Regression Project

DSC 531

February 17, 2026

This project uses data derived from IPEDS (Integrated Postsecondary Education Data System), in the corresponding Github repository.

1. Build your best model to predict graduation rates for baccalaureate degrees (based on completion within 150% of time), limit to institutions with an incoming cohort of at least 200.

   For this model you must include:

   (a) A specification sheet that includes all variables that were considered as potential predictors in any model selection attempt.

   (b) A data set of all model selection runs (method/criteria/steps) used in your efforts to arrive at a final model.

   (c) Your final model with interpretations of the effects and a discussion of why you chose it over other candidates.

   (d) Set up a GitHub repository with a file for the specs you used, a data set for the models attempted, and a file with the write-up for your final model.

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

   For this model(s) you must:

   (a) Base it 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.

   (b) Include specification sheet that includes all variables that were considered as potential predictors in any model selection attempt.

   (c) You do not have to provide a data set of all model selection runs used in your efforts to arrive at a final model, just give a general list of what methods and criteria you explored.

   (d) Your final model with interpretations of the effects and a discussion of why you chose it over other candidates.

   (e) Put these results into the GitHub repository also.