EPFL
ENAC TRANSP-OR
Prof. M. Bierlaire

Mathematical Modeling of Behavior
Fall 2016

EXERCISE SESSION 2

Exercise 1 We want to build a model that predicts the market penetration of electric vehicles (EV) as a function of the income level in a sample of 1000 individuals. Using the data shown in Table 1, perform the following tasks:

	Income			
EV	low	medium	high	
yes	15	50	135	200
no	200	450	150	800
	215	500	285	1000

Table 1: Contingency table of EV ownership conditional on income level

1. Using the maximization of the likelihood function, estimate the parameters π_1, π_2 and π_3, where:

$$\begin{aligned}\text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{low}) &= \pi_1 \\ \text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{medium}) &= \pi_2 \\ \text{Prob}(\text{EV} = \text{yes} \quad | \quad \text{income} = \text{high}) &= \pi_3\end{aligned} \tag{1}$$

   Comment on the values of the parameters. Do they make sense? *Hint: write the likelihood function and find its maximum.*

2. What is the final log likelihood of this model?

3. Estimate a model where you consider only two income categories: high on the one hand, and on the other hand low and medium together. Estimate a model where you consider only one income category. Among the three models, which one would you prefer? Why? *Hint: Calculate the final log likelihood associated with each model and perform likelihood ratio tests.*

4. Suppose now that after some economical growth, the income distribution is as follows: 75 individuals with low income, 400 individuals with medium income and 525 with high income. Considering the best of the three models, what would be the EV market penetration in this scenario?

Exercise 2

1. Provide three motivations for using discrete choice models in the context above. Why can we not use linear regression instead?

2. How would you model the EV example using linear regression? Compare both approaches.

Exercise 3 In a mode choice experiment the follow utility functions are defined for private motorized modes (pmm) and public transportation (pt):

$$\begin{aligned} U_{pmm,n} &= -\beta_c \cdot \text{cost}_{pmm,n} - \beta_t \cdot \text{time}_{pmm,n} + \varepsilon_{pmm,n} \\ U_{pt,n} &= -\beta_c \cdot \text{cost}_{pt,n} - \beta_t \cdot \text{time}_{pt,n} + \varepsilon_{pt,n} \end{aligned} \tag{2}$$

where $\text{cost}_{pmm,n}$ and $\text{cost}_{pt,n}$ are the cost of the trip by private motorized modes and public transportation respectively for individual n in CHF, and $\text{time}_{pmm,n}$ and $\text{time}_{pt,n}$ are their times in minutes. $\varepsilon_{pmm,n}, \varepsilon_{pt,n} \overset{iid}{\sim} \text{EV}(0,1)$

Our sample contains the following 10 observations:

Individual	Choice	time_{pmm}	time_{pt}	cost_{pmm}	cost_{pt}
1	pmm	10	20	2.3	1
2	pt	5	10	2.3	0.5
3	pmm	35	30	9	12
4	pmm	20	22	1.5	2
5	pt	6	7.5	2	1.25
6	pt	10	15	5	3.5
7	pt	8	5	3	2
8	pt	19	18	4	5
9	pt	22	19	7	8.5
10	pmm	8	8.5	3	9

The parameter estimates are $\beta_c = 1.38$ and $\beta_t = 0.363$

1. What is the value of time according to the model in [CHF/h]?

2. Plot these observations where the x-axis is $\text{time}_{pmm} - \text{time}_{pt}$ and the y-axis is $\text{cost}_{pmm} - \text{cost}_{pt}$.

3. Add to the previous plot the line $-\beta_c \cdot \text{cost}_{pmm} - \beta_t \cdot \text{time}_{pmm} = -\beta_c \cdot \text{cost}_{pt} - \beta_t \cdot \text{time}_{pt}$. What does its slope represent?

---

mbi/ ek/ afa /mpp