Building Binary Logistic Regression Models

Setpwise variable selection

Forward Stepwise

- Start with a null model
- Add predictors to the model **one at a time**. Choose the best model

Backward

- It is the reverse of forward : Start with all predictors and then drop one at a time and then select the best model

Model Selection

How to improve model

- By adding more independent variables?
- By deriving new variables from available set?
- By transforming variables ?
- By collecting more data?
- How do we choose best model from the list of fitted models with different parameters

AIC and BIC

- AIC and BIC values are like adjusted R-squared values in linear regression
- Stand-alone model AIC has no real use, but if we are choosing between the models AIC really helps.
- Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models
- If we are choosing between two models, a model with less AIC is preferred
- AIC is an estimate of the information lost when a given model is used to represent the process that generates the data

AIC and BIC

- AIC= $-2\ln(L)+ 2k$
  - L be the maximum value of the likelihood function for the model
  - k is the number of independent variables
- BIC is a substitute to AIC with a slightly different formula. We will follow either AIC or BIC throughout our analysis