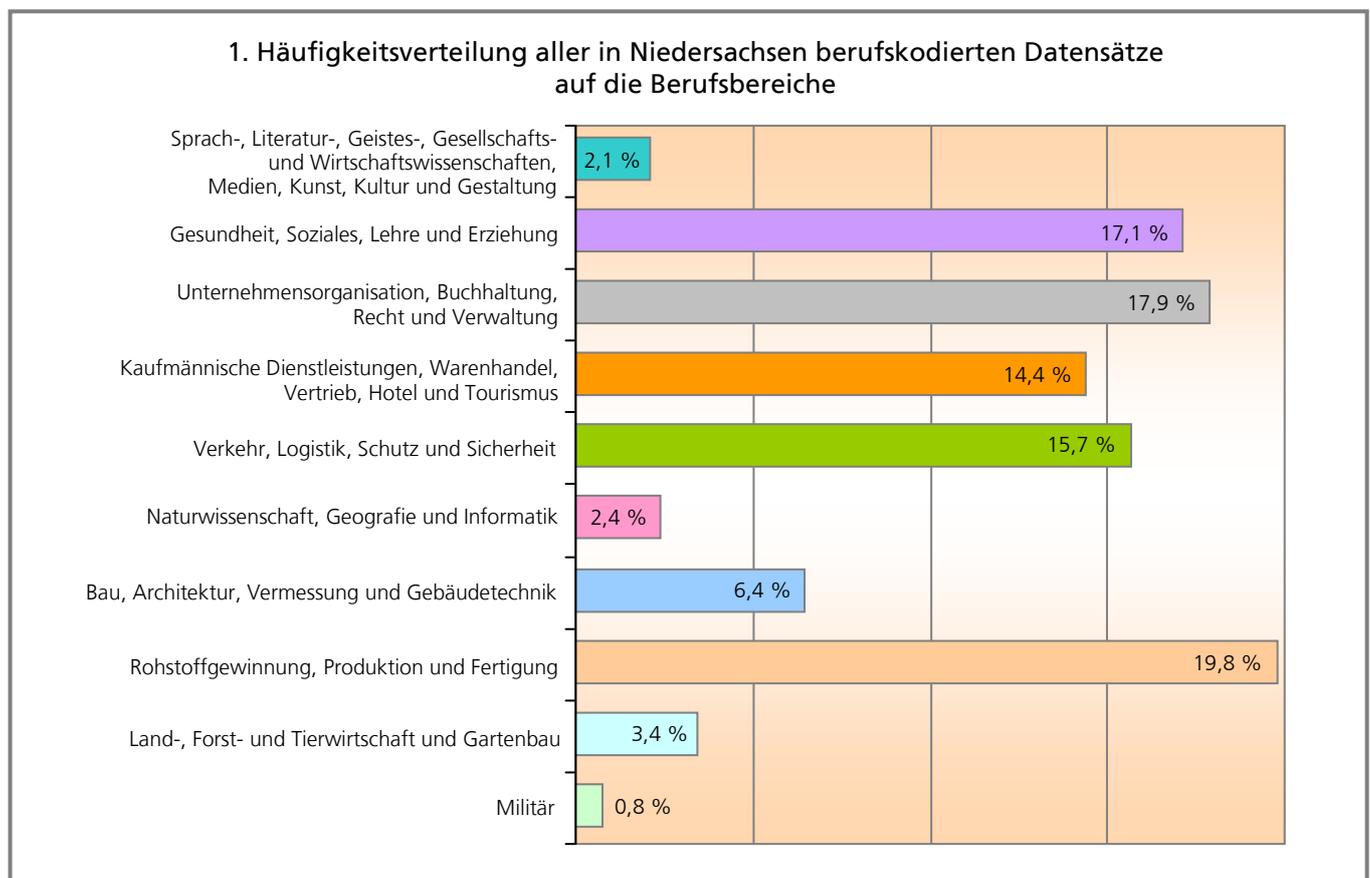
Neben dem Berufsbereich ist in der Kodierung auch das Anforderungsniveau verschlüsselt worden. Anhand der letzten Stelle des KldB 2010-Kodes lässt sich das jeweilige

Anforderungsniveau des kodierten Berufs ablesen. Es wurden vier Niveaus gebildet:

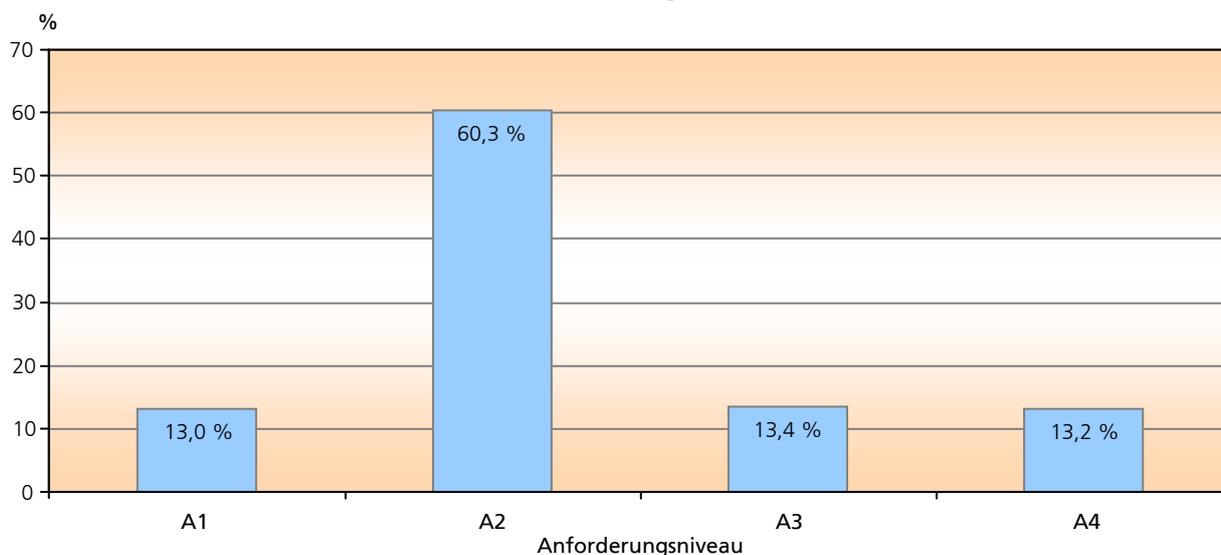
- **Anforderungsniveau 1**
(Endung mit xxxx1) = Helfer- und Anlernertätigkeiten,
- **Anforderungsniveau 2**
(Endung mit xxxx2) = fachlich ausgerichtete Tätigkeiten/Fachkraft,
- **Anforderungsniveau 3**
(Endung mit xxxx3) = komplexe Spezialistentätigkeiten,
- **Anforderungsniveau 4**
(Endung mit xxxx4) = hoch komplexe Tätigkeiten/Experte.

In Niedersachsen wurde mit großem Abstand die höchste Anzahl der Berufe mit dem **Anforderungsniveau 2** kodiert, was der Tätigkeit einer ausgebildeten Fachkraft entspricht. Die **Anforderungsniveaus 1, 3 und 4** unterscheiden sich in ihrer Anzahl nur unwesentlich voneinander (siehe Abb. 2).

Wenn auch die Verteilung des Anforderungsniveaus innerhalb der einzelnen Berufsbereiche durchaus variiert, so wurde über alle Berufsbereiche hinweg in Niedersachsen insgesamt mit durchschnittlich 60,3 % der größte Anteil der kodierten Berufe dem **Anforderungsniveau 2**



2. Prozentuale Häufigkeitsverteilung aller in Niedersachsen berufskodierten Datensätze nach Anforderungsniveau (KldB 2010)



zugeordnet. Einzig im Berufsbereich „Naturwissenschaft, Geografie und Informatik“ unterscheidet sich die Verteilung vom niedersächsischen Durchschnitt. In diesem Bereich wurde der größte Anteil der kodierten Berufsangaben mit 35,5 % den komplexen Spezialistentätigkeiten im **Anforderungsniveau 3** zugeordnet, dicht gefolgt von den Expertentätigkeiten im **Anforderungsniveau 4** mit 32,4 %.

Alles in allem verlief die manuelle Berufskodierung ohne größere Zuordnungsprobleme. Durch die Kodierung der Berufe im Rahmen des Zensus 2011 wird es möglich werden, differenzierte Aussagen über eine strukturelle Berufsverteilung sowie über die Verteilung der beruflichen Anforderungsniveaus in Niedersachsen zu tätigen. Endgültige Ergebnisse dazu werden im Frühjahr 2014 vorliegen.