Exercises 6 · Multiple regression and stepwise selection

(1) Beauty, or not, in the classroom

UT–Austin, like every other major university in the country, asks students to evaluate the quality of instruction they have received from their professors. In your career at UT, you will almost certainly have participated in this process, rating your professors on a scale of 1 (very unsatisfactory) to 5 (excellent). These ratings, in turn, are part of what administrators use to evaluate faculty performance, set salaries, promote instructors, and confer teaching awards. You therefore have a non-trivial say in the future direction of the university.

The file "profs.csv" contains data on course-instructor surveys from a sample of 463 courses at the University of Texas from 2000–2002. You are also given information about the individual courses and professors—including, most controversially, a rating of each professor's physical attractiveness, as judged by students. The data represent evaluations from 25,547 students and most major departments.[1]

[1] Data from "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity." Daniel S. Hamermesh and Amy M. Parker. *Economics of Education Review*, August 2005, v. 24 (4) pp. 369–76.

The variables included are:

minority: is the professor from a non-Caucasian ethnic minority?
age: the professor's age.
gender: a factor indicating the professor's gender.
credits: a factor indicating whether the course is a single-credit elective (e.g. scuba diving or ballroom dancing, coded "single") or an academic course (coded "more").
beauty: a rating of the professor's physical attractiveness, as judged by a panel of six students.[2]
eval: the professor's average teaching evaluation for courses in the sample, on a scale of 1 to 5.
division: whether the course is an upper or lower division course.
native: whether the professor is a native English speaker.
tenure: whether the professor is tenured/tenure-track, or not.
students: the number of students that participated in the evaluation.
allstudents: the number of students enrolled in the course.
prof: a unique numerical identifier for the professor being rated.

[2] The score was averaged across all six panelists, and shifted to have a mean of zero.

The fundamental question for you to address is: does it seem that teachers who are perceived as more attractive receive higher course-instructor evaluations, other relevant factors being equal? Use your knowledge of statistical modeling to address this question.

If you do not believe there is an effect, explain how you arrived at this conclusion. If, on the hand, you believe there is an effect, make sure you:

1. use a permutation test to assess whether the effect is different for male versus female teachers.

2. quantify the likely magnitude of the effect (with error bars) as well as the magnitude of any gender differences that may exist.

(2) Predicting demand in a bikesharing network

The data in bikeshare.csv, from the course website, contains data from the Capital Bikeshare system, in Washington D.C. The goal is to build a predictive model for system-wide demand for bikes, using important features about the time of year, time of day, the weather, and so forth.

There is a lot of information on the data set in bikeshare-key.txt, including a codebook for all the variables. Read this file to familiarize yourself with the data. The outcome variable is demand, as measured by the "cnt" variable. This is the total count of bikes rented in a given hour, including both "casual" and "registered" users of the system.

After reading this file, please address the following three questions.

(A) First fit a very simple model for demand ("cnt" variable) in terms of three variables: day of week, time of day, and temperature. Report R^2 for this model.

(B) Then use stepwise selection to build a predictive model for demand. Report R^2 for this model.

(C) Once you've selected a model by stepwise selection, use a train/test split strategy to estimate the root mean-squared predictive error (RMPSE, as defined in the course packet) both for the simple model above, and for the model you've selected by stepwise selection. Make sure you average over multiple train/test splits. Report your estimates for RMPSE of both models.

Here are several pieces of advice:

1. A lot of the categorical variables (like season, hour, day of week, etc) are given numerical codes here. Remember the "factor" command, which tells R to treat a numerical code as a categorical variable.

2. Interactions are important here. You can probably make some guesses up front about what the important interactions in the

data set will be. Make sure you allow the possibility of important interactions in your model.

3. These models will take a fairly long time to fit, especially if you include lots of interactions, because there are lots of parameters to estimate. Stepwise selection will take even longer. Therefore give your computer plenty of time to let stepwise selection run to completion while you do something else with your time.

4. It's OK for your scope of variable selection (i.e. the variables you consider) not to include ALL possible interactions. If you can rule out some interactions up front—perhaps, for example, using an ANOVA table on a very big model—this will certainly make stepwise selection run faster. However, remember that your time is a lot more valuable than the computer's time! So don't go crazy here trying to eke out a few saved minutes of computation time.