
Computer Lab I

Datasets and BIOGEME

Antonin Danalet, Amanda Stathopoulos

Outline

- Teaching assistants:
 - Amanda Stathopoulos
 - Antonin Danalet
- Course website: <http://transp-or.epfl.ch/courses/dca2012/>
- Semester projects: <http://transp-or.epfl.ch/studentProjects.php>
- Organization of the lectures
- Organization of the labs
- Case studies: Available datasets
- BIOGEME: step-by-step example

Pedestrian flow simulation in Lausanne train station

- Goal: analyzing the future expansion scenarios in Lausanne Gare
- Model already developed in VisWalk.
- Use more detailed data.
- Evaluate effects of infrastructure changes on pedestrian flows.
- Skills needed: bit of programming.
- Collaboration with SBB CFF FFS, PTV AG, and VisioSafe.
- If interested, contact Bilal Farooq, Marija Nikolic

Organization of the lectures

- **Your participation to the lectures:**
 - Read material before the beginning of the course!
Textbook and other reading material available on website
<http://transp-or.epfl.ch/courses/dca2012/schedule2012.php>
 - Prof. Bierlaire will give you feedback on the assignments from the labs during the course.

Organisation of the labs

- Case Studies
 - Choice of dataset
 - Testing and interpretation of example models
 - Material available on website
<http://transp-or.epfl.ch/courses/dca2011/labs.php>

Organisation of the labs

- **Your participation to the labs:**
 - We will send you the group assignments by email
 - On an **assigned** dataset, develop your own model specification
 - **By e-mail**, give back your results the **next Friday at noon**
 - **Max. 1 double-sided page** of assignment in PDF format
 - Model specification in .mod (text) format
 - **Output file** in **HTML** format
 - Remarkable results will be discussed at the beginning of the next lecture by Prof. Bierlaire
 - Individual corrections given at next exercise session
 - 3 assignments during the semester.
Schedule on the website.

Case Studies

- Goal: Study models
- Datasets to use models in practice:
 - Choice-Lab marketing
 - Netherland mode choice
 - Swissmetro
 - Choice of residential telephone services
 - Boeing
 - Optima

Case Studies

- Problem statement

Can the observed pattern of choice be explained in terms of basic economic variables, such as relative prices, income, and underlying individual characteristics (gender, age, etc.)?

Datasets

- Choice-Lab marketing

Data on clients of a business-to-business firm that collects and processes financial and customer data for their clients. The dataset includes choices of what information products were purchased by the client over time as well as the choice to remain as a client or drop as a client.

Datasets

- Netherlands mode choice

Data on intercity travelers' choices between the transport modes of rail and car.

Datasets

- **Swissmetro**

Data on travelers' choices of transport mode among a proposed underground system (Swissmetro), traditional train, and car.

Datasets

- Choice of residential telephone services

Data on households' choices of local telephone service.

Datasets

- Boeing

Data on flight options for a particular origin-destination trip.

Datasets

- Optima

Data on transportation mode choice in low-density areas of Switzerland.

More information: <http://transport.epfl.ch/optima>

BIOGEME

- Created by Michel Bierlaire
- State of the art software for estimating models in the field of discrete choice
- Open source
- All models presented in this course can be estimated with BIOGEME
- Website: <http://biogeme.epfl.ch>

BIOGEME

- Two versions are available for Windows and Mac OS X
 - GUI
 - DOS / command line
- We recommend the DOS / command line version

Lab 1

- Today
 - Read data descriptions available on the course web page
 - Step-by-step example with BIOGEME using the Netherlands Mode Choice dataset.

How to install Biogeme?

- `biogeme.exe` should be in `C:\Program Files\biogeme`.
- Open a DOS window (from the Start menu, select Run. In the dialog box, type `cmd` and select OK).
- In order to use BIOGEME from any directory on your computer, the above directory has to be in your “path” (environment variable).
- Type `path=%path%;C:\Program Files\biogeme` (in the DOS window).
 - Has to be typed every time the DOS window is opened
- To check if the installation has been successful, just type `biogeme` in the DOS window. A message displaying the version of BIOGEME should then appear.

How to invoke Biogeme?

- BIOGEME is invoked in a DOS command window under Windows using the following statement structure:

```
biogeme model_file sample_file.dat
```

- 2 types of file invoked: .mod & .dat
- The graphical version of Biogeme `guibiogeme.exe` (also available in `C:\Program Files\biogeme`) is invoked by a double-click on the executable file.

DOS Command Window

Here are some useful commands

- To select a drive (e.g. C), just type `C:` at the prompt.
- To connect to a directory (e.g. `C:\biogeme`), just type `cd C:\biogeme`.
- To see the content of a directory, use Windows Explorer, or type `dir`.
- In order to return to the previous (top) directory, type `cd ..`.

On Mac OS X (and Linux)

Here are some useful commands

- To go into a directory (e.g. `biogeme`), just type `cd biogeme`.
- To see the content of a directory, type `ls`.
- In order to return to the previous (top) directory, type `cd ..`
- To know where you are, type `pwd` (Print Working Directory)

How does BIOGEME work?

- BIOGEME reads:
 - a file containing the model specification
`model_file.mod`
 - a file containing the data `sample_file.dat`
- Both are text documents (`.txt`)
- `biogeme model_file sample_file.dat`
- BIOGEME automatically generates:
 - A file containing the results of the maximum likelihood estimation: `model_file.rep.`
 - The same file in HTML format: `model_file.html`

Example

- Netherlands mode choice
- Choice between rail and car
- 223 observations
- Travel times and travel costs are used as explanatory variables for the model, and the deterministic utility specifications are

$$V_{\text{car}} = \text{ASC}_{\text{car}} + \beta_{\text{cost}} \text{car}_{\text{cost}} + \beta_{\text{time}} \text{car}_{\text{time}}$$

$$V_{\text{rail}} = \beta_{\text{cost}} \text{rail}_{\text{cost}} + \beta_{\text{time}} \text{rail}_{\text{time}}$$



- Model is specified in `model_file.mod`

Example

Extract from the file containing the data `sample_file.dat`

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
...
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267

- 1 row = 1 observation
- 1 column = 1 variable

Estimate your first model

- Download the two files from the course web page to the directory of your choice (e.g. Desktop).
- In the DOS window, move to this directory using the `cd` command.
- Invoke BIOGEME:

```
biogeme model_file sample_file.dat
```

- Open the HTML file `model_file.html`.
- We briefly discuss it.