1   Methodology

So how exactly do we go about finding the variables that are relevant for graduation rates? It turns out that we can use a few different linear regression methods in order to accomplish this task.

1.1   Lasso Regression

Lasso Regression is essentially just a slight modification of Least-Squares Regression. In Least-Squares, the optimal solution is the beta vector that minimizes the sum of squared residuals. Because Lasso Regression's minimization problem is slightly different, the optimal solution for Lasso is the beta vector that minimizes both the sum of squared residuals (bias) and the absolute norm of the beta vector (model complexity). In addition to possibly performing better from a prediction standpoint, the solution to Lasso has the added benefit of setting multiple beta coefficients equal to zero. The variables that correspond to these beta coefficients of zero are usually interpreted as having a minimal impact on the response variable.

1.2   P-Values

While p-values are usually associated with hypothesis testing, they can actually also be used for variable selection as well. For instance, imagine that we're in the linear regression setting where we have p explanatory variables, n observations, and one response variable y. In order to find the variable with the most explanatory power, lets run p regressions where we regress y individually on every single explanatory variable. We then compare these p regressions and select the variable with the beta coefficient that has the smallest p-value. Lets call this variable X1. In order to find the variable with the 2nd most explanatory power, we can run p-1 regressions where we regress y on X1 and a second explanatory variable. We then select the variable with the beta coefficient that has the smallest p-value. This variable can be interpreted as the second most important explanatory variable. We can iterate this process until we're satisfied with the number of relevant explanatory variables we have. While this forward stepwise selection algorithm is greedy and not necessarily optimal, it performs fairly decently and is more feasible than an exhaustive approach.

1.3   BIC

BIC (Bayesian Information Criterion) is a criterion for model selection. In a subset of models, the model with the lowest BIC value is deemed the best model. BIC is composed of both a bias term and a model complexity term. As a result, in order to achieve a low BIC, a model has to do a decent job of minimizing bias without too much model complexity. While BIC is normally used for model selection, it can also do a decent job at variable selection. Using the forward stepwise selection algorithm described in the previous subsection,

we would sequentially add a variable to the model, and this was the variable that decreased BIC the most. We would stop adding variables to the model once they started to increase the BIC of our model.