Computer Lab III
Recapitulation and OPTIMA Case Study

Evanthia Kazagli, Anna Fernandez Antolin & Antonin Danalet

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne

September 30, 2014

Today

- Administrative information (rules, groups, number of pages, etc.).
- Summary of what you have learnt so far:
  - Types of coefficients (generic, alternative specific, socioeconomic).
  - Tests (likelihood ratio test, t-test).
- Help: dealing with missing data.
- You will start working on your first assignment.

Administrative information

- The report for the assignment is **compulsory**:
  - No report, no final exam!
- **Deadline** to submit the report: Friday, October 10, 2014, at noon.
- Submit by email to Anna:
  - the specification as a .mod file;
  - the .html file with the results; and
  - a .pdf (or Word) file containing the description of the model specification together with its underlying hypothesis and interpretation of results.
- 1-2 pages.
- Dataset to be used: Mode choice in Switzerland **(Optima)**.
- Groups of 4 persons, following the list in the email you received.

Dataset: Mode choice in Switzerland (Optima)

- Dataset: "optimaDataset.dat"
- Description of the data and variables available on the webpage:
  - General description;
  - List of variables.

OPTIMA dataset

- Objective: study mode choice in low density areas of switzerland.
- Data collection: Revealed Preferences (RP).
- Survey type: Mail survey.
- Context: Loop trips = cyclical sequence of trips.
- Dataset with 2265 observations including:
  - Trip features;
  - Socioeconomic variables.

OPTIMA data file

Table : extract 'optimaDataset.dat'

ID	TripPurpose	Choice	CostPT	CostCarCHF	age	LangCode
10350017	1	1	12.4	4.54	27	1
10350025	3	0	3	0.64	-1	1
10350075	1	1	24	3.38	63	1
10350085	1	1	10.8	1.66	57	1
10350086	1	1	9.6	2.71	58	1
10350100	3	1	3	1.63	80	1
10350120	3	1	24.8	6.98	64	1
10350125	3	1	4.8	0.99	68	1
10350125	3	1	2.2	0.19	68	1

Types of parameters

- In the linear formulation of utility functions, the βs are called coefficients or parameters. Different types:
  - Alternative specific constants (ASC).
  - Generic:
    - Appearing in all utility functions with equal coefficients.
    - Assume all choice makers have the same marginal utility among the alternatives.
  - Alternative specific:
    - Different coefficients among utility functions.
    - Capture the marginal utility specific to an alternative.
  - Alternative-specific socioeconomic:
    - Reflect differences in preference as functions of characteristics of the decision-maker.

Tests

Goal: test alternative specifications of the explanatory variables in the utility functions. Different tests:

- t-test
- Likelihood ratio test

t-test

- Goal: test whether a particular parameter in the model differs from some known constant –usually zero.
- Valid only asymptotically.
- t-test > 1.96 means significant parameter (95% confidence interval).

Likelihood ratio test (LRT)

- Goal: compare different specifications (i.e. models).
- Restricted model (e.g. some βs $= 0$ –null hypothesis) vs unrestricted model.
- Number of degrees of freedom (d.o.f.): difference between the number of estimated coefficients in the restricted and unrestricted model.
- χ^2 test with this number of d.o.f.: $-2(\mathcal{L}(\hat{\beta}_{unrestricted}) - (\hat{\beta}_{restricted}))$
- Find the LRT excel file in the Utilities tab on biogeme's official homepage.

Interpretation

- Is the coefficient significant?
- Are the signs reasonable?
  - Coefficients are expected to have a behavioral meaning, i.e. a negative coefficient means lower utility when the variable value increases, and higher utility when the variable value decreases (e.g. cost, travel time etc.).
  - The interpretation the other way around is the same (e.g. speed).

Dealing with missing data

- Section `[Exclude]` tells BIOGEME to NOT consider some observations.
- **Example** of `binary_generic_boeing.mod`
  `[Exclude] ArrivalTimeHours_1 == -1 || BestAlternative_3`
  1. Excludes missing data (-1) for variable ArrivalTimeHours_1
  2. Excludes alternative BestAlternative_3 (1 Stop with 2 different airlines)
- `[Exclude]` needs to be used in the Optima case study to exclude soft modes and only consider choice between public transportation and car for your assignment (binary logit model).

Dealing with missing data

- **Example**: if you want to use the gender variable (q17_gender).
- **Solution 1**
  - Exclude missing data (-1 and 99) from **the whole data set**
    → `[Exclude] q17_gender == 99 || q17_gender == -1`

Dealing with missing data

- **Example**: if you want to use the gender variable (q17_gender).
- **Solution 2 (better)**
  - Measure taste heterogeneity between men and women by introducing a term for missing data in the utility.
  - In section `[Expressions]` define:
    - `MissingGender = ( ( q17_Gender == -1 ) + ( q17_Gender == 99 ) ) > 0`
  - In section `[Utilities]` specify:
    - `+ Male_Opt2 * Male + MDGender * MissingGender`

Today's plan

Group work for your first assignment:

- Gather as a group;
- Work on the Optima dataset;
- Generate a .mod file (base);
- Continue improving the initial specification and use the tests to decide on the best model specification.

Lab assignment: Your goals for the report

1. Work with your group on your own specification of a Binary Logit on the Optima mode choice dataset;
2. Examine the data & and the variables' description and coding;
3. Formulate your own hypothesis and write your model specification;
4. Test your hypothesis;
5. Back to 3 until you obtain a satisfactory model specification;
6. Write a report (max double-sided page report);
7. Send your results before the dealine:
   - the .html file
   - .mod file
   - .pdf file.