Computer Lab II Biogeme & Binary Logit Model Estimation

Evanthia Kazagli, Anna Fernandez Antolin & Antonin Danalet

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

September 23, 2014

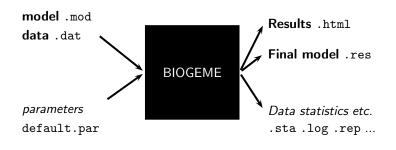


• Further introduction to BIOGEME

• Estimation of Binary Logit models



How does BIOGEME work?





- File extension .dat
- First row contains column (variable) names.
- One observation per row.
- Each row must contain a choice indicator.
- Example with the Netherlands transportation mode choice data: choice between car and train.



netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
219	1	35	2.416	6.4	1.283
220	1	30	2.334	2.083	1.667
221	1	35.7	1.834	16.667	2.017
222	1	47	1.833	72	1.533
223	1	30	1.967	30	1.267



1 5 / 26

996

EK, AFA, AD (TRANSP-OR)

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time
1	0	40	2.5	5	1.167
2	0	35	2.016	9	1.517
3	0	24	2.017	11.5	1.966
4	0	7.8	1.75	8.333	2
5	0	28	2.034	5	1.267
	Unique ide	entifier of ob	servations		
 219	Unique ide 1	entifier of ob 35	servations 2.416	6.4	1.283
	Unique ide 1 1			6.4 2.083	1.283 1.667
219	Unique ide 1 1 1	35	2.416		
219 220	Unique ide 1 1 1 1	35 30	2.416 2.334	2.083	1.667
219 220 221	Unique ide 1 1 1 1 1	35 30 35.7	2.416 2.334 1.834	2.083 16.667	1.667 2.017



EK, AFA, AD (TRANSP-OR)

 \equiv

996

netherlands.dat

id	choice	rail_cost	rail_time	car_cost	car_time			
1	0	40	2.5	5	1.167			
2	0	35	2.016	9	1.517			
3	0	24	2.017	11.5	1.966			
4	0	7.8	1.75	8.333	2			
5	0	28	2.034	5	1.267			
		Choice indi	Choice indicator, 0: car and 1: train					
219	1	35	2.416	6.4	1.283			
220	1	30	2.334	2.083	1.667			
221	1	35.7	1.834	16.667	2.017			
222	1	47	1.833	72	1.533			
223	1	30	1.967	30	1.267			



EK, AFA, AD (TRANSP-OR)

 \equiv

996

- File extension .mod
- Must be consistent with data file.
- Contains deterministic utility specifications, model type etc.
- The model file contains different [Sections] describing different elements of the model specification.



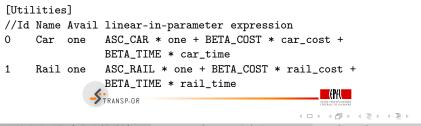
• How can we write the following deterministic utility functions in BIOGEME?

$$V_{car} = ASC_{car} + \beta_{time}time_{car} + \beta_{cost}cost_{car}$$
$$V_{rail} = \beta_{time}time_{rail} + \beta_{cost}cost_{rail}$$



[Choice] choice

[Beta]				
// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0



EK, AFA, AD (TRANSP-OR)

September 23, 2014 10 / 26

= nar

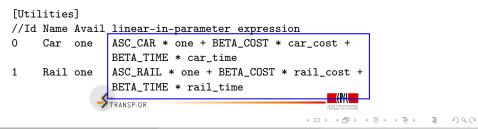
[Choice] choice [Beta] // Name DefaultValue LowerBound UpperBound status ASC_CAR 0.0 -100.0 100.0 0 ASC_RAIL 0.0 -100.0 100.0 1 BETA_COST 0.0 -100.0 100.0 0 BETA TIME 0.0 -100.0 100.0 0 [Utilities] //Id Name Avail linear-in-parameter expression 0 ASC_CAR * one + BETA_COST * car_cost + Car one BETA_TIME * car_time 1 Rail one ASC_RAIL * one + BETA_COST * rail_cost + BETA_TIME * rail_time TRANSP-OR + E + + E + < □ > < A > > EK, AFA, AD (TRANSP-OR) Computer Lab II

September 23, 2014 11 / 26

= nar

[Choice] choice

	[Beta]				
// Name		DefaultValue	LowerBound	UpperBound	status
	ASC_CAR	0.0	-100.0	100.0	0
	ASC_RAIL	0.0	-100.0	100.0	1
	BETA_COST	0.0	-100.0	100.0	0
	BETA_TIME	0.0	-100.0	100.0	0



EK, AFA, AD (TRANSP-OR)

September 23, 2014 12 / 26

[Choice] choice	What is o	ne?		
[Beta]	Which is t	the type of	model?	
// Name	DefaultValue	LowerBound	UpperBound	status
ASC_CAR	0.0	-100.0	100.0	0
ASC_RAIL	0.0	-100.0	100.0	1
BETA_COST	0.0	-100.0	100.0	0
BETA_TIME	0.0	-100.0	100.0	0
0 Car one 1 Rail one	BETA_TIME ASC_RAIL > BETA_TIME STRANSP-DR	one + BETA * car_time * one + BETA * rail_time	_COST * car_ A_COST * rai	1_cost +
EK, AFA, AD (TI	RANSP-OR)	Compute	er Lab II	September 23, 2014

୬ < ୍ୟ 13 / 26

```
[Expressions]
// Define here arithmetic expressions for name that are not directly
// available from the data
one = 1
[Model]
// Currently, only $MNL (multinomial logit), $NL (nested logit), $CNL
// (cross-nested logit) and $NGEV (Network GEV model) are valid keywords
//
$MNL
```



Model and Data Files

- How to read and modify model files?
- How to read data files?
 - GNU Emacs, vi, TextEdit (Mac) or Wordpad (Windows)
 - Notepad (Windows) should not be used!



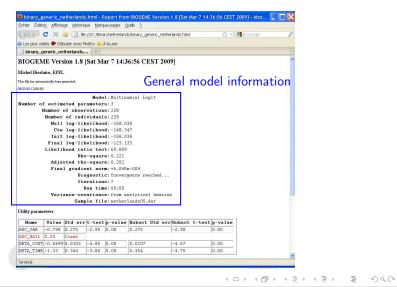
BIOGEME - Results - Netherlands dataset

Echier Édit	ion Afficha	ige <u>Historia</u>	que Marc	iue-pages 🤇	utils <u>2</u>						
< > -	C ×	🏠 🗋 能	:///C:/Bin	aryNetherland	s/binary_generic_ne	etherlands.html	公,	🛃 - Google	P		
Les plus v	isités 🌩 Dé	buter avec P	irefox 应	À la une							
binary_	generic_ne	therlands.	*						-		
BIOGE	ME Ver	sion 1.8	[Sat M	ar 7 14:3	6:56 CEST 2	009]			^		
Michel Bier	laire, EPFI										
This file has asto	matically been										
09/23/10 12:08:0											
			Mode	d · Multin	omial Logit						
Number o	f estim	ated par			antor pogro						
		of obset									
	Number	of ind:	ividual	s: 228							
	Null	log-lil	telihoo	d: -158.0	38						
	Cte	log-lil	kelihoo	d:-148.3	47				10		
	Init	log-lil	kelihoo	d:-158.0	38						
	Final	log-lil	kelihoo	d: -123.1	33						
	Likeli	hood rat	tio tes	t: 69.809							
		Rhe	o-squar	e:0.221							
	Adju	sted rhe	-squar	e:0.202							
	Fina	l gradie	ent nor	m:+6.045	a-004						
		Die	gnost	c: Conver	gence reached.	22					
		Ite	ratio	s: 7							
		2	Run tir	e:00:00							
	Vari	ance-co	varianc	e:from a	nalytical hess	ian					
		Sam	ple fil	e: nether	lands05.dat						
Utility para	neters										
Name	Value	Std err	t-test	n-value	Robust Std e	err Robust t-t	est n-value	1			
ASC_CAR	-0.798	0.270	-2.95	0.00	0.275	-2.90	0.00				
ASC_RAIL	0.00	fixed					Ċ.				
BETA_COST	r -0.0499	0.0103	-4.85	0.00	0.0107	-4.67	0.00	1			
BETA TIM		0.344	-3.86	0.00	0.354	-3.75	0.00				
		1		1	1	1	1	-	~		
Terminé											

EK, AFA, AD (TRANSP-OR)

September 23, 2014 16 / 26

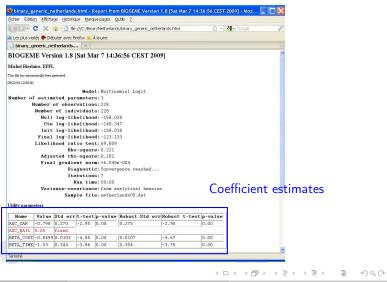
BIOGEME - Results



EK, AFA, AD (TRANSP-OR)

September 23, 2014 17 / 26

BIOGEME - Results



EK, AFA, AD (TRANSP-OR)

September 23, 2014 18 / 26

- Further introduction to BIOGEME
- Estimation of Binary Logit models



Binary Logit Case Study

Available datasets:

- Airline itinerary choice (Boeing)
- Choice-Lab marketing
- Mode choice in Netherlands
- Residential Telephone Services
- Mode choice in Switzerland (Optima)
- Descriptions available on the course webpage.
- Optima dataset does not contain .mod files. A specification has to be proposed as an assignment (next lab session).



How to go through the Case Studies

- Choose a dataset to work with (data descriptions are available on the course webpage).
- Copy the files related to the chosen dataset and case study from the course webpage.
- Go through the .mod files with the help of the descriptions.
- Run the .mod files with BIOGEME.
- Interpret the results and compare your interpretation with the one we have proposed.
- Develop other model specifications.



Course webpage

• http://transp-or.epfl.ch/

 \rightarrow Teaching \rightarrow Mathematical modeling of behavior \rightarrow Laboratories

- BIOGEME software (including documentation and utilities)
- For each Case Study:
 - Data files;
 - Model specification files;
 - Possible interpretation of results.



Today's plan

Independent work on 2-3 Case Studies

- choose a case study;
- estimate a model;
- interpret the results.
- ② Group work
 - gather in groups;
 - generate .mod file (base);
 - test an idea/ hypothesis.



Specifying models: Recommended steps

- Formulate a-priori hypothesis:
 - Expectations and intuition regarding the explanatory variables that appear to be significant for mode choice.
- Specify a minimal model:
 - Start simple;
 - Include the main factors affecting the mode choice of (rational) travelers;
 - This will be your starting point.
- Continue adding and testing variables that improve the initial model in terms of *causality*, and *efficiency* with respect to what actually happened in the sample.



Evaluating models

The main indicators used to evaluate and compare the various models are summarised here:

- Informal tests:
 - signs and relative magnitudes of the parameters β values (under our a-priori expectations);
 - *trade-offs* among some attributes and ratios of pairs of parameters (e.g. reasonable value of time).
- Overall goodness of fit measure:
 - adjusted rho-square (likelihood ratio index): takes into account the different number of explanatory variables used in the models and normalizes for their effect
 → suitable to compare models with different number of independent variables. We check this value to have a first idea about which model might be better (among models of the same type), but it is not a statistical test.



- E - E - E -

Evaluating models (cont.)

Statistical tests:

- *t-test values*: statistically significant explanatory variables are denoted by t-statistic values remarkably higher/ lower than ± 2 (for a 95% level of confidence);
- final log-likelihood for the full set of parameters: should be remarkably different from the ones in the naive approach (null log-likelihood and log-likelihood at constants); we ask for high values of likelihood ratio test $[-2(LL(0) LL(\beta))]$ in order to have a model significantly different than the naive one.

Test of entire models:

• likelihood ratio test $[-2(LL(\hat{\beta}_R) - LL(\hat{\beta}_U))]$: used to test the null hypothesis that two models are equivalent, under the requirement that the one is the restricted version of the other. The likelihood ratio test is X^2 distributed, with degrees of freedom equal to $K_U - K_R$ (where K the number of parameters of the unrestricted and restricted model, respectively).

