



# seriss

SYNERGIES FOR EUROPE'S  
RESEARCH INFRASTRUCTURES  
IN THE SOCIAL SCIENCES

Deliverable Number: 2.9

Deliverable Title: Database containing necessary information for computation of population margins

Work Package: 2 - Representing the population

Deliverable type: Other

Dissemination status: Public

Submitted by: SHARE ERIC

Authors: Claudio Rossetti (MEA)

Date Submitted: October, 2017

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.





[www.seriss.eu](http://www.seriss.eu)  @SERISS\_EU

SERISS (Synergies for Europe's Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe's social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe's grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Rossetti, C. (2017) *Database containing necessary information for computation of population margins*. Deliverable 2.9 of the SERISS project funded under the *European Union's Horizon 2020 research and innovation programme* GA No: 654221. Available at: [www.seriss.eu/resources/deliverables](http://www.seriss.eu/resources/deliverables)

# Database containing necessary information for computation of population margins

Claudio Rossetti

*LUISS Business School, Italy*

October, 2017

## Abstract

This report provides a documentation describing the database containing necessary information for computation of population margins used to compute calibrated weights in the first six waves of the Survey on Health, Ageing and Retirement in Europe (SHARE). The report is organized as follows. After a brief overview of the database and the documentation, we describe the data sources. Then we include a detailed description of the dataset. Finally, we describe a simple Stata code providing an example on how to use the database containing necessary information for computation of population margins.

## Contents

1. Overview of database and documentation .....	4
2. Data sources.....	4
3. The dataset .....	5
3.1 Region .....	6
3.2 Age .....	8
3.3 Gender .....	8
4. A simple example on how to use the dataset.....	8

## 1. Overview of database and documentation

The file *margins\_nuts1.dta* contains the dataset with necessary information for computation of population margins. The dataset is in Stata format, but we also exported into Excel. The database contains population and number of deaths by year, region, age and gender for all countries involved in the first six waves of the Survey on Health, Ageing and Retirement in Europe (SHARE). The list of included countries is the following:

AT	Austria
BE	Belgium
CH	Switzerland
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FR	France
GR	Greece
HR	Croatia
HU	Hungary
IE	Ireland
IL	Israel
IT	Italy
LU	Luxembourg
NL	Netherlands
PL	Poland
PT	Portugal
SE	Sweden
SI	Slovenia

Along with this documentation, the file *margins\_example.do* provides a Stata code with a simple example on how to use the database containing necessary information for computation of population margins. A description of this Stata code is provided at the end of the report.

## 2. Data sources

The data come from the Central Bureau of Statistics for Israel, and Eurostat for all other European countries involved in the first six waves of SHARE. The regional demographic statistics of the Eurostat dataset is updated regularly under the “Population and social conditions” theme. Specifically, the regional demographic statistics provides annual data on population, vital events (live

births and deaths), total and land areas of the regions and demographic indicators for regions and statistical regions at NUTS-2 levels for each EU Member State, Candidate and EFTA countries. The completeness of the figures depends on the availability of data received from the responsible national statistical institutes. Starting with March 2013, demographic statistics at regional level reflect the NUTS-2010 classification and the new statistical regions for Croatia. Most of the countries affected by the NUTS-2010 changes sent to Eurostat back time series for the new regional breakdown.

### 3. The dataset

The database contains populations and number of deaths from 2004 to 2015 for all countries involved in the first six waves of SHARE listed above. Population is defined as population on 1st January of each year. Below we show a snapshot of the dataset.

The screenshot shows the Data Editor (Browse) interface for the dataset 'margins\_nuts1.dta'. The main window displays a table with the following columns: year, country, nuts1, sex, age, deaths, and population. The data is filtered for the year 2004. The table shows 27 rows of data for the year 2004, with the first row highlighted. The status bar at the bottom indicates 'Ready', 'Vars: 7', 'Order: Dataset', 'Obs: 138,960', 'Filter: Off', and 'Mode: Browse'.

Year	Country	NUTS1	Sex	Age	Deaths	Population
2004	AT	AT1	M	30	19	23238
2004	AT	AT1	M	31	27	25004
2004	AT	AT1	M	32	23	26041
2004	AT	AT1	M	33	14	27285
2004	AT	AT1	M	34	30	28965
2004	AT	AT1	M	35	33	30287
2004	AT	AT1	M	36	42	30310
2004	AT	AT1	M	37	40	30817
2004	AT	AT1	M	38	58	30865
2004	AT	AT1	M	39	49	30902
2004	AT	AT1	M	40	52	30864
2004	AT	AT1	M	41	40	30283
2004	AT	AT1	M	42	52	29599
2004	AT	AT1	M	43	56	28583
2004	AT	AT1	M	44	81	27211
2004	AT	AT1	M	45	76	25554
2004	AT	AT1	M	46	95	25466
2004	AT	AT1	M	47	91	24551
2004	AT	AT1	M	48	108	23134
2004	AT	AT1	M	49	110	21966
2004	AT	AT1	M	50	101	21316
2004	AT	AT1	M	51	113	21239
2004	AT	AT1	M	52	137	20344
2004	AT	AT1	M	53	132	21143
2004	AT	AT1	M	54	164	21669
2004	AT	AT1	M	55	181	23129

Population and number of deaths in each country and year are provided over three dimensions, namely region, age and gender.

### 3.1 Region

Regions for European countries are statistical regions at NUTS1 level, defined according to the Nomenclature of territorial units for statistics NUTS 2010/EU-27. The Nomenclature of Territorial Units for Statistics (NUTS) was drawn up by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. The NUTS classification has been used in EU legislation since 1988, but it was only in 2003, after three years of preparation, when a European Parliament and Council Regulation on NUTS have been adopted. The NUTS nomenclature was created and developed according to the principles that NUTS favours institutional breakdowns and regional units of a general character, and it serves as a reference for the collection, development and harmonization of the European Union's regional statistics, for socio-economic analyses of the regions, and for the framing of EU regional policies.

The current NUTS nomenclature applicable from 1 January 2012 subdivides the economic territory of the European Union into 97 regions at NUTS 1 level, 270 regions at NUTS 2 level and 1,294 regions at NUTS 3. Below those, two levels of Local Administrative and populations at the three NUTS levels, for each Member Units (LAUs) have been defined. The upper LAU level State and for the European Union as a whole. LAU level 1 (formerly NUTS level 4) is defined only for the following countries: Bulgaria, the Czech Republic, Denmark, Germany, Estonia, Ireland, Greece, France, Cyprus, Lithuania, Luxembourg, Hungary, Malta, Poland, Portugal, Slovenia, Slovakia, Finland and the United Kingdom. The lower LAU level (formerly NUTS level 5) consists of around 120,000 municipalities or equivalent units in the 27 EU Member States (as of 2010).

The aim of the NUTS classification is to ensure that comparable regions appear at the same NUTS level. As population size has been defined in the regulation as a key indicator for comparability, each level inevitably contains regions that differ greatly in terms of area, economic weight or administrative powers. This heterogeneity across the EU often simply reflects the situation at Member State level. Below we show the map of the regions at NUTS1 level in the European Union.



Because we found some inconsistencies over time in population and number of deaths by NUTS1 codes for France and Greece, we did not use NUTS1 regions in these two countries. Thus, population and number of deaths in these two countries are not reported separately by NUTS1 region. We hope this can be somehow fixed in future releases of the Eurostat data. Moreover, because no NUTS

[www.seriss.eu](http://www.seriss.eu) GA No 654221 7

nomenclature is defined for Israel, we decided to use two codes, namely the code IL1 for Jews and others, and the code IL2 for Arabs.

### **3.2 Age**

Population and number of deaths in the dataset are also included separately by age. Age is defined as single years from the age of 30 to 88, plus the open-ended class aged 89 or over. Data from the Central Bureau of Statistics about population and number of deaths in Israel separately for Jews and others, and for Arabs is only available for age classes. Specifically, data are available for the age class 0-49 and then for five-year age classes from the age of 50 to 84, plus the open-ended class aged 85 or over. For this reason, to ensure homogeneity in the data with all other countries, we included population and number of deaths for single years of age for Israel too, by assuming uniform distribution of population and deaths in each year of the five-year age classes.

Finally, note that there are a few cases for which source data is only available for the open-ended class aged 85 or over. In order to ensure a greater detail, we decided to include in the dataset the open-ended class aged 89 or over for all countries and years. Specifically, data are only available for the open-ended age class 85 or over in the following cases: population in years 2012 and 2013 for Croatia, population in 2006 and 2007 for Poland, and both population and number of deaths for all years in Israel. Thus, note that in these few cases, the open-ended age class labelled as 89+ actually contains data for aged 85 or more.

### **3.3 Gender**

Population and number of deaths are included separately for males (M) and females (F). Total numbers (T) are also included.

## **4. A simple example on how to use the dataset**

Attached to this report we provide a Stata code with a simple example on how to use the database containing necessary information for computation of population margins. Below we show a snapshot of the Stata code.

