Computer Lab III

Summary

Evanthia Kazagli, Anna Fernandez Antolin, Xinjun Lai

Today

- Administrative stuff (rules, groups, number of pages, ...)
- Summary of what you've learnt so far
  - Types of variables (generic, specific, socioeconomic)
  - Tests (likelihood ratio test, t-test)
- Help: dealing with missing data
- You'll work on your first assignment

Administrative stuff

- Report is **compulsory**:
  - No report, no final exam!
- **Deadline** to submit the report: Friday, October 11, 2013, at noon
- Submit by email to Eva and Anna:
  - .html file (from BIOGEME)
  - .pdf (or Word) document containing a description of the model specification with its underlying hypotheses
- 1-2 pages.
- Data set: Mode choice in Switzerland (Optima)
- Groups of 4 persons, following the list in the email you received

Data set: Mode choice in Switzerland (Optima)

- Data set “optimaTOT3_valid.dat” on the website:
- Description of the data and variables available on the website:
  - General description
  - List of variables

OPTIMA

- Motivation: study mode choice in low density areas of switzerland
- Data-collection: Revealed Preferences (RP)
- Survey-type: Mail survey
- Collects 'loops' = cyclical sequence of trips
- database with 1906 observations
  - trip features
  - socioeconomic variables

OPTIMA - Data file

Table 1: extract 'optimaTOT3_valid.dat'

ID	TripPurpose	Choice	CoutTP1	CoutAutoCHF	age	CodeLangue
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 explanatory variables

- In linear formulation of utility function, βs are called coefficients or parameters. Different kinds:
  - Alternative specific constants (ASC)
  - Generic
    - Appearing in all utility functions with equal coefficients
    - Assume all choice makers have the same marginal utility between the alternatives
  - Alternative specific
    - Different coefficients between 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

- t-test
- Likelihood ratio test

Tests: t-test

- Goal: test whether a particular parameter in the model differs from some known constant, often zero
- Valid only asymptotically (since we work with nonlinear models)
- t-test > 1.96 means significant parameter (95% confidence interval)

Tests: Likelihood ratio test

- Goal: compare different specifications (i.e., models)
- Restricted model (e.g., some βs $= 0$) (null hypothesis) vs unrestricted model
- Number of degrees of freedom: difference between the number of estimated coefficients in the restricted and unrestricted models.
- χ^2 test with this number of freedom:
  $-2(\mathcal{L}(\hat{\beta}_{unrestricted}) - (\hat{\beta}_{restricted}))$

Interpretation

- Is the coefficient significant?
- Sign
  - Coefficients are expected to have a behavioral meaning: a negative coefficient means lower utility when the variable is high, and higher utility when the variable is low, e.g. travel time, cost.
  - The other way around: same interpretation

Dealing with missing data

- Section `[Exclude]` tells BIOGEME not to consider some observations.
- **Example** of `binary_generic_boeing.mod`
  - `[Exclude] ArrivalTimeHours_1 == -1 || BestAlternative_3`
  - Excludes missing data (-1) for variable ArrivalTimeHours_1
  - Excludes alternative BestAlternative_3 (1 Stop with 2 different airlines)
- The same needs to be done for the Optima case study: exclude soft modes, and keep public transportation and cars

Dealing with missing data

- **Example**: if want to use gender variable (q17_gender)
- **Solution 1**
  - Exclude missing data (-1 and 99) from **whole data set**
  - `[Exclude] ArrivalTimeHours_1 == -1 || BestAlternative_3 || 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 utility
  - `[Exclude]` section identical
    - `[Exclude] ArrivalTimeHours_1 == -1 || BestAlternative_3`
  - 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 .mod file (base)
- test an idea / hypothesis

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 data;
2. Examine the data & and the variables' description;
3. Write your own model with new variables;
4. Formulate your own hypothesis;
5. Test your hypothesis;
6. Back to 3 until you obtain a satisfactory model specification;
7. Write a report (max double-sided page report);
8. Send your results before the dealine:
   - the .html file
   - .mod file
   - .pdf file.