

PQHS 471 (2018) Midterm

The data are from a US census in the past. It has 15 variables. One of them, `income`, is our outcome. Most variables are self explanatory. The variable `fnlwgt` is the final sampling weight, a number often calculated in survey sampling.

The data have been split into a training set of 25000 observations and a test set of 7561 observations.

Study the relationship among the predictors and between the predictors and the outcome. Build prediction models using the training set. The test set should be used only for evaluating your final model(s).

Turn in your code as a file, and describe your results in a report.

The midterm is due by the midnight of Friday, March 9.