

Working with Test Data from the BA Employment Panel

The test data were generated by means of sampling, micro aggregation and anonymization of the original data. Certain variable distributions were simulated; furthermore, a number of individuals were also invented.

Test data are intended to enable you to write and test evaluation programs. **They are not eligible for analysis.**

Overview

1. Basically, the BA Employment Panel, and thus the test data, follow the principle that an individual enters into only one employment relationship. (If an individual has more than one job, a “main job” is identified and included in the dataset.) Normally, this specific employment is subject to social security. Beginning with wave 6 in the 2nd Quarter of 1999, however, persons who have worked in minor jobs below the level of liability for social security contributions have also been registered. As of the 2nd Quarter of 2003 (wave 22) the one-job principle was softened: the dataset was extended by side jobs (= minor employment taken up in addition).
2. Data on side jobs are available in separate files (carrying the suffix "nj"). The files are sorted by person ID (v3_s). How to merge corresponding observations from the main dataset with those from the side job dataset is shown in the example programs (<http://fdz.iab.de>).
3. Basic information on the variables contained in the test data is provided in the codebook of the BA Employment Panel, the FDZ Datenreport No. 1/2009. Particularities of the test data are described in the following.

Distributions of Test Data Variables

4. As far as possible, the marginal distributions of variables have been maintained; even common distributions have been kept for the following variable groups:
 - v4, v19, v55 (reason of notification, wage, transition zone indicator)
 - v8, v9, v10 (profession, vocational qualification, occupational status)
 - v26 - v36 (shares by age group)
 - v38 - v43 (shares by occupational status)
 - v44 - v49 (shares by vocational qualification)
5. For variables v8 (profession), v13 (nationality), v22 (industry 93) and v57 (industry 03) we ensured that each variable value that occurs in a given panel wave has also been included in the respective test data wave at least once. This unfortunately means that the marginal distributions of these variables are distorted. For all other variables, some values included in the original panel might be missing in the test panel. However, a test data variable in wave n will only adopt a value that is also included in the original panel wave n .

Employment Biography Structure

6. The structure of employment biographies - i.e. status "employed", "unemployed" or "no information" applicable on the reference date - is visible from the test data. Since the test data are anonymized, they are not consistent over time. Therefore, it might happen that, for example, the gender, nationality or qualification of an individual changes from wave to wave.

Sensitive Variables

7. The test data contain variables in their original state which are particularly sensitive with respect to compliance with the current data protection legislation, i.e. they are not aggregated. In the real data, these variables are available only in aggregated form, unless access to non-aggregated variables has been requested for separately.

Particular Features of Selected Test Data Variables

8. Variables v12 (age) and v19 (wage) are smoothed, i.e. the test data do not reflect extreme values of the original panel.

Missing Values

9. The BA Employment Panel differentiates between three missing values specified by three different labels: "mapping impossible" (value in Stata: .x, value in SPSS: -9); "not stated" (.z, -7); "not applicable" (.n, -6). These must be defined as missing values for analyses conducted under SPSS, e.g. using the command "mis val all" (-6,-7,-9). For variables v25 - v53, these missings are not explicitly labelled, but marked as system missings ("."), which also applies to variables v5 and v6 under SPSS.

Instructions for Remote Data Access:

10. Please follow the *Guidelines for On-Site Use and Remote Data Access at the FDZ of the BA at the IAB* when developing evaluation programs (http://doku.iab.de/fdz/access/Vorgaben_DAFE_EN.PDF).
11. Non-workable evaluation programs that appear not to have been tested in advance using the available test data will not be considered and will be returned for correction.