

Building a multiple regression model

stats545.1 Theory of linear models

Gunnar Stefansson

September 18, 2018

Introduction

- Have several independent variables
- Want to select some into regression
- Want to evaluate quality of resulting model
- Want to improve into a final model

Variable selection: Measuring quality

- R^2
- AIC
- BIC
- SSE
- MSE
- P -values

Example: Use the ecosystem data set and select a single variable in a simple linear regression to predict the growth of cod. Compare the various criteria.

Variable selection: Forward or backward

Model selection:

- All subset regression
- Forward stepwise regression
- Backwards stepwise regression

R commands: `add1`, `drop1`, `summary` - followed by `anova(fm.final, fm.full)`

Example: Use the ecosystem data set and conduct a forward stepwise regression to predict the growth of cod. Compare the various criteria for model selection.

Example: Use the ecosystem data set and conduct a backwards stepwise regression to predict the growth of cod. Compare the various criteria for model selection.

References Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. 1996. Applied linear statistical models. McGraw-Hill, Boston. 1408pp.

Belsey, D. A., Kuh, E. and Welsh, R. E. 1980. Regression diagnostics: Identifying influential data and sources of collinearity. John. Wiley and Sons, New York. 292pp.