Regression Project: EU-SILC analyses

Marcus Wurzer

2024-11-17

Data

We will use a data set generated from real Austrian EU-SILC (European Union Statistics on Income and Living Conditions) data. To get it, please install the `simFrame` package that includes the `eusilcP` data set.

You find a short description of all available variables writing `?eusilcP`, for more detailed information please have a look at the following document:

https://moodle.technikum-wien.at/mod/resource/view.php?id=1834126

Note: All the variables in `eusilcP` have an *n*-suffix which stands for *net*, but you may get the desired information by looking for the *gross* variant of the variable (having a *g*-suffix), e.g., `py120n`/`py120g` for net and gross sickness benefits, respectively.

Data subsets for the student's groups

Each group gets their unique problem, i.e., the analyses will differ with respect to the dependent variable, the independent variables and the (row) subset of `eusilcP` to be used. In the following table, look at the row with your group's number to see which dependent (`y`) and independent (`x1` to `x4`) variables you have to use and for which subset of the data set the analysis has to be done:

group	y	x1	x2	x3	x4	subset
1	py120n	main	region	hsize	age	South Austria
2	py120n	main	ecoStat	hsize	age	West Austria
3	py140n	main	citizenship	hsize	age	South Austria
4	py090n	main	citizenship	hsize	age	West Austria
5	py090n	main	ecoStat	hsize	age	West Austria
6	py100n	main	ecoStat	hsize	age	East Austria
7	py110n	gender	citizenship	hsize	age	West Austria
8	py010n	main	region	hsize	age	West Austria
9	py050n	gender	region	hsize	age	South Austria
10	py090n	gender	citizenship	hsize	age	West Austria
11	py110n	gender	region	hsize	age	West Austria
12	py110n	main	region	hsize	age	East Austria
13	py100n	main	citizenship	hsize	age	South Austria
14	py090n	gender	ecoStat	hsize	age	South Austria
15	py140n	gender	ecoStat	hsize	age	West Austria
16	py050n	main	region	hsize	age	South Austria
17	py140n	main	region	hsize	age	West Austria
18	py120n	gender	ecoStat	hsize	age	South Austria
19	py130n	main	region	hsize	age	West Austria
20	py140n	gender	region	hsize	age	South Austria
21	py010n	main	citizenship	hsize	age	East Austria

How do you create the required subset? Each of the values given in the `subset` column is a so-called First-level NUTS (or NUTS-1) of the European union, with NUTS standing for **Nomenclature of Territorial Units for Statistics**, see:

https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics

Thus, Austria has three NUTS-1 regions that are comprised of several NUTS-2 regions. You get information about the corresponding second-level NUTS for Austria here:

https://en.wikipedia.org/wiki/First-level_NUTS_of_the_European_Union

As can be seen, the second-level NUTS are just the nine federal states of Austria, and we can use the `region` variable given in `eusilcP` for subsetting. Please note that you still may have to use `region` as a predictor, but it will not have 9 levels anymore, but only 2, 3, or 4 (Important: You have to adapt the factor accordingly!).