Part A: MNL Model Estimation (70 points)

Instruction: Review the variable description for the data file MODEBIN3.csv, which is attached. This data file is available through Laulima. Use R/RStudio for the remainder of this section. Attach your final R script to your completed homework submission.

Include a table presenting the estimation results for all models tested in this assignment.

Initial Estimation: First, estimate a base market share model, without weights. What is the market share of each travel mode? Interpret each estimated parameter.

Include LOS Attributes: Next estimate a model with alternative specific constants and IVTT, OVTT and TOTCOST. Report and interpret the estimation results. Compute the value of time for in-vehicle and out-of-vehicle travel time. Conduct a statistical test comparing this model with the base market share model.

Estimation and Interpretation of other Model Specifications:

A) Introduce the level-of-service variables in different ways (keep the level-of-service variables generic in all specifications) and estimate the corresponding models. Report and interpret your estimation results for each model you tried in a table format that will allow for easy comparison.

   Conduct statistical tests among the models and select the preferred specification (explicitly state which pair of models are being compared in each test and how you arrived at the final specification).

B) Include variables representing the impact of automobile availability on mode choice to your final specification from part a. Try two different forms: 1) number of vehicles and 2) number of vehicles per worker. Which of these functional forms would you prefer to represent the effect of auto-availability? Give theoretical and statistical arguments to support your conclusion.

C) Use the preferred model from part B as the base. Explore additional specifications using other variables in the dataset to improve the overall goodness of fit consistent with theoretical notions. Report your results including reasons for trying different specifications, the results of your tests, and your recommendation for a preferred model.

D) Compute the values of in-vehicle travel time and out-of-vehicle travel time for your best overall specification. Also compute the marginal and elasticity effects of a change in each level-of-service variable for each mode on the mode choice probabilities (use mean values of variables for this computation). Present the results in the form of a matrix.

Aggregate Effect on Mode shares Due to Level-of-Service Changes: Consider a proposal to designate high-occupancy vehicle lanes during the peak period. It is estimated that this proposal would reduce transit and shared-ride in-vehicle time by 25% and, at the same time, increase drive-alone in-vehicle time by 30% (for each traveler in the Boston metropolitan area). Use your estimated model to determine the impact of such a transit-use incentive/auto-use disincentive policy on overall travel mode shares by transit, shared-ride, and drive-alone.

Part B: Nested Logit Estimation (30 points)

Preliminary Nested Logit Estimation: First, re-estimate nested logit models for the three nesting structures we covered in class. Interpret the logsum parameters; be as complete as you can.

Estimation and Interpretation of other Model Specifications: Introduce other variables and interpret your estimation results for each model you tried in a table format that will allow for easy comparison. Conduct statistical tests among the models and select the preferred.

List of Variable in Data Set

PERSID: Person Identification Number
UNO: 1 for all observations
IVTT: In-Vehicle Travel Time (minutes)
OVTT: Out-of-Vehicle Travel Time (minutes)
TOTTIME: Total Travel Time (minutes)
TOTCOST: Total Travel Cost ($)
TDUMMY: 1 if use of transit entails one or more transfers, 0 otherwise
CHOSEN: Mode Chosen
HHSIZE: Household Size
MALE: 1 if individual is male, 0 otherwise
NUMVEH: Number of vehicles in worker's household
WORKERS: Number of workers in worker's household
VEHWORK: Vehicles per worker in worker's household
NEWINC: Household income ($/year)
AEMPDENS: Employment density at worker's employment place
POPDENS: Population density at worker's residence
DIST: Travel distance to work (miles)
AGE: Age of traveler
WEIGHT: Weight to be given to the observation.
UNOTRN: 1 if record represents transit alternative, 0 otherwise
UNOSHR: