CE C265 & PBHLTH 285

Spring 2025

Assignment 3: Crash Severity Modeling

In this assignment you will develop a crash severity model using an ordered logit regression approach implemented in Python. Your objective is to identify and quantify the factors that influence the severity of crashes by analyzing a variety of explanatory variables such as road characteristics, driver behavior, environmental conditions, and vehicle types. You will use data extracted from TIMS (https://tims.berkeley.edu). Details on the data can be found at (https://tims.berkeley.edu/help/SWITRS.php).

Option 1

You will begin by reviewing the dataset and its variable definitions to form your initial hypotheses about how each factor might impact crash severity. For each variable, consider whether you expect its effect to be increasing, decreasing, or neutral, and record these expectations without revising them later based on your model outcomes.

Following this, conduct an exploratory analysis of the data to obtain summary statistics including means, standard deviations, and correlations among the variables. Use visualizations such as histograms, boxplots, and scatter plots to gain insight into the distributions and relationships within the data. This preliminary analysis will help you decide which variables are most relevant for your model.

In the model estimation phase, implement an ordered logit model to evaluate crash severity. Explain why this modeling approach is appropriate for your data and discuss any alternative specifications you explore, such as incorporating interaction effects or applying transformations to your variables. Be sure to report key diagnostic metrics such as log-likelihood, pseudo R-squared, and the statistical significance of the coefficients.

After estimating the models, interpret your results by discussing the rationale behind selecting your final model and providing a clear interpretation of the coefficients. Highlight which factors are statistically significant and describe the direction and magnitude of their effects on crash severity. Additionally, reflect on the limitations of your analysis—considering issues like data quality, potential omitted variable bias, and endogeneity—and suggest ways to address these concerns in future work.

Option 2

As an alternative to building your own model from scratch, a simple model code will be made available to you (on bcourses). With this code, you can calibrate the model across various

counties or cities and then compare and discuss the differences in the estimated coefficients across these geographic regions. This approach will allow you to explore how local factors might influence crash severity and to interpret the variations in the results. Regardless of the option you choose, you are expected to provide a thorough interpretation of the model outcomes, similar to option 1.

Submit your work in a concise report that includes a narrative discussion of your findings along with the Python code used for data processing, model estimation, and visualization. You may complete this assignment individually or in pairs.