



Wageindicator

in Argentina, Armenia, Azerbaijan, Belarus, Belgium, Brazil, Chile, China, Columbia, Denmark, Finland, France, Georgia, Germany, Guatemala, Hungary, India, Italy, Kazakhstan, Kirgizstan, Mexico, Netherlands, Paraguay, Poland, Russia, South Africa, South Korea, Spain, Sweden, Tajikistan, Turkmenistan, Ukraine, United Kingdom, United States, Uzbekistan

www.wageindicator.org

Cleaning of the wage data in the WageIndicator dataset

Kea Tijdens, Universiteit van Amsterdam/AIAS, Netherlands
Fikret Oez, Hans-Böckler-Stiftung/WSI, Germany
Marianne Oldenburg, Datamatch, Utrecht, Netherlands
12/04/2008

Contents

Introduction	1
Cleaning the waged working hours	1
Cleaning the reported gross and net wages	2
Cleaning the reported bonuses	2
Cleaning the reported wage period	2
Computing hourly wages	2
Appendix: The MIN-MAX range in hourly wages (Jan 2008)	3

Introduction

For the Annual release 2007 for all countries, the three authors undertook a large effort to clean the wage data in the dataset. Four variables are critical in the data cleaning, notably the waged working hours, the gross and net reported wage, the bonuses, and the wage period. This paper presents the steps undertaken. The reference to the variable names and the syntax numbers is made predominantly for the convenience of the authors, not for those of the readers.

Cleaning the waged working hours

- 1) The waged working hours **HRSWAG1** are computed, based on the reported waged working hours and checked against the reported contractual working hours, usual working hours, and standard working hours in the firm. These hours variables are computed from the time-format variables derived from the questionnaire (syntax 24c).
- 2) Waged working hours > 80.75 hours per week are set to missing (syntax 24c).

Cleaning the reported gross and net wages

- 3) If the reported net wage is larger than the reported gross wage, this is reversed (into **WAGEGR1, WAGENE1**, syntax 24d).
- 4) The odd values in the reported gross and/or net wages are set to missing, f.e. 999999999 (in **WAGEGR1, WAGENE1**, syntax 24d).
- 5) Annual wages above a maximum are set to missing, based on the computation of a maximum of 2100 hours per year times the maximum hourly wage (syntax 24d). It is derived from the MIN-MAX range for hourly wages, discussed in 18) and reported in the appendix.

Cleaning the reported bonuses

- 6) Every reported bonus is limited to maximum 4 times the reported gross wage (syntax 24f).
- 7) If the sum of bonuses **WASUM1** is $> 2/3$ of the reported gross wage, it is set missing (syntax 24f).
- 8) Then the sum of bonuses **WASUM1** (if >0) is deducted from the gross wage (if >0) (syntax 24f) (Recall that respondents are asked to report their wage including bonuses).
- 9) The sum of bonuses **WASUM1** (if >0) is deducted from the net wage (if >0). If the net wage then is < 0 , the net wage is set to 0 (syntax 24f).

Cleaning the reported wage period

- 10) The reported wage period is computed in **WAGEPER3**, including those who have reported an annual wage and who have reported the number of months in which this wage was earned.
- 11) For euro countries for each value in **WAGEPER3**, the wage boundaries are checked, and substituted if needed (syntax 24e).

Computing hourly wages

- 12) Hourly wages are computed for the 1st time (**WAGEGRHR, WAGENEHR**, syntax 25a).
- 13) Gross hourly wages $> 3*$ net hourly wages are checked (syntax 25a).
- 14) Hourly wages are computed for the 2nd time (**WAGEGRHR, WAGENEHR**, syntax 25a).
- 15) If missing, gross hourly wages are computed for observations with net hourly wages only, based on the mean gross/net ratio per country, based on previous releases of the WageIndicator data (syntax 25b).
- 16) If missing, net hourly wages are computed for observations with gross hourly wages only, based on the mean gross/net ratio per country, based on previous releases of the WageIndicator data (syntax 25b)

- 18) Extreme values for gross hourly wages are set to missing, thus hourly wages outside the MIN-MAX range that in the 2008 web-survey is set to instantly check reported wages during survey completion (WAGEGRHR, WAGENEHR, syntax 25c), see the Appendix.
- 19) Based on the hourly wages, the weekly, monthly, annual and the logarithm of the gross hourly wages are computed (**WAGEGRWE, WAGEGRMO, WAGEGRAN, WAGEGRHL**, syntax 25e - g)

Appendix: The MIN-MAX range in hourly wages (Jan 2008)

MINIMUM: Amount under which the reported hourly wage is considered unreliable

MAXIMUM: Amount over which the reported hourly wage is considered unreliable

Country Name	Currency	MINIMUM	MAXIMUM
Austria	EU Euro	1	150
Argentina	AR Peso	2	150
Belgium	EU Euro	1	150
Brazil	BR Cruzeiro	2	150
China	CN Renminbi	2.6	261.2
Denmark	DK Krone	7.5	1,000
Finland	EU Euro	1	150
France	EU Euro	1	150
Germany	EU Euro	1	150
Hungary	HU Forint	100	25000
India	IN Rupee	21	4500
Italy	EU Euro	1	150
Korea, Rep.	KR Won	1,000	125,000
Mexico	MX Peso	3	600
Netherlands	EU Euro	1	150
Poland	PL Zlotty	1.5	500
Russian Federation	RU Rouble	8.08	1443.42
South Africa	ZA Rand	5	3000
Spain	EU Euro	1	150
Sweden	SE Krona	9.5	1,400
United Kingdom	GB Pound	0.75	120
United States	US Dollar	2	300