

3.2: Variable Selection

Dr. Bean - Stat 5100

Why might we prefer simpler models to complex ones? *Should* we prefer simpler models to more complex ones?

Why is variable selection not something we would normally want to use in an experimental setting?

Clearly, it would be better to compare all possible models, rather than a subset (in stepwise methods). Why do stepwise methods even exist?

In what ways might multicollinearity “mess-up” stepwise selection approaches?

(Open-ended discussion, no right answer) How do we decide on the “right” model when each variable selection technique suggests a different model?