**CE C265 & PBHLTH 285**

**Spring 2025**

**Assignment 2: Estimating a bicycle exposure model (15 points)**

**Submitted by: Chun-hin (Anson) Sit, Jesus Hinojosa**

**Feb 2025**

In this assignment, you will estimate a exposure model, selecting the appropriate explanatory variables (features) and the model that performs the best, i.e. predicts the counts most accurately. The dependent variable (target) in the data is Annual Pedestrian Traffic at intersections on California state highway system and the explanatory variables (features) consist of infrastructure, demographic, climate, network connectivity, transit, employment/land use and other characteristics. Some of these characteristics are available at different radii of catchment area (buffer) around the intersection, half-mile, quarter-mile, and one-tenth of a mile. This dataset might include variables that would not be useful in the modeling process and miss some that could potentially have been useful.

Please answer the following questions:

1. Based on the columns, please list some initial hypotheses on how each variable might affect pedestrian traffic, that is whether it would have an increasing or decreasing effect, or no influence for increases in that explanatory variable. Please do not revise this after estimating your model. It is always good to note what your initial hypotheses were and how your result might conflict with some of those, since your results are dependent on the variations captured in your data set ***(2).***

Answer: (Jesus)

1. Perform initial data descriptive analysis. Include mean, standard deviations for the dependent variable (target) and the explanatory variables (features). Display correlations among them and plot a few dependent variables with the explanatory variable to observe the functional form. How does your data descriptive analysis and your knowledge of pedestrian traffic influence the variables you select for your model ***(3)***?

Answer: (Jesus)

1. In the modeling stage, please begin with estimating a linear regression model, and then the count models, and any other model that you would like to estimate. For the initial model, you may want to include the explanatory variables one by one to observe the effect that each has on the significance of the other. Remember that the R-squared for LR models would mostly increase as you add more variables, whatever those may be ***(5).***

Answer: (Jesus)

1. Please display some of the models you estimated including your final model. What were the reasons for selecting this model as your final model ***(2)?***

Answer: (Jesus)

1. Please discuss any limitations in your model and how you could correct those. Discuss issues such as endogeneity, simultaneity etc. ***(3).***

Answer: (Jesus)

Please submit your responses in the form of a short report/write-up including your code. You could do this assignment on your own or team up with **one** other student.