

ESM 206B
Hampton
6 April 2009

Lab 1 – Model selection using stepwise regression in JMP

- 1) Open FakeCyano_data.xls – a simulated data set including water samples from 90 hypothetical water bodies. Your objective is to determine which variables among these predictors creates the best model describing algal biomass.
- 2) Run multiple regression with all predictors (analyze-> fit model, with algal biomass as response – standard least squares in the “personality” box). For now, do not consider interactions (to keep it a little more simple).
- 3) How does the model fit? (r^2 , p-value for whole model, any problems implied by plot of residuals or leverage plots?)
- 4) From this model what do you surmise about the effect of predictors? (r^2 , p-values, partial regression coefficients “Estimate” in JMP)
- 5) Compare the partial regression coefficients from this multiple regression to simple linear regressions for the two strongest predictors in this model. Looking at the data sheet again, go to Analyze ->Fit y by x, run a simple linear regression for each of the two strongest predictors against algal biomass – how do the regression coefficients for these two simple models compare to the partial regression coefficients from the full multiple regression model?
- 6) Do a stepwise regression to find a more parsimonious model. Analyze – fit model – stepwise, run, adjust prob to enter (try 0.15), prob to leave (try 0.15), “forward” stepwise is the default, Go.
- 7) Look at Step History, this tells you the order in which they were entered or removed. What can you infer about the univariate x-y relationships from the ordering of the variables in Step History?
- 8) Record AIC from this window.
- 9) Make model, Run Model. Do your conclusions differ? (R^2 , Partial Regression Coefficients?) Anything troubling in residuals or leverage plots?
- 10) If you haven’t already, examine multicollinearity. If you find multicollinearity, how will you decide your course of action?
- 11) Let’s get the AIC for the full model to see how it compares to the smaller model that resulted from the stepwise regression. Do you still have the “Fit Stepwise” window open? If not, go back to your data sheet, Analyze-Fit Model – Stepwise in “personality” box – Run Model. If you still have “Fit Stepwise” window open, go there now. Now hit “Enter All” – with all predictors in the model, you can read off the AIC value for the full model. Is it better or worse than the one selected through stepwise model selection (i.e., Step 8 above)?
- 12) Click through the checked variables (the boxes under “Current Estimates” – “Entered” – including them in the model turns them blue) – when you add and remove variables manually, do you find a better model than the one selected through stepwise selection, judging from AIC? No need to be exhaustive, just explore.