

**Home Assignment 4: Discrete Choice Models**Due until Jan. X<sup>th</sup> 2020 (24hours)

## 1 Objectives

The objective of this home assignment is to evaluate your understanding of logit discrete choice models applied to choices of more than two alternatives. Two different models were taught in the module of the TDM course. The two have their own special characteristics which you should be able to recognize and apply in this home assignment.

## 2 Software

The reference software to use for developing this work will be Python using Biogeme.

## 3 Data

The data that you will use is the result of the London Travel Demand Survey which consisted of a household travel survey collected by the Transport for London. The objective was to be able to predict how people react to different network conditions to allow for a more efficient management and investment planning of the transportation network. Details about the study's framework and dataset are referred in [1].

The database contains details about 81 086 trips made by 31 954 individuals across 17 616 households. Each observation contains details about the socio-demographics of each person and information about different transportation modes. Each respondent answered their preferred mode from the 4 alternatives (walking, cycling, public transportation and driving).

This dataset is available as a .dat file. The variable names, description and coding is presented in the next table:

Table 1: Data variables.

Attribute	Description
trip_id	Unique ID for each trip
household_id	Unique ID for the household
person_n	ID of the person

*Continued on next page*

Table 1 – *Continued from previous page*

<b>Attribute</b>	<b>Description</b>
trip_n	ID of the trip
travel_mode	Mode of travel chosen by LTDS trip (1: walk, 2: cycle, 3: public transport, 4:drive)
purpose	Journey purpose of the trip (1: home-based work, 2: home-based education, 3: home-based other, 4: employers' business, 5: non-home-based other)
fueltype	Fuel type of the proposed vehicle for the trip (1: petrol car, 2: diesel car, 3: hybrid car, 4: petrol LGV, 5: diesel LGV, 6: average car)
faretype	Public transportation fare type (: full-fare, 2: 16+, 3: child, 4: disabled, 5: free)
bus_scale	Bus fare scale of person making trip (0: free bus journeys, 0.5: half price, 1: full price)
survey_year	Survey year: (1: 2012/13, 2: 2013/14 or 3: 2014/15)
travel_year	Trip travel year, from 2012 to 2015
travel_month	Month of year of travel (from 1 -January- to 12 -December)
travel_day	Day of the month of travel (from 1 to 31)
day_of_week	Day of the week of travel (from 1 -Monday- to 7 -Sunday)
start_time	Linearized start time of the trip
age	Age of the traveller
female	Gender of the traveller (1 female, 0 otherwise)
driving_license	Whether the traveller has a driving license (1 yes, 0 otherwise)
car_ownership	Car ownership of the household (0: no cars in household, 1: less than one car per adult, 2: one or more cars per adult)
distance	Straight line trip distance
dur_walking	Duration of walking route in hours
dur_cycling	Duration of cycling route in hours
dur_pt_access	Duration walking to/from first/last stop on the public transport route in hours
dur_pt_rail	Duration spent on rail services on public transportation in hours

*Continued on next page*

Table 1 – *Continued from previous page*

Attribute	Description
dur_pt_bus	Duration spent on bus services on public transportation in hours
dur_pt_int	Total duration of public transport interchanges interchanges in hours
pt_interchanges	Total number of public transport interchanges (rail–rail, bus–bus, bus–rail and rail–bus)
dur_driving	Duration of the driving route in hours
cost_transit	Estimated cost of the whole public transport route in GBP
cost_driving_fuel	Estimated fuel cost for driving route in GBP
cost_driving_ccharges	Estimated congestion charge cost of driving route in GBP
driving_traffic_percent	Predicted traffic variability on driving route

## 4 Your tasks

- 1) Formulate hypothesis on the relationship between the socio-demographic variables of the respondent and their modelled choice of transportation mode behaviour. Confirm or not these hypotheses through the models' results.
- 2) Determine different utility specifications for the 4 alternatives in a Multinomial Logit (MNL) model discrete choice structure. You are encouraged to build new variables using the ones which were given and test them too. You are also encouraged to try multiple utility specifications to compare different models.
- 3) Measure the goodness of fit of your best model and comment on those values. Produce willingness to pay (WTP) estimations which are meaningful and useful for transportation planning practitioners. Explore different elasticities in the produced models.
- 4) Propose and test Nested Logit (NL) model tree structures for the previous utility functions specifications. Choose one and complete the model adding again explanatory variables to the utility specification of each alternative and to the utility of each branch. Comment on the results of your experiments and compare them to the MNL outputs.

## 5 Report

Your final report should include among other things:

- 1) Description and discussion of the factors that affect the choice of different transportation modes, based on the database provided;

- 2) Present the models obtained, respective statistical tests, goodness of fit and other relevant indicators;
- 3) Discussion about the models and statistical tests validation;
- 4) Comparison between MNL and NL, and discussion about the possible limitations of the models;
- 5) A discussion of the similarities and differences between the causal inferences from your “best” specification and your *a priori* considerations.

The report should be printed in .pdf format (no word files) and sent through email (fmoura@tecnico.ulisboa.pt) until the deadline referred above.

## 6 Final Comments

- 1) The following criteria will be applied for grading:
  - a) Your understanding of the problem (e.g. causal relationships);
  - b) Your understanding of the different Logit models for multiple choices;
  - c) Your understanding of statistical testing (explain what the statistics mean).
- 2) Remember that you must always examine and comment on your results. Computer outputs without explanations are not acceptable.
- 3) There are no strict formatting rules except that you should write a concise report that is not longer than 15 pages (approx.), or around 7 500 words without appendixes.

**Final Note:** There is no fully right or fully wrong model. Your objective is to take meaningful conclusions on the behaviour of decision makers regarding their mode choice.

The task of finding the perfect model can be very time consuming and in the end we are never sure of the outcome, so keep in mind that it is possible to give the same grade to two different models which have focused more on some aspects than others. Overall your understanding of the subjects is what counts most.

## References

- [1] T. Hillel, M. Z. E. B. Elshafie, and Y. Jin, “Recreating passenger mode choice-sets for transport simulation: A case study of london, uk,” *Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction*, vol. 171, no. 1, pp. 29–42, 2018. [Online]. Available: <https://doi.org/10.1680/jsmic.17.00018>