

## NOTES TO THE DATA FILE

### Variables in the data file

*num* – the questionnaire id

*hhid* - the household id

*reg* – region

*ur* – urban/rural indicator

*stratum* – the stratum number

*cluster* – the cluster number

*mount* – mountain/lowland indicator

*w1f* – weights to be used for consumption expenditure variables

*w2p* – weights to be used for all variables except consumption expenditure variables

*complet* - indicator of whether the questionnaire has been fully completed.

*food\_exp* – monthly expenditure on eating in the household (term J3 in the formula (\*) below)

*lnf\_exp* – long-term non-food expenditure calculated on monthly basis (term J1 in the formula (\*) below)

*edu\_exp* - education expenditure calculated on monthly basis (term E in the formula (\*) below)

*hc\_exp* – health care expenditure calculated on monthly basis (term F12 in the formula (\*) below)

*fd\_o\_exp* – expenditure on eating out of home calculated on monthly basis (term J4 in the formula (\*) below)

*cnf\_exp* - current non-food expenditure calculated on monthly basis (term J5 in the formula (\*) below)

*tot\_exp* – total per month consumption expenditure of the household (term C in the formula (\*) below)

*fs* – family size

*eqadult* – number of equivalent adults in the household

*ea\_sc08* - number of equivalent adults in the household allowing for scale effect with exponent 0.8

*exp\_ea08* – per month per equivalent adult consumption expenditure of the household allowing for scale effect with exponent 0.8.

### Weighting scheme

In total, 4808 households were actually interviewed. Of them, in 4646 households the questionnaire was completed fully and in 162 households the questionnaire was completed just partly. Specifically, in partly completed questionnaires Parts A-H and K as well as Table J1 of the questionnaire were completed during the first visit while the completion of the weekly expenditure and food consumption data (Tables J2-J7) during the second visit turned out to be impossible due to various reasons (absence of household members, refusal to continue providing information or some other reason). In the provided database the variable “completion” is included which marks by 1 those households where the questionnaire was fully completed and

by 2 those where the questionnaire was just partly completed. In order to not to lose any piece of obtained information, two sets of weights have been calculated: one set of weights (w1f) is assigned to only those 4646 households where the questionnaire was fully completed and it may be used for analyzing all variables, especially in cases where consumption expenditures are to be analyzed. The second set of weights (w2p) is assigned to all responding households and it may be used for analyzing all variables except those concerning consumption expenditures and food intake (variables from Tables J2-J7 and the variables *food\_exp*, *fd\_o\_exp*, *cnf\_exp*, *tot\_exp*, *exp\_ea08* from the data file). Thus when analyzing a set of variables not containing the above mentioned variables, the weights w2p may be switched while when analyzing consumption expenditures or related issues, the weights w1c should be used. Anyway, the weights w1c are universal and they may be used for any kind of analysis.

The weights have been calculated by strata according to the following formulas:

$$w1f_s = \frac{N_s - r_s}{N_s} \frac{H_s}{n_s(f)}$$

$$w2p_s = \frac{N_s - r_s}{N_s} \frac{H_s}{n_s(f) + n_s(p)},$$

where

$H_s$  is the number of households in the  $s$ -th stratum according to the Department of Statistics data;

$N_s$  is the sample size in the  $s$ -th stratum;

$r_s$  is the number of non-existing households from the sample in the  $s$ -th stratum;

$n_s(f)$  is the number of fully completed questionnaires in the  $s$ -th stratum;

$n_s(p)$  is the number of partly completed questionnaires in the  $s$ -th stratum.

### **Calculation of the number of equivalent adults in a household**

For calculation of the number of equivalent adults the following table, used by Department of Statistics, was applied

Age	Gender	Equivalent Adult coefficient
<8		0.64
>=8 and <16		1
>=16 and <65	Male	1
>=16 and <60	Female	0.84
>=65	Male	0.88
>=60	Female	0.76

The scale effect exponent was set to 0.8, as done by Department of Statistics.

## Calculation of the total per month consumption expenditure of a household

The total per month consumption expenditure of a household is calculated according to the formula

$$C = F12 + E + J1 + J3 + J4 + J5 \quad (*)$$

Here **C** is the total consumption expenditure of the household, while (the terms under the summation signs denote the contents of the corresponding cells of the questionnaire or, if there is a row of cells under that heading, the sum of contents of all the cells in the row):

1. **F12** is health care expenditure per month calculated by the formula

$$F12 = \frac{1}{12} \sum_{i=1}^{12} F12.i$$

2. **E** is education expenditure per month calculated by the formula

$$E = \frac{1}{12} \sum_{i=8}^{13} Ei$$

3. **J1** is long-term non-food expenditure per month calculated by the formula (the number of addends may exceed 89)

$$J1 = \frac{1}{12} \sum_{i=1}^{89} J1.i$$

4. **J3** is food expenditure in the household per month calculated by the formula (number of addends may exceed 76)

$$J3 = \frac{365.25}{12 \times 7} \sum_{i=1}^{76} J_{2.i} \times p(i);$$

here  $p(i)$  is the median price of the  $i$ -th product in the country, calculated on the basis of the data of Table J2. See below the description of calculation of the prices  $p(i)$ .

5. **J4** is food expenditure outside home per month calculated by the formula

$$J4 = \frac{365.25}{12 \times 7} \sum_{i=1}^{15} J4.i$$

6. **J5** – is current non-food expenditure per month calculated by the formula

$$J5 = \frac{365.25}{12 \times 7} \sum_{i=1}^{15} J5.i$$

### Calculation of the prices $p(i)$

For the  $i$ -th product, from Table J2 the price of every purchase is calculated by dividing the value by quantity (only those observations are used for which both quantity and value are known). The weighted (by the sampling weights) median of the obtained prices is taken as  $p(i)$ .