Exercises 6 · Quantifying uncertainty, continued; multiple regression

Due Monday, March 2, 2015

(1) Prediction uncertainty in the normal linear regression model

Refer to the "Newspapers' walkthrough on the class website.[1] This will show you how to use the normal linear regression model to compute standard errors for estimated model parameters and prediction intervals for forecasting future observable quantities. Once you're comfortable with this walkthrough, continue with this problem.

[1] http://jgscott.github.io/teaching/r/newspapers/newspapers.html

Download the data set "shocks.csv." This data was taken by Monroe Shocks and Struts, a company that manufactures high-performance shock absorbers for top-end cars. Monroe offers a range of shock absorbers for cars of various sizes. These different shocks are distinguished from one another by their "rebound," a number which describes how aggressively the vibrations from the road are absorbed by the shock. Having an accurate understanding of a shock's rebound is important for safety; you don't want to put shocks designed for an SUV on a small car, or vice versa.

As part of its manufacturing process, Monroe tests each shock absorber to make sure it performs to the required rebound specification. They have one very accurate test of the shock's rebound, but this test is expensive. They also have a cheaper test, but this is less accurate.

In "shocks.csv," you have rebound readings on 35 different shock absorbers for both the expensive test and the cheap test. If the cheap test can accurately predict the result of the expensive test with minimal uncertainty, then it's OK to use the cheap test. But if it can't, then the expensive test must be used instead.

(A) Suppose the company is willing to use the cheap test as long it can predict at least 90% of the total variation in the readings given by the expensive test. In light of this data, should they use the cheap test? Why or why not?

(B) Now suppose the company adopts a more specific standard, and decides it is willing to use the cheap test if both of the following criteria are met under the assumptions of the normal linear regression model. First, the slope of the regression line for the expensive test, given the cheap test, is close to 1, as measured by a 95% confidence interval. Second, the 95% prediction interval for the value of

the expensive test, given the cheap test, is no wider than 16.5 units of rebound, as measured from center to endpoint. (Or, measured from endpoint to endpoint, the interval can be no wider than 33 units of rebound.) This criterion must be met for readings of the cheap test (x) in the low (510), middle (550), and high (590) end of the rebound scale. That is, if the prediction interval for y is too wide at any of these three different x values, then the cheap test is not precise enough and cannot be used.

In light of the data and these criteria, should the company use the cheap test? If not, what criterion was missed and how? What assumptions (in addition to linearity) must the company make about their data in order for the confidence interval and the prediction intervals to be valid? Do these assumptions seem reasonable in light of the data?

(2) Multiple regression

Complete the walkthrough on the wage gap at `http://jgscott.github.io/teaching/r/salary/salary.html`, which introduces you to multiple regression in R. Then use what you've learned, together with earlier lessons, to address the following question.

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

Your goal is to investigate the determinants of vote undercount, or the difference between the number of ballots cast and the number of le-

gal 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. Construct a good statistical model for vote undercount (or some transformation thereof) to address this question.[2] Use this model to assess the marginal effect of voting equipment in explaining the undercount pattern across Georgia in 2000, adjusting for other relevant factors. Make sure to include an assessment of your uncertainty surrounding the probable effect size of voting equipment.

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

(3) 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. This gives you 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.[3]

[3] 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.[4]
eval: the professor's average teaching evaluation for courses in the

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

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 or on the tenure track.
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.

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 multiple regression to address this question. [5]

[5] 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;
2. assess whether it is different for (a) male versus female teachers, and (b) lower- versus upper-division courses; and
3. play "devil's advocate" and make your best case for what else might be causing the association you claim to see.