

ECON 8320 – Tools for Data Analysis

Assignment 11 [25 points]

Modeling data to draw inference or to make predictions is difficult and requires an understanding of the data to be modeled, the models available, and iteration as that model is implemented and improved. For this assignment, you will pose two questions, and use the models and tools available through Statsmodels and Scikit-Learn to answer those questions.

1. Choose a question, using any data available through the ACS or NFL databases, that can be answered with regression analysis.
 - (a) Collect the data required to answer that question
 - (b) Process the data in preparation for modeling
 - (c) Generate a model that can answer your question, focusing on improving the explanatory power of the model (use R^2 or a pseudo- R^2 to assess explanatory power)
2. Choose a factor, using any data available through the ACS or NFL databases, that you would like to be able to predict.
 - (a) Collect the data required to predict that variable
 - (b) Process the data
 - (c) Generate a predictive model of your variable, and focus on maximizing the accuracy score of your predictions. Be sure to use some form of cross-validation in your modeling.

Problem (1) should be completed using statsmodels, and Problem (2) should be completed using sklearn. Please submit code, as well as a short writeup of your findings (no more than 2 pages double spaced).