EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

LAB SESSION 7

In lab sessions 7 and 8, you will work with the same topic *Validation and forecasting*. In these sessions, you will work with one of the specifications for the *Netherlands mode choice* case study developed in previous labs and you will perform different tasks related to the validation of the model and the forecasting and calculation of relevant indicators. The purposes of lab sessions 7 and 8 are the following:

- perform an outlier analysis using the simulation file from Biogeme to generate the probabilities and any statistical software for the analysis,
- compute relevant indicators to better understand the behavior of the population, and
- evaluate different hypothetical scenarios,

First, download the file `Netherlands_Val_For.zip`. It contains:

1. the **data** file `netherlandsRP.dat`,
2. the **model** file `Netherlands_Base_Model.ipynb`, and the results of the estimation in a `pickle` file,
3. the **model** file `Netherlands_Base_Simul.ipynb`, for you to use as template to perform the simulation.

Specifically, in this lab session 7, you will work on the following tasks.

1 Aggregation

Since the procedure that has been used to collect the sample is assumed to be *stratified random sampling*, we need to associate a weight with each group or stratum, and then with each individual. Create an additional column for the dataset `netherlandsRP.dat` containing the individual weights. Define the new dataset as `netherlandsRP_w.dat`.

2 Simulation file

Now we ask you to complete the simulation file `Netherlands_Base_Simul.ipynb` and using your newly created dataset `netherlandsRP_w.dat` in order to answer the following questions:

1. Compute the predicted market shares for car and rail with stratified random sampling.

2. Compare the predicted market shares with the actual choices. More precisely calculate the following shares:

   - share of users choosing car with a higher (unweighted) probability for rail, and
   - share of users choosing rail with a higher (unweighted) probability for car.

   Try to find the possible causes.

---

jp / th / an / no / mpp / mw