



## Arbeiten mit den Testdaten des BA-Beschäftigtenpanels

Die Testdaten wurden durch Ziehung einer Substichprobe, durch Mikroaggregation und Verfremdung der Originaldaten erzeugt. Die Verteilungen mancher Merkmale wurden simuliert, ferner sind einige Personen erfunden worden.

Die Testdaten sollen es Ihnen ermöglichen, Auswertungsprogramme für Analysen mit dem BA-Beschäftigtenpanel zu schreiben und zu testen. **Sinnvolle inhaltliche Analysen können Sie mit den Testdaten nicht vornehmen.**

### Überblick

1. Für das BA-Beschäftigtenpanel und somit für die Testdaten gilt prinzipiell, dass pro Person (nur) ein Beschäftigungsverhältnis aufgenommen wird. Dies ist in der Regel eine sozialversicherungspflichtige Beschäftigung, ab dem 2. Quartal 1999 (Welle 6) sind aber auch ausschließlich geringfügig Beschäftigte erfasst. Ab dem 2. Quartal 2003 (Welle 22) ist das o. g. Prinzip dadurch durchbrochen, dass auch ein zusätzliches, geringfügiges Beschäftigungsverhältnis (= Nebenjob) aufgenommen wird – sofern vorhanden.
2. Die Daten zu den Nebenjobs liegen in getrennten Dateien mit der Endung „.nj“ vor. Alle Dateien sind nach der Personen-ID (v3\_s) sortiert. In den Beispielprogrammen (<http://fdz.iab.de>) finden Sie eine Vorgehensweise, wie Sie Hauptbeschäftigungsverhältnisse mit Hilfe der Personen-ID mit den dazugehörigen Nebenjobs verknüpfen können.
3. Grundlegende Informationen zu den Variablen der Testdaten können Sie dem Codebuch des BA-Beschäftigtenpanels, dem FDZ-Datenreport 1/2009 entnehmen. Die Besonderheiten der Testdaten sind im Folgenden aufgeführt.

### Verteilungen der Testdatenvariablen

4. Die Testdaten wurden so konstruiert, dass die Randverteilungen der Merkmale möglichst gut erhalten geblieben sind, bei den folgenden Variablengruppen sogar die gemeinsamen Verteilungen:
  - v4, v19, v55 (Meldegrund, Einkommen, Gleitzoneindikator)
  - v8, v9, v10 (Beruf, Ausbildung, Berufsstellung)
  - v26 – v36 (Anteilswerte bzgl. Altersgruppen)
  - v38 – v43 (Anteilswerte bzgl. Berufsstellung)
  - v44 – v49 (Anteilswerte bzgl. Ausbildung)
5. Bei den Variablen v8 (Beruf), v13 (Staatsangehörigkeit), v22 (WZ 93), v57 (WZ 03) ist sichergestellt, dass in jeder Testdatenwelle jede in der entsprechenden Panelwelle vorhandene Ausprägung mindestens einmal auftritt. (Dadurch sind allerdings die Randverteilungen dieser Variablen verzerrt.) Bei den anderen Variablen können Ausprägungen, die im Originalpanel vorhanden sind, im Testpanel fehlen. Eine Testdatenvariable in Welle  $n$  nimmt aber nur solche Werte an, die auch in der Originalpanelwelle  $n$  vorkommen.

## Zeitliche Struktur

6. Die Struktur von Erwerbsverläufen – Status „Beschäftigung“, „Arbeitslosigkeit“ oder „keine Information“ an den Stichtagen – ist in den Testdaten sichtbar. Aufgrund der Anonymisierung sind die Testdaten aber nicht im Zeitablauf konsistent. Beispielsweise kann eine Person von Welle zu Welle Geschlecht, Staatsangehörigkeit, Ausbildung etc. wechseln.

## Sensible Merkmale

7. Die Testdaten enthalten die datenschutzrechtlich besonders sensiblen Merkmale sowohl vergrößert als auch in ihrer Originalform. Bei einem Gastaufenthalt/einer Fernrechnung stehen diese Merkmale nur in vergrößerter Form zur Verfügung, es sei denn, die Nutzung von nichtvergrößerten Variablen wurde gesondert beantragt.

## Besonderheiten einzelner Testdaten-Variablen

8. Die Variablen v12 (Alter) und v19 (Entgelt) erscheinen geglättet, das heißt in den Testdaten fehlen die extremen Werte des „echten“ Panels.

## Fehlende Werte

9. Das BA-Beschäftigtenpanel unterscheidet drei verschiedene *missing values* und verwendet drei unterschiedliche Labels: „keine Zuordnung möglich“ (Wert in Stata: .x, in SPSS: -9); „keine Angabe“ (.z, -7); „nicht zutreffend“ (.n, -6). Diese müssen für Auswertungen in SPSS als *missing values* definiert werden, bspw. durch den Befehl: `mis val all (-6,-7,-9)`. In den Variablen v25 – v53 finden sich anstelle dieser Missings nur Systemmissings („.“), in der SPSS-Version gilt dies zusätzlich auch für die Variablen v2, v5 und v6.

## Hinweise zum Fernrechnen:

10. Beachten Sie bitte beim Schreiben von Auswertungsprogrammen *die Vorgaben des FDZ der BA im IAB zur Nutzung von Datenfernverarbeitung und Gastaufenthalt* ([http://doku.iab.de/fdz/access/Vorgaben\\_DAFE.PDF](http://doku.iab.de/fdz/access/Vorgaben_DAFE.PDF)).
11. Fehlerhafte Auswertungsprogramme, die nicht mit den Testdaten kontrolliert wurden, werden zur Überarbeitung ohne Ergebnisübermittlung zurückgeschickt.