



COMMENTS ON THE FIRT ASSIGNMENT

In this document you will find some general comments about the most common mistakes that we noticed after going through the first assignments you gave in and the peer-review session. If you have any questions or comments about it you can send an email to anna.fernandezantolin@epfl.ch.

1. When developing a model you should make a hypothesis and then see if the results that you obtain match your intuition, *not the opposite!* Modeling is **not** about including variables that don't make sense, obtaining results and then trying to justify them. Remember that you are assuming utility maximization under budget and time constraints and that transport is considered as a normal consumption good: demands for it are necessarily downward sloping functions of prices and times!
2. As a first modeling attempt, put attributes of alternatives in their corresponding utility function, i.e., assume that demands are functions of only their own attributes.
3. As you are modeling preferences, think about interacting individual specific variables with attributes: different people have different sensitivities to attributes (old and young people may have different perception of travel time).
4. Think about presenting first overall goodness-of-fit of your model by comparing it to specifications without any variable and with intercept terms only.
5. You should be careful where you include the variables that you want to test. E.g. if you want to test the effect of waiting time of the PT alternative, you should include it in the PT utility and not in the CAR utility!
6. It is good to have attributes of both alternatives. In this way, your model will be useful for forecasting.
7. Avoid sentences like: "The rho squared/likelihood has a value of X so our model is good". These values are meaningless by themselves. You need to compare them to those of other models. In particular, for the likelihood you can't compare the values just by looking at them, you need to do the likelihood ratio test.
8. Avoid excluding respondents whenever you can. It is not needed to exclude respondents that did not answer about income if income is not in your model. If you do it, you lose information for no reason, as you throw away observations that include information about the rest of the variables that you have in your model.
9. For numerical reasons, it is good practice to scale the data so that the values of the estimates are close to one.

10. The attitude questions (EnvirXX, MobilXX, ResidChXX, LifStyXX) are indicators and cannot be included directly in the model. They are supposed to be used with the latent variable approach that you will learn in the last part of the course.
11. When interpreting dummies be careful on how you comment on them. Just for illustration purposes, assume you have the following utility functions:

$$U_1 = ASC_1 + \beta_{time}time_1 + \beta_{cost}cost_1 + \beta_{male}male$$

$$U_2 = \beta_{time}time_2 + \beta_{cost}cost_2$$

With $ASC_1 = -2$ and $\beta_{male} = 1$. If time and cost are equal for both alternatives the values for the utilities will be the following (time and cost will cancel out):

- **Male:** $U_1 - U_2 = -2 + 1 = -1 < 0 \Rightarrow U_1 < U_2$
- **Female:** $U_1 - U_2 = -2 = -2 < 0 \Rightarrow U_1 < U_2$

The utility of alternative 2 is higher for both men and women. However, since β_{male} is positive, you can conclude that men have a higher utility towards alternative 1 *compared to women*. If you forget the "compared to women" it is not clear if it is compared to alternative 2 or compared to women, and it is not true that utility of alternative 1 is higher compared to alternative 2, as just shown.