

CHOICE MODELING IN R

CLASS 2

Wiktor Budziński

EXERCISE

Run a basic MNL model
for the *netherlands.dat*
dataset

Exercise Week 2.R

1. Use only RP part of the data
 - database\$rp variable
2. Control for
 1. In-vehicle travel time
 2. Out-of vehicle travel time
 3. Travel cost
3. Should there be an ASC in the model?
4. What is the WTA for a travel time increase of 15 minutes?

CHOICE MODELING

MULTINOMIAL LOGIT



MNL VS. CL

A lot of textbooks differentiate between multinomial logit and conditional logit

- MNL is often assumed to have only individual-specific variables

$$U_{ij} = \mathbf{z}_i \boldsymbol{\gamma}_j + \varepsilon_{ij}$$

- CL focuses on alternative-specific variables

$$U_{ij} = \mathbf{x}_{ij} \boldsymbol{\beta} + \varepsilon_{ij}$$

This difference is rather arbitrary, usually forced by the software which expect different data types for different models

- MNL uses a short format
 - One row per choice, dependent variable is categorical (no. of levels equal to no. of alternatives)
- CL uses a long format
 - Multiple rows per choice (equal to no. of alternatives), dependent variable is binary
 - Not true for Apollo

In modern choice modeling literature, this model is usually called MNL

MULTINOMIAL LOGIT

Last week, our MNL specification was simplified

- It only concerned attributes – alternative-specific variables
 - E.g., Legroom for the itinerary choice
- Often, we also want to control the individual-specific variables
 - E.g., socio-demographics: age, gender, etc.
- These have to be treated differently than attributes in the choice model
 - As they are “constant” for a given choice set, their coefficients cannot be identified in the same way as constant

There are 3 main ways to incorporate individual-specific variables

- As alternative-specific effects
- As interactions with attributes
- As explanatory variables of scale

MULTINOMIAL LOGIT

Alternative-specific effects

$$U_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_j + \varepsilon_{ij}$$

Interactions with attributes

$$U_{ij} = \mathbf{X}_{ij}(\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\Gamma}) + \varepsilon_{ij}$$

As explanatory variables of scale

$$U_{ij} = \exp(\mathbf{Z}_i\boldsymbol{\gamma})(\mathbf{X}_{ij}\boldsymbol{\beta}) + \varepsilon_{ij}$$

MULTINOMIAL LOGIT

The first two approaches are usually interpreted as an observed preference heterogeneity

- Either for given alternative or attribute
- Different people have different taste
- For example, individuals with higher income may be more likely to travel by a car, all other things held constant

Modeling scale is interpreted as explaining error term's variance (heteroskedasticity)

- Note, that it only works with individual-specific variables
- Although the model is identifiable if alternative-specific variables enter scale, it does not have such an interpretation

These interpretations are to some extent driven by our assumptions as preference heterogeneity and scale heterogeneity cannot be fully disentangled in choice models

EXERCISE

Analyze consumers' preferences for the air travel using MNL
CMiR_2.R file

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

EXERCISE

Analyze consumers' preferences for the air travel using MNL
CMiR_2.R file

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 [8737] 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			

Let's go to  

CHOICE MODELING

IIA PROPERTY

IIA PROPERTY

The multinomial logit assumes that error terms are uncorrelated across different alternatives

This combined with the extreme value distribution leads to the so called “*independence from irrelevant alternatives*” property

- Relative choice probability of two alternatives is not affected by other alternatives
 - Adding or dropping some alternatives should not affect it

$$\frac{P(Y_i = j)}{P(Y_i = k)} = \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\gamma}_j)}{\exp(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\gamma}_k)}$$

In real data this assumption is often violated

- Some alternatives are likely to be correlated due to some unobserved factors
- For example, car vs. public transport

IIA PROPERTY

IIA assumption can be tested using Hausman-McFadden test

It compares estimates from the basic model with the estimates from the model with a single alternative dropped

- If IIA holds, then this should not affect the results
- You have to be careful if you have ASCs in the model

Test statistic is given by $H = (\boldsymbol{\beta}_r - \boldsymbol{\beta}_f)' (\mathbf{V}_r - \mathbf{V}_f)^{-1} (\boldsymbol{\beta}_r - \boldsymbol{\beta}_f)$

- It has chi squared distribution under the null hypothesis that the coefficients did not change

EXERCISE

Test IIA assumption for
the estimated MNL
model

CMiR_2.R file

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

EXERCISE

Test IIA assumption for
the estimated MNL
model

CMiR_2.R file

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			

Let's go to  