Exercises on noise addition with the Cell Key Method using τ-ARGUS.

We will be using τ-ARGUS to apply CKM in different ways.

The dataset we will be using in the exercises is `test_data_10k_tau.csv` with the corresponding meta-data file `test_data_10k_tau.rda`.

In this folder you will also find four ptable files, hierarchy files for NUTS3, LAU2, COC.M, POB.M and AGE.H.

Exercise 1

Examine the metadata file `test_data_10k_tau.rda`.
What do the specification tags (between "<"and ">") mean?

Note: using a free format file (`.csv`) as input microdata to τ-ARGUS is much slower than a fixed format file as input microdata, but only during loading the microdata.

Exercise 2

a. Start τ-ARGUS and load the microdata file `test_data_10k_tau.csv`
b. Examine the metadata from the "Specify Metadata" window

What ptable file is specified to be used by default for frequency count tables?

c. Specify the frequency count table "SEX" by "COC.M" by "NUTS3"

How many cells does this table have?

d. Change view such that you see NUTS3 by COC.M and SEX defines the layers
e. Apply CKM with ptable ptab1

How many cells were not changed? What percentage of cells did not change?

f. Save the results in CKM format to `test_data_10k_tau_ptab1.tab` with perturbed and original values as well as the differences
g. Have a look at the report file `test_data_10k_tau_ptab1.html`, the file with the protected table `test_data_10k_tau_ptab1.tab` and the metadata file `test_data_10k_tau_ptab1.rda`.

What are the mean values for AD, RAD and DR for this table? Calculated over non-empty cells and over all cells.

Exercise 3

Apply CKM with ptable ptab2 to table “SEX” by “COC.M” by “NUTS3”.

How many cells does this table have?

How many cells were not changed? What percentage of cells did not change?

What are the mean values for AD, RAD and DR for this table? Calculated over non-empty cells and over all cells.

Exercise 4

Apply CKM with ptable ptab6 to table “SEX” by “COC.M” by “NUTS3”.

How many cells does this table have?

How many cells were not changed? What percentage of cells did not change?

What are the mean values for AD, RAD and DR for this table? Calculated over non-empty cells and over all cells.

Compare the results of the information loss measures with the results from Exercise 3 (using ptab2). Can you explain the differences?

Exercise 5

Specify table “NUTS2” by “SEX” by “AGE.M” by “HST” by “POB.L”.

Apply CKM with ptable ptab3.

How many cells does this table have?

How many cells were not changed? What percentage of cells did not change?

What are the mean values for AD, RAD and DR for this table? Calculated over non-empty cells and over all cells.