November 25, 2025

1 Questions discussed in class

Here are the solutions discussed during the interactive session. The correct answer is in bold.

Question 1: Scale

Why is the scale of the logit model not identified?

1. Because it is normalized to 1. No. It is the consequence, not the cause.
2. Because it does not matter. No. The scale does matter.
3. **Because only the order of utility matters, not the value.** Indeed. The choice probability is given by

$$P_n(i|\mathcal{C}_n) = \Pr(U_{in} \geq U_{jn} \forall j \in \mathcal{C}_n),$$

and is not affected by a change of scale:

$$P_n(i|\mathcal{C}_n) = \Pr(\mu U_{in} \geq \mu U_{jn} \forall j \in \mathcal{C}_n),$$

for any $\mu > 0$. Note that it is true not only for logit, but for any random utility model.

4. I don't know.

Question 2: generic vs alternative specific

We consider a mode choice model with two variables: travel cost and travel time. For each of these variables, should the coefficient be generic or alternative specific, based on behavioral considerations? More specifically, which one of the following models is the most behaviorally meaningful?

1. Model 1: both parameters are generic:

$$U_{in} = \beta_c \text{cost}_{in} + \beta_t \text{time}_{in},$$
$$U_{jn} = \beta_c \text{cost}_{jn} + \beta_t \text{time}_{jn},$$

   that involves two unknown parameters.

2. Model 2: the cost parameter is generic and the time parameter is alternative specific:

$$U_{in} = \beta_c \text{cost}_{in} + \beta_{ti} \text{time}_{in},$$
$$U_{jn} = \beta_c \text{cost}_{jn} + \beta_{tj} \text{time}_{jn},$$

   that involves three unknown parameters.

3. Model 3: the time parameter is generic and the cost parameter is alternative specific:

$$U_{in} = \beta_{ci} \text{cost}_{in} + \beta_t \text{time}_{in},$$
$$U_{jn} = \beta_{cj} \text{cost}_{jn} + \beta_t \text{time}_{jn},$$

   that involves three unknown parameters.

4. Model 4: both parameters are alternative specific:

$$U_{in} = \beta_{ci} \text{cost}_{in} + \beta_{ti} \text{time}_{in},$$
$$U_{jn} = \beta_{cj} \text{cost}_{jn} + \beta_{tj} \text{time}_{jn},$$

   that involves four unknown parameters.

5. **All models are equally behaviorally meaningful.**

6. None of them is behaviorally meaningful.

All are behaviorally meaningful. If time (cost, resp.) is considered as a resource, one minute is one minute, irrespectively how you spend it. Therefore, the time (cost, resp.) coefficient should be generic. But a behavioral argument based on a different perception of time in one alternative (the car, say), or the other (the bus, say) may also support an alternative specific specification. In that case, we ackowledge the fact that one minute spent in the car may not be perceived in the same way as one minute spent in the bus.
Therefore, each potential specification is equally meaningful from a behavioral point of view. Hypothesis testing with real data will help selecting the most appropriate specification.

Question 3: Segmentation

Consider a model involving only one variable (travel time, say). And there is a time coefficient for males and one for females. We have a sample of 200 males and 200 females. The estimates are $\beta_m = -0.123$ and $\beta_f = -0.096$. We collect more data from another 100 females and re-estimate the same model with the sample of 500 individuals. Will the parameters have the exact same value or not? Which one of the following cases are you expecting to happen?

1. β_m same value (-0.123), β_f same value (-0.096),
2. β_m same value (-0.123), β_f different value, (this is the correct answer)
3. β_m different value, β_f same value (-0.096),
4. β_m different value, β_f different value.

We consider two distinct populations: males and females, where the parameters for each population are estimated independently. During the second wave of estimation, the data available for estimating the coefficient for males are identical to those used in the first wave. As a result, the estimator for males remains unchanged.

In contrast, for females, additional data are available during the second wave compared to the first. Consequently, it is highly unlikely that the estimated value will be exactly the same, as the increased amount of data generally improves the precision of the estimator.

However, this reasoning no longer applies if the model includes parameters shared between males and females.

Question 4: t-test

Consider a transportation mode choice model. The utility of the public transportation alternative is specified as

$$\beta_t \text{travelTime} + \beta_c \text{travelCost} + \beta_h \text{headway}.$$

The results of the estimation are

	Estimates	t-test
β_t	-0.01	0.98
β_c	-0.2	-1.95
β_h	-0.15	2.06

The specification that the analyst should consider is

1. $\beta_t$travelTime + $\beta_c$travelCost + $\beta_h$headway,
2. $\beta_c$travelCost + $\beta_h$headway,
3. $\beta_h$headway.

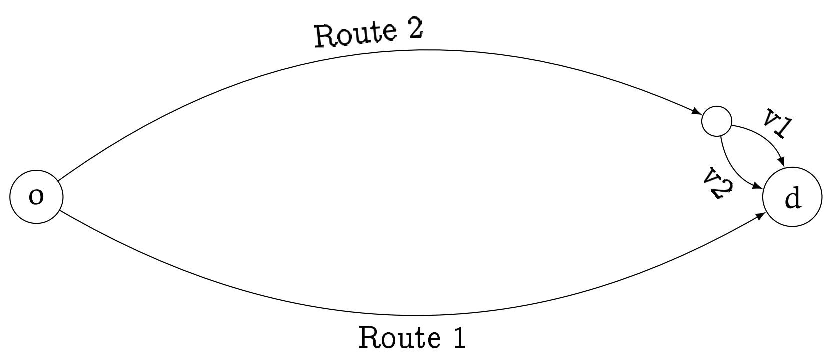
The correct answer is the first one. Indeed, these three variables are key variables, and we know that they are indeed influencing the choice. The hypothesis that $\beta_t = 0$ does not make sense, and should not even be tested. The fact that the t-test is low is probably due to a lack of variability in the data. Instead of removing the variables, more data must be collected to improve the precision of the estimates.

2 Additional questions

Route choice

We consider two routes linking an origin and a destination. The two routes have exactly the same travel time T. The second route includes two variants for a small portion of the itinerary. We consider a logit model with three alternatives (route 1, route 2 variant 1, and route 2 variant 2) where travel time is the only explanatory variable. What is the probability predicted by the model for a given individual to choose route 1?

1. $\approx 1/2$,
2. $\approx$ **1/3**,
3. $\approx 1/4$,
4. ≈ 0.



Intuitively, we expect each of the two routes to have roughly 50% probability to be chosen. If we apply the logit model, we obtain

- utility of route 1: βT,
- utility of route 2, variant 1: βT,
- utility of route 2, variant 2: βT.

Therefore,

$$P_n(1) = \frac{e^{\beta T}}{e^{\beta T} + e^{\beta T} + e^{\beta T}} = \frac{1}{3}.$$

The reason why we do not obtain the intuitive value of 1/2 is due to the assumption of independence of the error terms associated with the derivation of the logit model. Indeed, the two variants of route 2 share all the unobserved variables associated with route 2. Therefore, the error terms are certainly not independent, in this example.

Captivity

Consider a binary model choice model between car (i) and train (j) for commute: $P_n(i|\{i,j\})$. Individuals without a driving license are said to be captive, as they have no choice. They have to take the train. The analyst does not have information about the possession of a driving license. But she knows the age of the respondents, and she knows that, if an individual is under 24, the probability to have a driving license is 45%. Using the model, how can the analyst calculate the probability for such an individual to use the car?

We have to decompose the model into each possible scenarios:

$$\begin{aligned}\mathrm{Pr}(\mathrm{car}) =& \mathrm{Pr}(\mathrm{car}|\mathrm{license})\,\mathrm{Pr}(\mathrm{license}) \\ &+ \mathrm{Pr}(\mathrm{car}|\mathrm{no\ license})\,\mathrm{Pr}(\mathrm{no\ license}) \\ =& \mathrm{Pr}(\mathrm{car}|\mathrm{license})0.45 + 0 \cdot (1 - 0.45) \\ =& 0.45\,\mathrm{Pr}(\mathrm{car}|\mathrm{license}).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{Pr}(\mathrm{bus}) =& \mathrm{Pr}(\mathrm{bus}|\mathrm{license})\,\mathrm{Pr}(\mathrm{license}) \\ &+ \mathrm{Pr}(\mathrm{bus}|\mathrm{no\ license})\,\mathrm{Pr}(\mathrm{no\ license}) \\ =& \mathrm{Pr}(\mathrm{bus}|\mathrm{license})0.45 + 1 \cdot (1 - 0.45) \\ =& 0.45\,\mathrm{Pr}(\mathrm{bus}|\mathrm{license}) + 0.55.\end{aligned}$$

Such a model is sometimes called a latent class model, as the class of the individuals is not observed.

Interaction

Consider a model with three alternatives, that includes two alternative specific constants:

$$\begin{aligned} V_{1n} &= \cdots + c_1 \\ V_{2n} &= \cdots + c_2 \\ V_{3n} &= \cdots \end{aligned}$$

We consider two segments in the population based on age. Consider the indicator "young" that is 1 if individual n belongs to the segment of young people, and 0 otherwise. The indicator "old" is similarly defined. Note that, for each individual, exactly one of the two indicators is 1. Which specification should be used to capture the interaction of age with the constant?

1. $$\begin{aligned} V_1 &= \cdots + c_1^y \text{ young} + c_1^o \text{ old} \\ V_2 &= \cdots + c_2^y \text{ young} + c_2^o \text{ old} \\ V_3 &= \cdots \end{aligned}$$

2. $$\begin{aligned} V_1 &= \cdots + c_1 + c_1^o \text{ old} \\ V_2 &= \cdots + c_2 + c_2^o \text{ old} \\ V_3 &= \cdots \end{aligned}$$

3. $$\begin{aligned} V_1 &= \cdots + c_1^o \text{ old} \\ V_2 &= \cdots + c_2^o \text{ old} \\ V_3 &= \cdots \end{aligned}$$

4. $$\begin{aligned} V_1 &= \cdots + c_1^y \text{ young} \\ V_2 &= \cdots + c_2^o \text{ old} \\ V_3 &= \cdots \end{aligned}$$

Both specifications 1 and 2 are valid, and equivalent. In the first specification, we associate a different set of constants with each segment. As there are two segments and two constants, it means 4 parameters. The second specification involves also 4 parameters. In that case, we consider the "young" segment as the reference. It means that c_1 and c_2 are the constants for young people. The parameters c_1^o and c_2^o capture the **difference** between the constants of the "old" segment and the constants of the reference segment. Therefore, the constants for old people are $c_1 + c_1^o$ and $c_2 + c_2^o$.

Tests

To test that the true value of one parameter is 1, I can use

1. a t-test,
2. a likelihood ratio test,
3. a Cox test,
4. a Davidson-McKinnon J test,
5. the adjusted likelihood ratio index,
6. none of them,
7. any of them.

Any of these tests can be applied. However, they are not necessarily equivalent. It means that it may happen that two tests generate contradictory outcomes.

Prediction

It is common practice to predict that the chosen alternative is the one associated with the highest probability. Why is it not valid?
This would be equivalent to assume that the choice is associated with the highest deterministic utility. Therefore, it amounts to ignore the error term. In terms of prediction, it may lead to significant errors.
Consider a case where we want to predict the choice of 1000 individuals between two alternatives. The true model, that generated the data, associates a probability of 51% with alternative 1, and of 49% with alternative 2, for all individuals. Assume that the analyst is able to recover the true model. Using the highest probability method, the analyst will predict that all individuals will select alternative 1, while, in reality, only 510 individuals will choose it.
When the number of alternatives increases, the error may be more severe. Assume that we have N individuals and J alternatives. The true model associates with alternative 1 probability

$$P(1) = \frac{1+\delta}{J},$$

and with each other alternative i,

$$P(i) = \frac{J-1-\delta}{J(J-1)} = \frac{1}{J}\left(1 - \frac{\delta}{J-1}\right),$$

where $0 < \delta < J-1$ is a small number. It is a valid probability function as $0 \leq P(i) \leq 1$ and

$$\sum_i P(i) = \frac{1+\delta}{J} + (J-1)\frac{J-1-\delta}{J(J-1)} = 1.$$

As $P(1) > P(i)$ for all $i \neq 1$, the use of the highest probability method will again predict that N individuals will select alternative 1. In reality, the number is

$$N\frac{1+\delta}{J},$$

so that the error is

$$N(1 - \frac{1+\delta}{J}),$$

and it increases with N and J.

Projecting a probability to zero or to one is sometimes inevitable, but should be done at the latest possible moment, as precious information is lost. In particular, it should not be done before aggregation.

Prediction with SP data

Stated preferences data cannot be used for forecasting market shares. Why?
Forecasting market shares can be done using sample enumeration. A sample of individuals, representative of the population, is selected. For each of them, the value of all explanatory variables x is observed, and the choice model is used to predict their choice. Then aggregation is performed to obtain market shares.
It is actually sometimes possible to use the same sample that was used for estimation for that purpose. But two conditions must be respected. First, the sample must be weighted to correct for possible under- or oversampling of some strata in the population. This is not always necessary for estimation (as we will see in the lecture on sampling), but it is always necessary for prediction. Second, the sample must represent a real market situation. This implies that only Revealed Preferences (RP) data can be used. Stated Preferences (SP) data cannot be used for forecasting market shares.
In SP data, each individual responds to several hypothetical choice situations (like in the Swissmetro data set). The values of the x are "engineered" in order to maximize the statistical significance of the parameters. They do not represent any real market situation.
Consider a situation where the population of interest is all the travelers from Lausanne to Zurich. In the real market, the travel time by train is the same for all the travelers. Therefore, RP data will not allow the analyst to estimate the coefficient for travel time by train, as there will be no variability in the data. She must rely on SP data for estimation. But when predicting market shares, the actual travel time between Lausanne and Zurich must be used. The values of travel time in the SP data are the results of the experimental design and cannot be used for that purpose.