Exercises on Targeted Record Swapping using µ-ARGUS.

We will be using µ-ARGUS to apply TRS in different ways. Analysing the results can be done using your preferred software (SPSS, SAS, R, Python, …). The output of µ-ARGUS will be an ascii-file along with some metadata.

The dataset we will be using in the exercises is `test_data_10k_mu.csv` with corresponding meta-data file `test_data_10k_mu.rda`.

Exercise 1

Examine the metadata file `test_data_10k_mu.rda`. What do the specification tags (between "<"and ">") mean?

Note: using a free format file (`.csv`) as input microdata to µ-ARGUS is much slower than a fixed format file as input microdata.

Exercise 2

a. Start µ-ARGUS and load the microdata file `test_data_10k_mu.csv`.
b. Examine the metadata from the "Specify Metadata" window.
c. Set things up to be able to use Targeted Record Swapping (TRS).
d. Apply TRS with
    i. Similarity profile "Size"
    ii. Hierarchy structure "NUTS1", "NUTS2"
    iii. Risk defining variables "AGE.M"
    iv. K-anonymity k = 3
    v. Swap rate 0.05
    vi. Seed 2501
e. Save the results to `test_data_10k_muSafe.saf`.
f. Check the report file `test_data_10k_muSafe.html`, the microdata file `test_data_10k_muSafe.saf` and metadata file `test_data_10k_muSafe.rds`.

Exercise 3

Analyse the results from Exercise 2 (you may adjust and use the R script `RscriptAnalysis.R`):

a. Calculate table "NUTS2" by "SEX" from original data and from perturbed data. Compare the results by eye.
b. Calculate AD, RAD and HD per NUTS2 region for table "NUTS2" by "SEX".

Exercise 4

Repeat exercise 2 and 3 but in this case with swap rate = 0.2. Compare the results with the results from swap rate = 0.05.

Exercise 5

Repeat exercise 2 but in this case make sure that all regional variables are swapped consistently (without adjusting the set of hierarchy-variables). Check the consistency.

Exercise 6

Repeat exercises 2

i. Similarity profile “Size”, “HST”
ii. Hierarchy structure “NUTS1”, “NUTS2”
iii. Risk defining variables “COC.M”, “POB.M”
iv. Swap rate 0.05
v. Seed 2501

Repeat exercise 3 for table “NUTS2” by “SEX” by “AGE.M”.

Explain the RAD results.