Exercises 6 · Multiple regression

Due Monday, March 7, 2016

(1) Vote undercount

The data in "georgia2000.csv" contains Georgia's county-level voting data from the 2000 presidential election. You might recall that the 2000 election was among the most controversial in history, and turned on an esoteric set of issues surrounding voting machines, vote counts, and the Equal Protection Clause of the Constitution.

This file contains the following information for all 159 counties in Georgia:
votes: number of votes recorded
ballots: number of ballots cast
equip: voting equipment (lever, optical, paper, punch card)
poor: coded 1 if more than 25% of the residents in a county live below 1.5 times the federal poverty line; coded 0 otherwise.
perAA: percent of people in the county who are African-American
urban: indicator of whether county is predominantly urban (1)
atlanta: indicator of whether the county is in Atlanta (1)
gore: number of votes for Gore
bush: number of votes for Bush

This problem is about understanding vote undercount, or the difference between the number of ballots cast and the number of legal votes recorded. There can be many different reasons for undercount. Voters may have chosen not to vote for any presidential candidate; they may have voted for more than one candidate, in which case their votes were disqualified; they may have misunderstood the instructions on the ballot; or the equipment may have simply failed to register their choices.

One possibility that worries state election boards is that certain kinds of voting machines (paper, lever, etc.) will undercount valid ballots at higher rates, and that some precincts are unable to afford better machines. The questions at issue here are twofold:

(1) Do certain voting machines lead to systematically higher rates of vote undercount than others, adjusting for other important factors?

(2) If so, do the worse machines (i.e. those leading to higher undercount rates) tend to occur more frequently in counties that are socioeconomically disadvantaged, or have higher proportions of African Americans?

Use your knowledge of statistical modeling to address these questions.[1] Make sure to include an assessment of your uncertainty in your conclusions.

[1] Remember that larger counties might have more undercounted ballots just because they are larger, and had more ballots to begin with!

(2) 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.[2]

[2] 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.[3]
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.

[3] 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.[4]

[4] 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. quantify its likely magnitude (with error bars!); and
2. assess whether the effect is different for (a) male versus female teachers, and (b) lower- versus upper-division courses.