15.774/15.780 Fall 2021

Analytics of Operations Management

Problem Set 3 - Choice Modeling & Assortment Optimization

Due date: October 8, 2020

Instructions:

1. Submit a PDF file with your solutions to Canvas before the assigned deadline. Write your name and MIT ID on your submission.
2. All plots must have clear and easy-to-read axis labels and legends,
3. **Include relevant code in the PDF submission even if the question doesn't explicitly ask for it**. This means we can give you partial credit even if the output is wrong. if appropriate.

Problem 1. Choice Modeling (40 pts)

In this problem, we want to build a model to represent a choice between the usage of two transportation systems (bus or car). We use the data stored in `commuteFall2020.csv`, which contains the choices made by 400 commuters. The data is stored in *wide* format, and contains variables:

- `Car.Travel` travel time by car (in minutes);
- `HHIncome` is annual household income (in dollars);
- `ParkingCost` is the car parking cost (in dollars);
- `BusFare` is bus fare in dollars;
- `Bus.Travel` is the travel time by the bus (in minutes);
- `Bus.Wait` is the average waiting time for the bus (in minutes).
- `Choice` whehter the commuter choise `CAR` or `BUS`

We assume that for each person, the utility of taking either transportation method is **linear** in **total journey time** (travel time for a car, or *travel plus waiting time* for a bus), **daily cost** (parking cost for car or bus fare), and **household income**. We will assume that there is no outside option, i.e. a person *must* choose either car or bus.

Therefore we want our `mlogit` model to use the formula `Choice ~Time + Cost | Household - 1`. Recall from Recitation 4 what this means: (i) On the left of the tilde, `Choice` is the categorical choice we want to model (ii) Before the `'|'` we have things that are features of the alternatives (i.e. products) (iii) To the right of the `'|'` we have things that are features of the person making the choice (iv) The -1 indicates to R that we don't want intercepts.

1. We need to do some pre-processing to get the data in the right format for `mlogit`. You will need to create some new columns and name them correctly. Show your code for doing this. (5pts)
2. Fit an MNL model to the data and show the model coefficients. (5pts)
3. Table 1 shows the characteristics of commutes for two subpopulations. Think of subpopulation L as people living near the center of a city and subpopulation H as people living in the suburbs. People in both subpopulations all work in the downtown area. We assume that the same MNL model applies to both subpopulations. Based on the information in Table 1, compute the probabilities that an individual belonging to each subpopulation will choose to take a car to work. Show what equations you used. Hint: You should provide 2 numeric answers for this question. (10 pts)

Table 1: Characteristics of commute for two subpopulations

	Segment L		Segment H	
Variable	Car	Bus	Car	Bus
Car travel time (min)	25	-	40	-
Parking cost ($)	3.00	-	8.00	-
Bus travel time (min)	-	40	-	60
Bus waiting time (min)	-	5	-	10
Bus fare ($)	-	0.80	-	2.00
Household annual income ($)	40,000		80,000	

4. In order to increase bus ridership, the local transport authority has considered the following options:

   (a) Reducing the bus fare by 50 cents in both subpopulations.

   (b) Increasing the number of buses so that the waiting times would be cut in half for both subpopulations.

   (c) Doubling parking costs.

   For each subpopulation, how will each of these strategies (a, b, c above) affect the **probability of taking the bus**, and which option will be most effective in increasing bus ridership? (20 pts)

Problem 2. (Assortment Optimization) (40 pts)

For this question we will consider a set of mobile phones to be released sometime in the future. The phones are Galaxy S99, Galaxy S100, iPhone 99 and iPhone 100. We will assume that a MNL choice model has be estimated based on choice data from consumer surveys. The weights of the choice model are shown in Table 2 (note that these are w's, i.e. you don't need to exponentiate them). Also shown in the table are the revenues for each phone.

Table 2: MNL weights and revenue for different mobile phones.

Phone	w	Revenue [USD]
Galaxy S99	2.0	400
Galaxy S100	2.2	499
iPhone 99	2.1	500
iPhone 100	2.5	600

1. First let's find the optimal assortment. Assume there is always a no purchase option with weight $w_0 = 1$. Use the data in Table 2. Calculate the revenue of the revenue-ordered (RO) assortments

for assortment sizes of one, two, three, and four phones. What is the assortment which maximizes revenue? (10 pts)

2. It turns out the MNL model is totally wrong here. In fact, there are two types of customers. One type will NEVER buy iPhones, and the other will NEVER buy Galaxy. A customer is a "NEVER iPhone" with 50% probability, and a customer is "NEVER Galaxy" with 50% probability. Consider the MNL model for each customer type given in Table 3.

Table 3: MNL weights and revenue for different mobile phones for each customer type.

Phone	NEVER Galaxy w	NEVER iPhone w	Revenue [USD]
Galaxy S99	0.1	2.0	400
Galaxy S100	0.1	2.2	499
iPhone 99	2.1	0.01	500
iPhone 100	2.5	0.01	600

Include a "No purchase" option with $w_0 = 1$ for both subpopulations. Calculate the expected revenue of the optimal assortment from (1), but using the numeric values in Table 3. Remember, a customer belongs to either type with 50% probability. What is the difference between this expected revenue and the expected revenue from part (1)? (5 pts)

3. Now imagine that you use information about the customer to figure out their type and you show each type a different assortment. Assume each type follows an MNL model using the weights in **Table 3**. Assume there is always a no purchase option with weight $w_0 = 1$ and use the RO assortment technique you learned in class.

   (a) What is the optimal assortment for each customer type? (20 pts)

   (b) What is expected revenue when you show each type their own optimal assortment? (You can just average the revenue for each type.) (5 pts)