

Lab I

Choice data and datasets

Evanthia Kazagli
Meritxell Pacheco & Anna Fernandez Antolin

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

September 19, 2017



Outline

1 Part 1

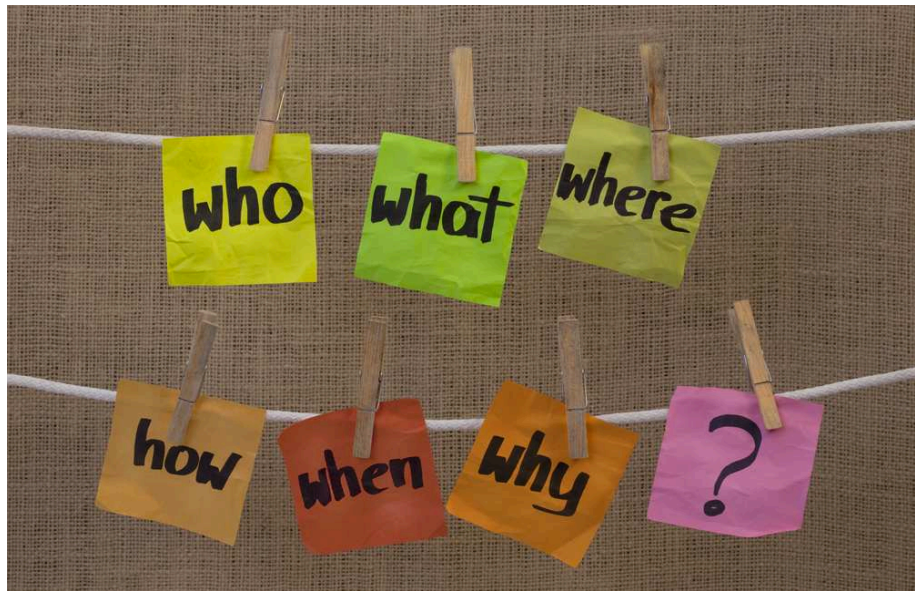
- Useful information
- Organization of the course
- Organization of the labs
- Your participation and involvement
- The project assignment

2 Part 2

- Choice data



Part 1



Teaching assistants

- Meritxell Pacheco
- Evanthia Kazagli
- Anna Fernandez Antolin



Useful information

- Course webpage:
<http://transp-or.epfl.ch/courses/dca2017/>
- Homework info:
<http://transp-or.epfl.ch/courses/dca2017/homework.php>
- Exam info:
<http://transp-or.epfl.ch/courses/dca2017/exam.php>

During the semester you will have to submit one assignment.



Organization of the course

- First part of the course:
 - you are expected to learn the theory by yourselves, using the online documents and videos that are provided.
- Second part of the course:
 - covers more advanced material with traditional ex-cathedra lectures.
- Exercises and laboratories are organized every week.



Organization of the course

- The organization of the semester is described here: Course schedule:
<http://transp-or.epfl.ch/courses/dca2017/schedule2017.php>
- During the five weeks appearing in white, there will be no lecture, but there will be **laboratories from 10:15 to 12:00**.
- Course material available here:
<http://transp-or.epfl.ch/courses/dca2017/slides.php>
- Textbook and additional reading material available on the course webpage under the **Additional Material** tab.



Organisation of the labs

The necessary material is available on the course webpage
<http://transp-or.epfl.ch/courses/dca2017/labs.php>

- 1 Computer labs using biogeme:
 - Work with one dataset.
 - Test and interpret the provided example models.
 - Specify and interpret your own models.
- 2 Exercises with pen and paper.
- 3 Homework practice quizzes to be solved during the next lab.



Your participation to the labs

- The labs and the assignment will be organized in groups.
- ~~The groups will be determined by the teaching assistants and communicated via email to enrolled participants. The groups will be determined by the teaching assistants and communicated via email to enrolled participants.~~
- We ask you to make groups of **4 people** and to send one email per group to meritxell.pacheco@epfl.ch with the names of the group members before **October 3th**. For those not allocated to a group by then, we will randomly create the groups.
- Work jointly with your group.

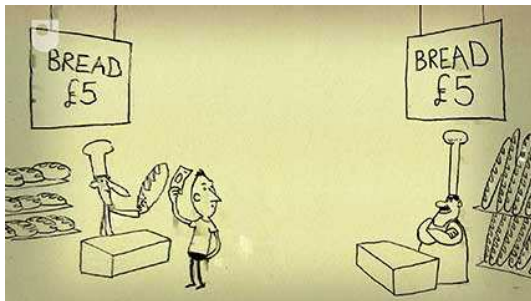


The project assignment

- It accounts for **10% of the final grade**.
- **By e-mail**, submit your project:
 - Report of **maximum 2 double-sided pages** in .pdf format.
 - Model specification in .py format.
 - **Output file** in .html format.
 - The excel file you used for the analysis.
- **Deadline: November 17th**.



Part 2: Choice data



Sampling

- Identify the population of interest.
- In general, it is not possible to collect data about each individual.
- Identify a list of N representative individuals.
- Various sampling methods are presented later in this course.
- Collect choice data for each individual in the sample.

Choice context

Revealed preferences

- Observe actual behavior.
- Real market situations.
- Example: scanner data in supermarkets.

Stated preferences

- Hypothetical situations.
- Choice context defined by the analyst.
- Example: Swissmetro.

Revealed preferences

Data about the decision-maker: socio-economic characteristics

- Age, income, level of education, etc.
- Collected in any survey.
- Not specific to choice models.
- Collect those that seem relevant for the analysis.

Choice set

- Identify the list of alternatives considered by the respondent.
- Context dependent.
- Awareness difficult to observe.

Revealed preferences

Data about the alternatives

- Utility^a is a latent concept, cannot be observed.
- Value of the attributes.
- Particularly difficult for non chosen alternatives.

^aOne assumption of the Discrete choice theory that is studied in this course is that the decision maker associates a utility with each alternative. Utility is a function that captures the attractiveness of an alternative. It is presented in next weeks class.

Observed outcome

- The chosen alternative

Stated preferences

Data about the alternatives

- Constructed by the analyst.
- Provided for each alternative
- Experimental design.

Preferences

- Choice
- Ranking
- Rating
- Allocation

Preference data

Consider the following beers

- 1 Cardinal
- 2 Kronenbourg
- 3 Orval
- 4 Tsing Tao



Choice

What is your preferred option?

Preference data

Consider the following beers

- 1 Cardinal
- 2 Kronenbourg
- 3 Orval
- 4 Tsing Tao



Ranking

Rank the beers, from the best to the worst

Preference data

Consider the following beers

- 1 Cardinal
- 2 Kronenbourg
- 3 Orval
- 4 Tsing Tao



Rating

Associate a rate from 0 (worst) to 10 (best) with each beer

Preference data

Consider the following beers

- 1 Cardinal
- 2 Kronenbourg
- 3 Orval
- 4 Tsing Tao



Allocation

Distribute 100 points among the beers

A transportation example

Boeing Commercial Airplanes

- 2004—2005.
- Designed by Boeing staff with the assistance of Jordan Louviere of the University of Technology, Sydney.
- Objective: understanding the sensitivity that air passengers have toward the attributes of an airline itinerary.
- Recruitment: intercepting customers of an internet airline booking service that searches for low-cost travel deals

Boeing Commercial Airplanes

Pick Your Preferred Flight

Three flight options are described for your trip from Chicago to San Diego. These are options that might be available on this route or might be new options actively being considered for this route as well as replacing some options that are offered now. The options differ from each other in one or more of the features described on the left.

Please evaluate these options, assuming that everything about the options is the same except these particular features. Indicate your choices at the bottom of the appropriate column and press the Continue button.

FEATURES	Non-Stop (Option 1)	1 Stop (Option 2)	1 Stop (Option 3)
Departure time (local)	6:00 PM	4:30 PM	6:00 PM
Arrival time (local)	8:14 PM	8:44 PM	9:44 PM
Total time in air	4 hr 14 min	4 hr 44 min	4 hr 44 min
Total trip time	4 hr 14 min	6 hr 14 min	5 hr 44 min
Legroom <input type="checkbox"/>	typical legroom	2-in more of legroom	4-in more of legroom
Airline [Airplane]	Depart Chicago Continental Airlines [B737] to San Diego	Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego	Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego
Fare	\$565	\$485	\$620
1. Which is MOST attractive?	<input checked="" type="radio"/> Option 1	<input type="radio"/> Option 2	<input type="radio"/> Option 3
2. Which is LEAST attractive?	<input type="radio"/> Option 1	<input checked="" type="radio"/> Option 2	<input type="radio"/> Option 3
3. If these were the ONLY three options available, I would NOT make this trip by air. <input type="radio"/> Yes <input checked="" type="radio"/> No			

RP data: advantages

- Real life choices
- Possibility to replicate market shares
- Decision-makers have to assume their choice
- “A bike or a Ferrari?” — “A Ferrari, of course!”
- Real constraints involved

RP data: drawbacks

- Limited to existing alternatives, attributes and attributes levels.
- Lack of variability of some attributes
- Lack of information about non chosen alternatives
- High level of correlation
- Data collection cost
- In general, one individual = one observation

SP data: advantages

- Exploring new alternatives, attributes and attributes levels
- Control of the attributes variability
- Control on all alternatives
- Control on the level of correlation
- One individual can answer several questions

SP data: drawbacks

- Hypothetical situations
- Cannot be used for market shares
- Decision-makers do not have to assume their choice
- Real constraints not involved
- Credibility
- Valid within the range of the experimental design
- Policy bias (example: “every body else should take the bus”)
- Justification bias (or inertia)
- Framing: phrasing of the question matters
- Anchoring: one variable explains it all
- Fatigue effect



Summary

- Both revealed and stated preferences data have pros and cons
- RP: real behavior
- SP: control of the experiment
- It is common to combine them

