Mathematical Modeling of Behavior 2024

Michel Bierlaire

December 17, 2024

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

Question 1: 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. $\beta_t \text{travelTime} + \beta_c \text{travelCost} + \beta_h \text{headway}$,
2. $\beta_c \text{travelCost} + \beta_h \text{headway}$,
3. $\beta_h \text{headway}$.

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.

Question 2: 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 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?

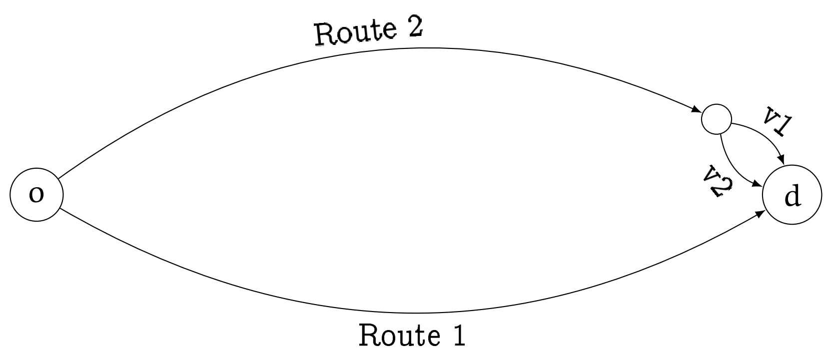
1. β_m same value (-0.123), β_f same value (-0.096),
2. β_m same value (-0.123), β_f different value, (this is the correct answer)
3. β_m different value, β_f same value (-0.096),
4. β_m different value, β_f different value.

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: 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. $\approx \mathbf{1/3}$,
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,

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

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.

Question 5: 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:

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

Similarly,

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

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

Question 6: 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 $c_2 + c_2^o$.