TU DORTMUND
Fakultät Statistik
Prof. Dr. Jörg Rahnenführer
JProf. Dr. Antonia Arsova
Dr. Birte Hellwig
M.Sc. Julia Duda
M.Sc. Sven Pappert

Winter semester 2020/2021

Project I
Descriptive analysis of demographic data

The *International Data Base* (*IDB*) of the *U.S. Census Bureau* contains various demographic data (currently from 1950 to 2050) on all states and regions of our world that are recognized by the US Department of State and have a population of 5000 or more. The sources of the database are information from state institutions, such as censuses, surveys or administrative records, as well as estimates and projections by the U.S. Census Bureau itself.

The dataset in the file `census_2019_1999.csv` contains a small extract from the IDB. It includes life expectancy and fertility rates for 228 countries and territories from 1999 and 2019. For the exact definitions of these variables see `https://www.census.gov/programs-surveys/international-programs/about/glossary.html`. Life expectancy is stratified by sex. The countries and territories are divided geographically into 21 regions. For further details regarding data collection see `https://www.census.gov/programs-surveys/international-programs/about/idb.html`.

Tasks:

1. Describe the frequency distributions of the variables. Consider also the differences between the sexes.
2. Are there bivariate correlations between the variables?
3. Are the values of the individual variables comparatively homogeneous within the regions and heterogeneous between the regions?
4. How have the values of the variables changed over the last 20 years, i.e. comparing 1999 with 2019?

For tasks 1–3, consider only the year 2019. This project serves to practice the use of explorative and descriptive methods. Therefore, use appropriate statistical measures and graphical methods for the analysis in all parts of the project.

Submission

Submission of the report and the corresponding (executable and commented) program code until Friday, November 27, 2020, 08:30 am, in Moodle.