Regression Project

DSC 531

February 21, 2025

This project uses data derived from IPEDS (Integrated Postsecondary Education Data System), in the corresponding Github repository. The questions for this project are:

1. Build your best model to predict whether an institution has a graduation rate above the median for all institutions.
2. Build your best model to predict whether an institution has a graduation rate above the median for institutions with an incoming cohort of at least 200.
3. Build your best model to predict whether an institution has a graduation rate above the median for institutions with an incoming cohort of at least 400.

Your results for each must:

1. Be based on a median cut-off computed to three decimal places–*e.g.* if a median was computed as 0.531725, use 0.532 (or 53.2%) as the cutoff.
2. Include specification sheet that includes all variables that were considered as potential predictors in any model selection attempt.
3. You do not have to provide a data set of all model selection runs (method/criteria/steps) used in your efforts to arrive at a final model, just give a general list of what you explored.
4. Your final models with a discussion of why you chose them over other candidates and interpretations of their effects.
5. A comparison of the models for the three scenarios–what do they do differently, what is the same.
6. You may use the variable specifications we set forth in class, or you may modify them. For whatever you chose to do, set up a Git repository with a file for the specs you used and a file with the write-up for your final models.