5 · *More on regression*

Due Monday, February 23, 2015

(1) 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. Once you're comfortable with this walkthrough, use what you've learned to create your own version of Figure 5.7 (on page 111) of the course packet.

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

(A) Choose your favorite parameters β_0, β_1, and σ to serve as the "ground truth" for your simulation. Set up a Monte Carlo simulation that will generate data sets of size n from the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \, , \quad \epsilon_i \sim \mathrm{N}(0, \sigma^2) \, .$$

Play around with your parameters and sample size until you're satisfied that the true sampling distributions are dispersed enough to be interesting, but not so dispersed to the point where the least-squares estimate for a sample of size n is uninformative.

(B) Now modify your Monte Carlo simulation so that it:

1. Simulates a data a set.
2. Fits the least-squares estimate to the simulated data.
3. Forms two confidence intervals for the slope of the least-squares line, one using bootstrapping and the other using normal-theory formulas.[2]
4. Checks whether each confidence interval covers (i.e. contains) the true value used to simulate the data set.

[2] You don't need to explicitly apply the formulas; let the software do the work.

Once you're confident the simulation is working, run the Monte Carlo simulation at least 1000 times. This will take awhile, because each time requires a full run of bootstrapping. Turn in the following items: 1) a paragraph that describes the frequentist coverage property in your own words; 2) your script for Part B; and 3) the output of your simulation that assesses whether the confidence intervals you are constructing for each simulated data set approximately satisfy the frequentist coverage property.

(2) Prediction uncertainty in the normal linear regression model

For this problem, use 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?

(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]

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 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]

[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.

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

[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.