Introduction
Lab I: Choice data and datasets

Michel Bierlaire
Meritxell Pacheco

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

September 18, 2018

TRANSP-OR

Outline

1. Introduction to the course
   - The team
   - Useful information
   - Organization of the course
   - Organization of the labs
   - Project assignment

2. Choice data and datasets
   - Sampling
   - Choice context
   - Revealed preferences
   - Stated preferences
   - A transportation example

TRANSP-OR

Introduction to the course

The team

Michel Bierlaire

Meritxell Pacheco

Yuki Oyama

Thibaut Richard

Nicholas Molyneaux

Useful information

- Course webpage:
  http://transp-or.epfl.ch/courses/dca2018/
- Self-learning material:
  https://www.edx.org/course/introduction-choice-models-epflx-discretechoicex
- Homework info:
  http://transp-or.epfl.ch/courses/dca2018/homework.php
- Exam info:
  http://transp-or.epfl.ch/courses/dca2018/exam.php

  During the semester you will have to submit one assignment.

TRANSP-OR

Organization of the course

- First part of the course:
  - self-learning material
  - edX platform
  - Q&A session

- Second part of the course:
  - ex-cathedra lecture
  - Tuesdays 08:15-10:00

- Sessions of written exercises and computer laboratories are organized every week: Tuesdays 10:15 - 12:00

Organization of the labs

MATHEMATICAL MODELING OF BEHAVIOR

Home   Schedule   Homework   Laboratories   Clickers   Discussion group   Exam   Course material   Additional Material

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 exercises to be solved during the next lab (some weeks)

TRANSP-OR

ÉCOLE POLYTECHNIQUE
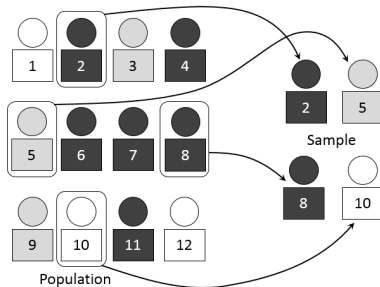FÉDÉRALE DE LAUSANNE

Project assignment

- **25% of the final grade**
- Organize yourselves in groups of 4 students
- Send an email with the group members: **meritxell.pacheco@epfl.ch**
- More information and deadlines: **Laboratories** tab of the website

TRANSP-OR

Choice data and datasets

Sampling

- Identify the population of interest
- In general, it is not possible to collect data about each individual
- Identify a list of representative individuals (sample)
- Collect choice data for each individual in the sample

Decision-maker: socio-economic characteristics

- Collected in any survey
- Not specific to choice models
- Gather the ones that seem relevant for the analysis
- Example: age, income, level of education

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
- Several scenarios can be created for each respondent
- Preferences are expressed through statements or intentions
- Example: Swissmetro

TRANSP-OR

Revealed preferences

Choice set

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

Data about the alternatives

- Value of the attributes
- Particularly difficult for non-chosen alternatives

Observed outcome

- Chosen alternative

Stated preferences

Choice set
- Constructed by the analyst
- May contain hypothetical alternatives
- May vary across scenarios and respondents

Data about the alternatives
- Defined by the analyst and provided for each alternative
- Experimental design

Observed outcome
- Preferences: choice, ranking, rating, allocation

TRANSP-OR

Stated preferences: preference data (1)

Consider the following beers

1. Cardinal
2. Kronenbourg
3. Chimay
4. Tsing Tao

Choice
What is your preferred option?

Stated preferences: preference data (2)

Consider the following beers

1. Cardinal
2. Kronenbourg
3. Chimay
4. Tsing Tao

Ranking

Rank the beers, from the best to the worst

Stated preferences: preference data (3)

Consider the following beers

1. Cardinal
2. Kronenbourg
3. Chimay
4. Tsing Tao

Rating

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

Stated preferences: preference data (4)

Consider the following beers

1. Cardinal
2. Kronenbourg
3. Chimay
4. Tsing Tao

Allocation
Distribute a 100 points among the beers

TRANSP-OR

RP data: advantages and drawbacks

✔
- Real life choices (decision-makers have to assume their choice)
- Possibility to replicate market shares
- Real constraints are involved

✗
- Limited to existing alternatives, attributes and attribute levels
- Lack of variability of some attributes
- Lack of information about non-chosen alternatives
- High level of correlation
- High data collection cost
- In general, one observation per individual

TRANSP-OR

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SP data: advantages and drawbacks

✔
- Exploring of new alternatives, attributes and attribute levels
- Control on the alternatives, attribute levels, level of correlation
- One individual can answer several questions (observations)

✗
- Hypothetical situations (real constraints are not involved)
- It cannot be used to calculate market shares
- No need to assume the choice by decision-makers (credibility)
- Valid within the range of the experimental design
- Policy bias ("everyone else should go by bus"), justification bias
- Framing (phrasing of the question)
- Anchoring (one variable explains it all), fatigue effect

A transportation example (1)

Boeing Commercial Airplanes

- 2004–2005
- Designed by Boeing staff with the assistance of Jordan Louviere (University of Technology, Sydney)
- **Goal**: understanding the sensitivity of air passengers towards the attributes of an airline itinerary
- **Recruitment**: intercepting customers of an internet airline booking service (low-cost travel deals)

TRANSP-OR

A transportation example (2)

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 ☐	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?	● Option 1	● Option 2	● Option 3
2. Which is LEAST attractive?	● Option 1	● Option 2	● Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. ● Yes ● No