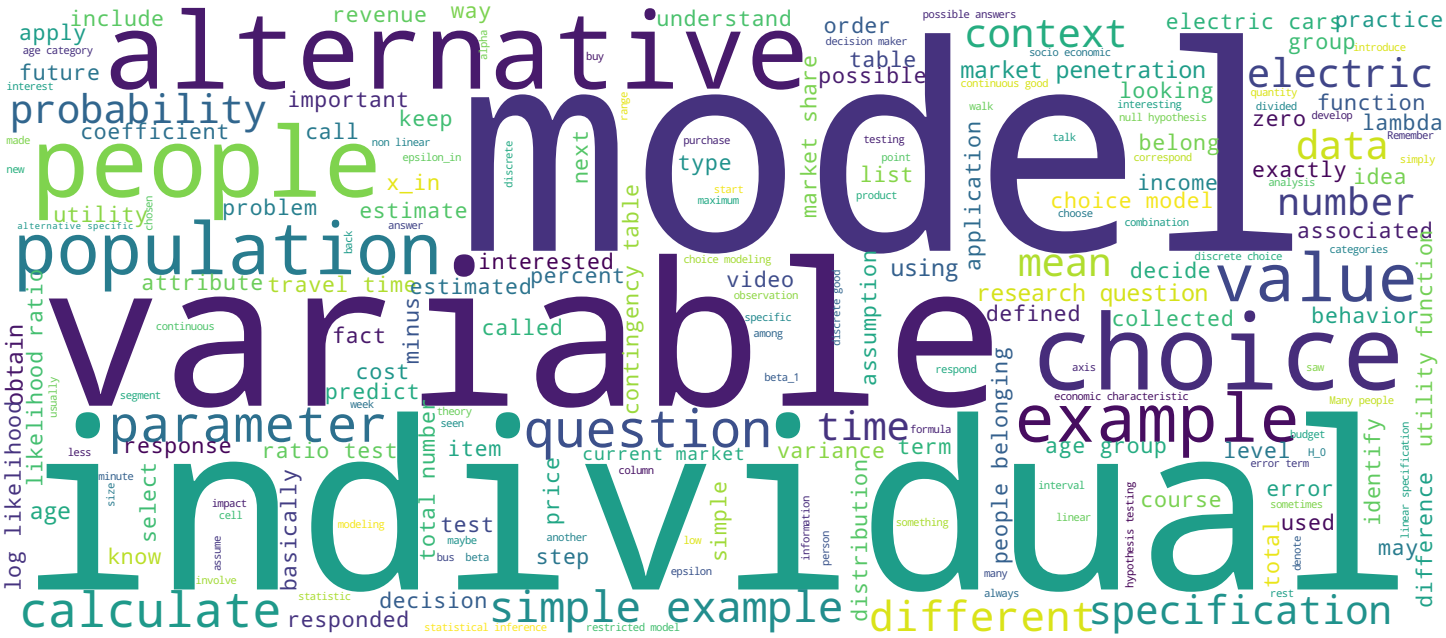
Introduction to behavior modeling

Simple example

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE



EPFL

Simple example: introduction

What I would like to do for the rest of the week is to walk you through a simple example. Actually, a very simple example. But that would involve all the steps that we will be investigating during the whole course. The idea is that this simple example would help you to understand the various issues that are important when we develop a choice model.

Notes

Summary

0m 04s

Simple example

Objectives

Introduce basic components of choice modeling:

- definition of the problem
- data
- model specification
- parameter estimation
- model application

Application

Analysis of the market for electric cars

The objectives of the simple example is to introduce you to the basic components of choice modeling. We will start by defining the problem. We will talk about data, about model specification, parameter estimation, and model application. And the context where we will apply the simple example, is an application where we would like to analyze the market for electric cars.

Notes

Summary

0m 28s

Choice problem

Choice

Consumer's choice to

- own an electric car
- own a car with combustion engine

Research questions

- what is the current market penetration of electric cars relative to combustion engine cars?
- how will the penetration change in the future?

Because we are interested in the market of electric cars, the choice that we are interested in is the choice of each individual to purchase an electric car or not. And the research question that we have is: what is the current market penetration for electric cars. But also, what will be the future market penetration for electric cars. How will this market evolve in the future? And we will see how to address these research questions in the context of this choice problems by going through all the steps of model development.

Notes

Summary

0m 54s

Data

Population

- adults aged 20 and above
- in Switzerland
- owning a car

Sample

- 2500 individuals
- randomly selected

The first thing that we need to do is to observe actual behavior and to collect data. Before we do that, we need to identify the population of interest. It has to be clearly identified. In our simple example, we defined a population of interest as all adults who are age 20 and above, who are living in Switzerland, and who own a car. Now, this population is big. Many people will correspond to this definition. Therefore we cannot interview all of them and we identify a sample of 2500 individuals. These individuals are randomly selected within the full population.

Notes

Summary

1m 31s

Questions

Is your car electric?

- Yes,
- No.

What is your age range?

- 20–39
- 40–64
- 65+

In practice, these people, these individuals who are in the sample, receive a questionnaire, or they are interviewed by somebody, and they have to respond to many questions. In the context of this simple example, we will keep the list of questions very short. We will ask them two questions. The first question is: Is your car electric? Remember, the population is defined as all the owners of a car. So everybody owns a car. So the question is: Is your cat electric? Yes or no. And the second question is about their age: What is your age range? And we define three ranges: 20 to 39, 40 to 64 and 65 and above. Again these questions are designed to keep the example simple.

Notes

Summary

2m 14s

Data

Contingency table

	Age		
	20–39	40–64	65+
Electric	65	55	5
Not electric	835	1045	495

After we have collected the data, we have, for each individual in the sample, the response to each of the two questions. So we know if the car is electric or not, and we know if the age is in one of the three categories that we have defined. In this simple example, we can actually organize all these responses into this table that we call a contingency table. The table has two rows, corresponding to the two possible answers to the first question, and it has three columns, corresponding to the three possible answers of the second question. It means that each single individual in the sample is associated with exactly one cell of this table, depending on the answer he gave to these questions. So therefore we can count how many individuals correspond to each combination of responses. So, for instance, there are five individuals who responded that they own an electric car and they have an age which is 65 or above. So this contingency table contains all the information we have collected from the data. No we can make some analysis of the table. For example, we can calculate how many people in the sample belong to each of the age group.

Notes

Summary

3m 02s

Data

Contingency table

	Age			
	20–39	40–64	65+	
Electric	65	55	5	125
Not electric	835	1045	495	2375
	900	1100	500	2500

125 / 2500 = 5 %

We have 900 people belonging to the age group 22 to 39, 1100 people belonging to the age group 40 to 64, and 500 people belonging to the age group 65 and above. We can also calculate the row totals. If we calculate the number of people who responded that they own an electric car, we obtain 125. And if we calculate the total number of people who responded that they do not own an electric car, we obtain 2375. And of course if you make the sum of these two numbers, you obtain 2500 which is the total number of people in the sample. And obviously you will obtain the same total if you sum these three columns totals here. So this is really interesting because we can actually respond to the first research question by looking at this table. The first research question was: What is the current market penetration in the population? Well, we can immediately see what is the market penetration in the sample. We have 125 people who responded that they own an electric car, among a total of 2500. So we have a total of 5 percent of the people who responded that they have an electric car. So this is the market share in the sample.

Notes

Summary

4m 29s

Data

Contingency table

	Age		
	20–39	40–64	65+
Electric	65	55	5
Not electric	835	1045	495
	900	1100	500

125 / 2500 = 5 %

2375

2500

Now, what we will do is to use statistical inference because we have selected the people completely randomly in the population, we will infer that this percentage that we observe in the sample, is the same in the population. So using this contingency table and statistical inference, we give the answer to the first research question: 5 percent is the market penetration of electric cars now, currently. Good. But now, how do we predict? The second research question is: how will this market penetration evolve in the future? Well, in order to predict, we need a model. We cannot use data only. We need a model. And this is what we will do next.

Notes

Summary

5m 56s