EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2020

COMPUTER SESSION 1

The objective of this first lab session is to get familiar with the datasets that will be used during the course. To do so, we recommend to perform the following tasks:

1. Go through the description of the case studies that will be used during the lab sessions. The datasets, together with the associated descriptions, can be found under *Datasets* in the *General* section in Moodle. Each description file contains the following information:

   - description of the choice context,
   - methodology of the data collection,
   - description of the available data, and
   - statistics on certain data variables.

2. Understand the structure of the data file (`.dat`) and compute descriptive statistics for the Netherlands dataset. Notice that the Netherlands dataset contains records for both RP and SP choices.

   For this task, open the `netherlands.dat` file using your preferred software (e.g., Excel, R, Python) and answer the following questions:

   (a) What does each row of this file correspond to?

   (b) What does each column of this file correspond to?

   (c) Identify which variables are related to (i) attributes of the alternatives and (ii) socioeconomic characteristics of the decision-makers.

   (d) How are the variables related to the socioeconomic characteristics of the decision-makers coded?

   (e) Compute the mean, mode, standard deviation, min and max for each variable.

   (f) Visualize the qualitative variables (e.g., gender) by generating charts.

   (g) Visualize the attributes of the alternatives (e.g., travel time) by using histograms.

   (h) Investigate the correlation between the dependent and the independent variables.

   (i) Use scatter plots to investigate bias towards specific alternatives. For instance, try to plot the travel time of the rail alternative for those who chose car against the travel time of the car alternative for those who chose rail.

---

mpp/th/rk/no/jp/mw