Department of City and Regional Planning
CPLN 505 Planning By Numbers
Professor: Megan Ryerson

ASSIGNMENT 3: CATEGORICAL DATA MODELS
(due April 20, 2022, 11:59 p.m.)

A. **MODELING LAND USE CHANGE:** The dataset, Chester Urban Growth ('Chester_Urban_Growth.csv' in 'Datasets' folder on Canvas. For metadata on this csv, see the .xlsx file) includes selected GIS-derived land cover, land use, transportation facility, and demographic data for Chester County, PA for 1992 and 2001. Each row in the dataset corresponds to a 500-meter raster cell location.

   1. Read the data into R. Run descriptive statistics as necessary.

   2. Create a new binary (0/1) variable, [CHNG_URB] indicating those raster cells that were either farmland or pasture land or forest in 1992, and which converted to urban uses by 2001. Indicate these cells with a 1. All other cells should be coded to 0.

   3. Build a best "lean & mean" binomial logit model that identifies the determinants of agricultural/pasture/forest land-to-urban land use change between 1992 and 2001 (This is the 0/1 variable created in step 2). Use whatever independent variables you think appropriate, and be sure to explain your logic and model development steps.

   4. Explain the results of your best model in a few paragraphs for non-statisticians. How well does your model fit the data? Which factors are most important in explaining land use change in Chester County from 1992-2001? How do you know? Do you have any ideas for improving the performance of your model?

   5. Include a few plots or tables that show the probabilities versus changes in key variables in your model.

B. **MODE CHOICE MODELING:** How do a person's household, demographic, and trip-characteristics affect their choice of transportation mode to work? To answer this question, we will use data from the DVRPC HHTS.

   1. Develop the best ("leanest and meanest" and unbiased) binomial LOGIT model you can explaining whether a commuter **drove to work** or used another mode. Use any and all independent variables you think appropriate, including dummy variables. Summarize your model results in a paragraph. Include a few plots or tables that show the probabilities of each choice versus changes in key variables in your model.

NOTE: THIS MAY REQUIRE YOU TO LOOK THROUGH THE META_DATA. It requires selecting the observations from the trip table that correspond to a person's work trip. A person may have multiple work trips – how will you handle this? (I suggest taking the first one-way work trip of the day, but there are others methods to handle this).

   2. Develop a second binomial LOGIT model explaining whether a commuter walked/biked to work or used another mode. Use any and all independent variables you think appropriate, including dummy variables.

- Summarize your model results in a few paragraphs. Include a few plots or tables that show the probabilities of each choice versus changes in key variables in your model.

- How does this model compare to the results of the models from question B1? Venture some reasons for the differences.

3. Develop the best ("leanest and meanest" and unbiased) multi-nomial LOGIT model you can explaining whether a traveler drove to work, carpooled, took public transit, or walked or biked. Use any and all independent variables you think appropriate, including dummy variables.

- Summarize your model results in a few paragraphs. Include a few plots or tables that show the probabilities of each choice versus changes in key variables in your model.

- How well does your model predict observed mode choices? For which modes does it perform best? Worst? How do you know?

- What ideas do you have for additional variables (that are not available in the data) that might make for better models?