**Variable Selection**

select the best subset of predictors to explain the data in the simplest way

unnecessary predictors add noise to the estimation of other quantities that interest us, wasting degrees of freedom

collinearity is caused by having too many variables provide the same information

save time any money by not measuring redundant predictors

Model Hierarchy

lower order terms should not be removed before higher order terms of the same variable

is not significant but is

remove term to get reduced model

if you make a scale change from to , model becomes

the first order term reappears

scale changes shouldn’t make any important changes to the model

**Stepwise Testing**

compares successive models

doesn’t have to be 5%, 15-20% cutoff may work best

possible to miss the optimal model because adding/dropping one variable at a time

no multiple testing correction so p-values should not be assumed to be the Type I Error

procedures are not directly linked to final objectives of prediction or explanation so may not help solve the problem of interest

Backward Elimination/Top Down

Step 1 start with all the predictors in the model

Step 2 remove the predictor with the highest p-value >

Step 3 refit the model

Step 4 repeat Steps 2-3 until all non-significant predictors are removed

Forward Selection/Bottom Up

Step 1 start with no predictors

Step 2 add the predictor with the lowest p-value <

Step 3 refit the model

Step 4 repeat Step 2-3 until no new predictors can be added

Stepwise Regression

combination of backward elimination and forward selection

at each stage, a variable may be added or removed

top down stepwise regression alternate drop step and add step

bottom up stepwise regression alternate add step with drop step

**Criterion Testing**

find the model that optimizes the measure of goodness of fit and reduces the model’s complexity

prefer small values of AIC and BIC with smaller and a small number of parameters

larger models fit better so will have smaller but use more parameters

need to find a balance between and number of parameters

BIC penalizes larger models more heavily and tend to prefer smaller models compared to AIC

Akaike Information Criterion (AIC)

Bayes Information Criterion (BIC)

**Adjusted**

adding variables can only decrease and increase

is not a good criterion because it always prefers the larger model

use significant changes of and the criterion

adding a predictor will only increase if it has some predictive value

**Mallow’s Statistics**

average mean square error of prediction =

is from the model with all the predictors

is from a model with parameters

for the full model,

if a -predictor model fits, then and

if a model has a bad fit, will be much larger than

models with a good fit will have small and a around or less than