



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR TECHNICAL REPORT 2012-001
APRIL 2012

Reshaping of Human Fertility Database data from long to wide format in Excel

Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)
Dmitry A. Jdanov (jdanov@demogr.mpg.de)

For additional material see www.demogr.mpg.de/tr/

© Copyright is held by the authors.

Technical reports of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in technical reports are attributable to the authors and do not necessarily reflect those of the Institute.

Reshaping of Human Fertility Database data from long to wide format in Excel

by Vladimir M. Shkolnikov and Dmitry A. Jdanov

Abstract

In the Human Fertility Database (HFD, 2012), age-specific data are presented in the long format, with both the years (or the years of the mothers' births) and the mothers' ages at birth in descending order, and with the age-specific fertility rates (or birth counts) corresponding to various combinations of years and ages placed in one long column. For constructing figures and performing other actions in MS Excel, it is useful to have the same data in wide format with the years (or the years of the mothers' births) descending rightwards, and with age-specific fertility rates for these years placed in separate columns arranged next to each other. We provide a simple VBA/Excel program that reshapes long-format age-specific HFD data into a wide format.

Keywords: age-specific fertility rates, long format, wide format, reshape

Background

For each country, the Human Fertility Database (HFD) provides on the page entitled "Age-Specific Data" cohort and period data on birth counts, female population exposure, and age-specific and cumulative fertility rates (Figure 1). These data are most useful for users who perform their own analyses of fertility trends and age patterns.

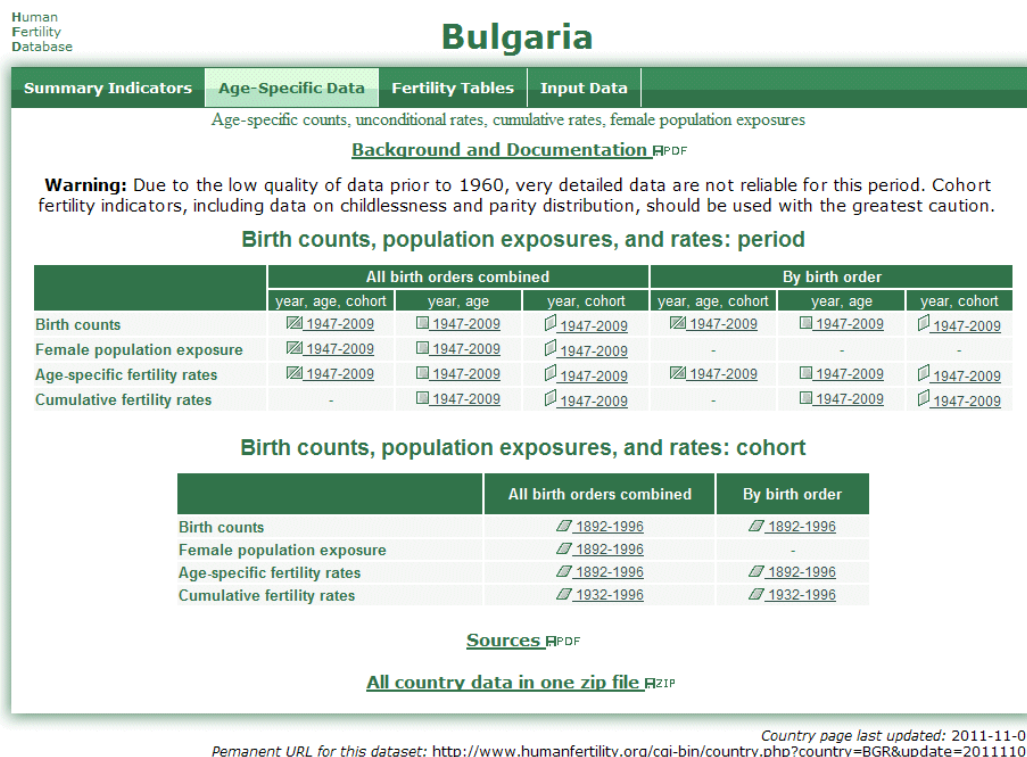


Figure 1. The HFD "Age-Specific Data" page for Bulgaria.

All of the datasets available on the Age-Specific Data page have the same long format. In them, the data elements (e.g., age-specific rates or birth counts) for all of the calendar years or years of the mothers' births are placed in one column. As an example, Figure 2 shows fragment of long-format

ASCII file with period fertility rates by calendar year and the age of the mothers for Bulgaria.

Year	Age	ASFR
1947	12-	0.00000
1947	13	0.00009
1947	14	0.00148
1947	15	0.00579
1947	16	0.01747
1947	17	0.03915
1947	18	0.06974
1947	19	0.10710
1947	20	0.14818
1947	21	0.17945
1947	22	0.19836
1947	23	0.20687
1947	24	0.21029
1947	25	0.19504
1947	26	0.17450
1947	27	0.15670
1947	28	0.14374
1947	29	0.13567
1947	30	0.12946
1947	31	0.11090
1947	32	0.09723
1947	33	0.08711
1947	34	0.07915
1947	35	0.06818
1947	36	0.05814
1947	37	0.04814
1947	38	0.03884
1947	39	0.03064
1947	40	0.02276
1947	41	0.01720
1947	42	0.01374
1947	43	0.01129
1947	44	0.00922
1947	45	0.00703
1947	46	0.00515
1947	47	0.00277
1947	48	0.00123
1947	49	0.00048
1947	50	0.00018
1947	51	0.00007
1947	52	0.00003
1947	53	0.00001
1947	54	0.00000
1947	55+	0.00000
1948	12-	0.00000
1948	13	0.00010
1948	14	0.00155
1948	15	0.00625
1948	16	0.01852
1948	17	0.04084
1948	18	0.07210
1948	19	0.11121
1948	20	0.15428
1948	21	0.18830
1948	22	0.20854
1948	23	0.21582
1948	24	0.21511

Figure 2. Period age-specific fertility rates by calendar year and age of the mothers. A fragment of HFD data for Bulgaria.

In this file, the calendar years and ages are given in the first two columns, and the age-specific fertility rates (ASFRs) for every combination of calendar year and age are given the last column. The file begins from the ASFR values in 1947 for ages 12 and younger, 13, 14, ..., 54, 55 and older. This is followed by the ASFR values in 1948 for ages 12 and younger, 13, etc. While users of statistical packages are very familiar with this universal data format, users who work in Excel may find it difficult to operate with such data. For many purposes, it would be more useful to transform these data into a rectangular matrix of ASFRs, with columns corresponding to calendar years and rows corresponding to the mothers' ages at birth. The latter data shape can be readily used in Excel for constructing a figure with fertility age curves by year and various calculations.

Figure 3 presents another fragment of age-specific data from the HFD in the long format. These are cohort ASFRs by the mother's year of birth (cohort), the mother's age at birth, and birth order (1, 2, 3, 4, 5+).

Cohort	Age	ASFR	ASFR1	ASFR2	ASFR3	ASFR4	ASFR5p
1933	12-	-	-	-	-	-	-
1933	13	-	-	-	-	-	-
1933	14	0.00153	0.00151	0.00002	0.00000	0.00000	0.00000
1933	15	0.00670	0.00656	0.00013	0.00000	0.00000	0.00000
1933	16	0.02110	0.02042	0.00067	0.00000	0.00000	0.00000
1933	17	0.04585	0.04322	0.00263	0.00001	0.00000	0.00000
1933	18	0.07657	0.06820	0.00799	0.00030	0.00007	0.00000
1933	19	0.11371	0.09406	0.01685	0.00245	0.00033	0.00003
1933	20	0.15566	0.12059	0.02958	0.00476	0.00061	0.00011
1933	21	0.17666	0.12132	0.04546	0.00905	0.00070	0.00012
1933	22	0.18849	0.11167	0.06224	0.01310	0.00147	0.00001
1933	23	0.18380	0.08943	0.07305	0.01645	0.00445	0.00042
1933	24	0.17334	0.06599	0.07852	0.01997	0.00631	0.00255
1933	25	0.15626	0.04558	0.07782	0.02157	0.00779	0.00349
1933	26	0.13414	0.03165	0.06874	0.02057	0.00841	0.00476
1933	27	0.11895	0.02288	0.06244	0.01932	0.00868	0.00562
1933	28	0.10017	0.01669	0.05182	0.01668	0.00832	0.00666
1933	29	0.08472	0.01266	0.04223	0.01476	0.00773	0.00733
1933	30	0.06963	0.00946	0.03308	0.01241	0.00694	0.00774
1933	31	0.05333	0.00707	0.02333	0.00943	0.00567	0.00784
1933	32	0.04182	0.00560	0.01666	0.00713	0.00438	0.00805
1933	33	0.03246	0.00445	0.01198	0.00527	0.00349	0.00727
1933	34	0.02735	0.00365	0.00925	0.00518	0.00290	0.00637
1933	35	0.02296	0.00294	0.00708	0.00530	0.00236	0.00529
1933	36	0.01811	0.00233	0.00529	0.00421	0.00182	0.00446
1933	37	0.01352	0.00172	0.00370	0.00288	0.00131	0.00390
1933	38	0.00965	0.00139	0.00250	0.00187	0.00097	0.00292
1933	39	0.00729	0.00115	0.00175	0.00128	0.00078	0.00233
1933	40	0.00530	0.00083	0.00117	0.00089	0.00056	0.00185
1933	41	0.00380	0.00053	0.00084	0.00060	0.00035	0.00147
1933	42	0.00250	0.00035	0.00051	0.00035	0.00021	0.00108
1933	43	0.00150	0.00023	0.00024	0.00023	0.00013	0.00068
1933	44	0.00086	0.00013	0.00011	0.00017	0.00008	0.00038
1933	45	0.00035	0.00005	0.00005	0.00006	0.00004	0.00015
1933	46	0.00022	0.00003	0.00003	0.00004	0.00003	0.00010
1933	47	0.00013	0.00003	0.00002	0.00002	0.00001	0.00006
1933	48	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1933	49	0.00002	0.00001	0.00000	0.00000	0.00000	0.00000
1933	50	0.00003	0.00002	0.00000	0.00000	0.00000	0.00000
1933	51	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000
1933	52	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1933	53	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1933	54	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000
1933	55+	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1934	12-	-	-	-	-	-	-
1934	13	0.00009	0.00009	0.00000	0.00000	0.00000	0.00000
1934	14	0.00164	0.00162	0.00002	0.00000	0.00000	0.00000
1934	15	0.00724	0.00710	0.00014	0.00000	0.00000	0.00000
1934	16	0.02138	0.02070	0.00068	0.00000	0.00000	0.00000
1934	17	0.04502	0.04222	0.00278	0.00001	0.00000	0.00000
1934	18	0.07804	0.06963	0.00810	0.00027	0.00003	0.00000
1934	19	0.11795	0.09863	0.01651	0.00249	0.00027	0.00004
1934	20	0.15922	0.12357	0.03022	0.00504	0.00034	0.00003

Figure 3. Cohort age-specific fertility rates by the year of the mother’s birth, the age of the mother, and birth order. A fragment of HFD data for Bulgaria.

Such HFD data can be transformed into five rectangular (age)x(year) matrices corresponding to birth orders 1, 2, 3, 4, 5, and higher.

This technical report provides an Excel spreadsheet for reshaping long-format data into wide format.

How to use it

The “Reshape-Wide_2.xls” spreadsheet has two worksheets: “ReadMe” with a short explanatory note, and “InputData.” The input (long) HFD data should be placed in the green range on the InputData worksheet.

This can be done in several ways. For example, on the HFD Age-Specific Data webpage of the country of interest, you can click on the dataset that you want to reshape. The selected file will appear on the screen. Select the entire range of ASFRs (you may press Ctrl-A to select all data) and press Ctrl-C for copying.

Open a blank spreadsheet in Excel and paste the data in it by clicking on the Paste button (Home tab in Office-2007) or by pressing Ctrl-V. Split the data by column using the “Data\Text to Columns” menu (Data tab\Text to column button). When splitting the text by column, you can select “space” as the column delimiter. Select the data range and copy it to the green range in the InputData worksheet.

Press the green button in the InputData worksheet to transform data into wide format. The output wide data will appear in a newly opened spreadsheet, which will be activated. This spreadsheet can then be saved in any folder.

Go back to the InputData page in Reshape-Wide_2.xls and press the yellow button to clean the input range.

References

HFD 2012. The Human Fertility Database. Available at <http://www.humanfertility.org>