

# Problem Set 1

Topics in Advanced Econometrics (ResEcon 703)  
University of Massachusetts Amherst

**Due: September 24, 10:00 am ET**

## Rules

Email a single .pdf file of your problem set writeup, code, and output to `mwoerman@umass.edu` by the date and time above. You may work in groups of up to three, and all group members can submit the same code and output; indicate in your writeup who you worked with. You must submit a unique writeup that answers the problems below. You can discuss answers with your fellow group members, but your writeup must be in your own words. You can use any “canned” routine (e.g., `lm()`, `glm()`, and `mlogit()`) for this problem set.

## Data

Download the file `travel_datasets.zip` from the course website ([github.com/woerman/ResEcon703](https://github.com/woerman/ResEcon703)). This zipped file contains two datasets, `travel_binary.csv` and `travel_multinomial.csv`, that you will use for this problem set. Both datasets contain simulated data on the travel mode choice of 1000 UMass graduate students commuting to campus. The `travel_binary.csv` dataset corresponds to commuting in the middle of winter when only driving a car or taking a bus are feasible options (assume the weather is too severe for even the heartiest graduate students to ride a bike or walk). The `travel_multinomial.csv` dataset corresponds to commuting in spring when riding a bike and walking are feasible alternatives. See the file `travel_descriptions.txt` for descriptions of the variables in each dataset.

## Problem 1: Linear Probability Model

Use the `travel_binary.csv` dataset for this question.

- Model the choice to drive to campus during winter as a linear probability model. Include only alternative-specific data (and not demographic data) in this model. Report the estimated coefficients and heteroskedastic-robust standard errors from this model. Interpret these results, including how a marginal change in each variable would affect the probability of driving.
- Do any of your coefficients look like they might be equal (in absolute value)? Test that the two time coefficients are equal (in absolute value) and then test that the two cost coefficients are equal (in absolute value). Interpret the results of these tests and explain why they make intuitive sense. (Hint: There are many ways to conduct a Wald test in R. I like the `linearHypothesis()` function from the `car` (companion to applied regression) package.)

- c. Again model the choice to drive to campus during winter as a linear probability model, but now force the two cost variables to have the same coefficient (in absolute value). Report the estimated coefficients and heteroskedastic-robust standard errors from this model. Interpret these results, including how a marginal change in each variable would affect the probability of driving. (Hint: An easy way to force coefficients to be equal is to construct a new variable from the existing data.)
- d. One potential problem with a linear probability model is that predicted probabilities can fall outside the  $[0, 1]$  range. Using the model from part (c), how many graduate students have infeasible choice probabilities? Given these results, are you worried about using a linear probability model in this case?

## Problem 2: Binary Logit Model

Use the `travel_binary.csv` dataset for this question.

- a. Model the choice to drive to campus during winter as a binary logit model with the same variables as in part (c) of problem 1. Report the estimated coefficients and standard errors from this model. Interpret these results, including how a marginal change in each variable would affect the probability of driving, and compare to the linear probability model you estimated in part (c) of problem 1. Also calculate the dollar value of an hour spent on each travel mode and explain these time-value results.
- b. Demographic information might also affect a travel mode decision. For example, individuals with different incomes might have different sensitivities to cost. Again model the choice to drive to campus during winter as a binary logit model, but now divide cost by income to allow for this kind of heterogeneity. Report the estimated coefficients and standard errors from this model. Interpret these results, including how a marginal change in each time or cost variable would affect the probability of driving. Also calculate the dollar value of an hour spent on each travel mode, which now varies by income; calculate these values for incomes of \$15,000, \$25,000, and \$35,000.
- c. The family status of a graduate student might also affect their decision making. Estimate the binary logit model of part (b) separately for single students and for married students. Report the estimated coefficients and standard errors from each model. Interpret these results and compare your estimated coefficients across the two models. Also calculate and compare the time-value of each travel mode for single students and for married students; again calculate these values for incomes of \$15,000, \$25,000, and \$35,000. How can you explain that one set of results is very different across these two models, but the other set of results are roughly similar?
- d. The university has a strong commitment to environmental sustainability and would like to convince graduate students to take the bus rather than drive to campus. One proposal is to introduce express bus lines that would speed up the longest bus commutes. The express buses would reduce bus travel time by 10 minutes for the long-distance bus routes; long-distance bus routes are those that cost \$3 to ride. But new bus routes are costly, so the university will only implement this plan if an extra 10% of graduate students start taking the bus. Using your results from part (b), how many students will switch from driving to taking the bus because of these express routes?
- e. An alternative proposal is to reduce the cost of these long-distance bus routes by \$0.50. Using only the results from part (b) and no new calculations, you should be able to form a good idea of how this proposal compares to the express bus routes. Do you think this \$0.50 subsidy will yield more or fewer new bus riders than the express bus routes, and how did you come to this conclusion?

Now calculate the effect of this subsidy. Using your results from part (b), how many students will switch from driving to taking the bus because of this subsidy?

### Problem 3: Multinomial Logit Model

Use the `travel_multinomial.csv` dataset for this question.

- a. Model the travel mode choice to commute to campus during spring as a multinomial logit model with the same variables as in part (b) of problem 1. Report the estimated coefficients and standard errors from this model. Interpret these results, including the elasticity of each alternative with respect to the cost to drive, the cost to take the bus, and the time to take the bus (i.e., you should discuss 12 elasticities, 4 alternatives  $\times$  3 elasticities each). You should notice a pattern in these elasticities. Describe this pattern and how it relates to the logit model. Also calculate the time-value of each travel mode; again calculate these values for incomes of \$15,000, \$25,000, and \$35,000.
- b. Age is another demographic variable that might affect a graduate student's commute decision, and it might have differential effects on the travel modes. For example, aging might have no direct effect on the utility of driving to campus, but biking through the rolling hills of the Pioneer Valley might become more difficult as a student ages. Again model the travel mode choice to commute to campus during spring as a multinomial logit model, but now include age with alternative-specific coefficients. Report the estimated coefficients and standard errors from this model. Interpret these results, including the elasticity of each alternative with respect to the cost to drive, the cost to take the bus, and the time to take the bus. Also calculate the time-value of each travel mode; again calculate these values for incomes of \$15,000, \$25,000, and \$35,000.
- c. The university is again considering the express bus route proposal from part (d) of problem 2. Using your results from part (b), how many students will begin taking the bus because of these express routes in the spring when biking and walking are also options? Also calculate how many students will no longer use each alternative because of the bus (i.e., how many current bikers, how many current drivers, and how many current walkers will start taking the bus). Compare these results to your results from part (d) of problem 2 and explain why they are similar or different.
- d. The university may consider implementing express bus routes even though they do not cause an extra 10% of graduate students start taking the bus. Being an altruistic public institution, the university will begin running express buses if the value they provide graduate students exceeds the cost. The university expects the express routes will cost an extra \$500 each day. Using your results from part (b), should the university implement these express bus routes? (Reminder: This is a random sample of 1,000 graduate students, not the entire graduate student population.)