

StatR 502 Homework 5

Gregor Thomas

Due Thursday, Feb. 12, 2015, 6:30 pm

Submission guidelines: please submit a PDF if possible, otherwise a self-contained HTML file, and optionally your `.Rmd` file. As always, ask in the discussion forum if you're having trouble!

(1) Abalone Models

- (a) Using the abalone data from HW 4, search for “best” models using stepwise regression (or other model search strategies such as `leaps`) to find “best” models. Use at least two search methods (e.g., forward, backward, both, using AIC, using BIC) with different starting points. Do you get the same final model with the different methods?
- (b) Use one of your search methods from (a) on a subset of the data excluding the big outlier. How does it change your results? Do you think it's worth omitting the outlier, or would you prefer another strategy?

Note: for both (a) and (b), you should either transform the response variable or use a Poisson GLM. Note that AIC/BIC cannot help you compare a Poisson GLM with a Gaussian LM—the different error assumptions make the likelihoods incomparable.

Book Problems

Do G&H Chapter 7 problems **1, 2 and 4** (pp. 152). For number 1, also do part (d) below:

(d) Simulation can be used test hypotheses, even generate p-values. We can consider the situation described in the problem as a *model*. Perhaps we have another basketball player and we have a null hypothesis that his shooting percentage is 60%, just like the first player. He's talking a big talk, so we have an alternative hypothesis that his shooting percentage is $>60\%$. We test the new player, having him take shots until he misses two in a row. He takes 15 shots (i.e., 13 shots without two misses in a row, then shots 14 and 15 are both misses). Under the null model, what is the probability of taking at least 14 shots? Do you think the new player is better than the original player?

You shouldn't need to do any new simulations for part (d), just examine the simulations you've already done.