#### Philadelphia Household Travel Survey

University of Pennsylvania Erick Guerra

**Purpose:**

The purpose of this assignment is to learn to understand and work with the 2012 Philadelphia household travel survey. Follow the instructions from the labs to load and examine the data in R. Please submit your assignment and a clean version of your code to Canvas. The last three questions are a refresher for interpreting regression outputs and require no code.

1. Draw a random household from the second ten households in your data. If your data frame is named dat, this command will do it for you: sort(unique(dat$HH\_ID))[sample(11:20, 1)]

Write a brief description of the household. Be sure to include: their income, race, composition, county of residence, and whether they own a car.

1. Describe the daily activities and travel of the household’s members using the trip data.
2. Provide an estimate of the total number of bicycle trips represented by the survey data. Note that you should use the Person Weight to make this estimate. This number represents the total number of people each person in the survey is supposed to represent. For example, if a person’s Person Weight is 65, then that person represents 65 people in the Philadelphia region.
3. Make a table that shows the mode choice for residents from the county of the household from question 1. Note that you will have to make a choice about how to group the modes together. Note also that I asked for the home county not the origin of a trip county. Look for this in the data dictionary.
4. Make a graphic that shows the relationship between household income and the age of a household’s vehicle for all households. (Hint: merge the household data to the vehicle data with the merge() command. Use an internet search or type ?merge into the console to learn how to use the command.)
5. Make a table or graphic that shows the relationship between household income and mode choice for all households. (The boxplot() command is a nice option, but a table is just fine.) Again, you must choose how to combine modes.
6. The last three questions reference the output of an OLS regression model predicting daily boardings at US light rail stations.
   1. According to the regression model, each additional job within a half mile of a station correlates with how many more daily transit riders per day?
   2. If there are an average of 1,793 boardings per station and 3,130 jobs around each station, what percentage increase in ridership does the model predict due to a doubling of the number of jobs around a station (hint: elasticity).
   3. Describe the relationship between AM peak service frequency and light rail boardings.

|  |  |
| --- | --- |
|  | |
|  | *Dependent variable:* |
|  | Daily light rail boardings |
| People within a half mile | 0.051 |
|  | (0.034) |
| Jobs within a half mile | 0.179\*\*\* |
|  | (0.030) |
| AM peak service frequency | 105.822\*\*\* |
|  | (9.887) |
| Bus connections nearby | 115.562\*\*\* |
|  | (22.915) |
| Constant | -1,112.977\*\*\* |
|  | (232.457) |
| Observations | 406 |
| R2 | 0.408 |
| Adjusted R2 | 0.402 |
| *Notes:* Standard errors in parentheses | \*p\*\*p\*\*\*p<0.01 |