General Linear Model, Psychology 610/710

Friday, September 15, 2017

Threat - Description of the study

A social psychologist would like to examine if "stereotype threat" could be partially responsible for the difference in academic achievement between children with high and low socio-economic status. Previous research has shown that there is a widespread stereotype that poor people are less intelligent than rich people. The psychologist suggests the following hypothesis: when children from poor families are in a testing situation (that is, in an evaluative situation), they feel threatened by the stereotype, and this threat decreases their performance in the test. In other words, if one were able to remove the perceived threat, the difference in performance between children from high and low socio-economic status should be reduced or even disappear.

In order to test his hypothesis, the psychologist conducts a study in a combined middle and high school. 24 students are randomly selected for participation in the study (N = 24). The students take three tests measuring their academic achievement in English (Test 1), Mathematics (Test 2), and General Knowledge (Test 3). Before the tests, the psychologist says to half of the participants:

"You will take a series of intelligence tests. Previous studies have shown that these tests are sensitive to socio-economic status: that is, children from modest backgrounds score on average lower than students with more well-off parents" (threat present condition).

The psychologist says to the other half of the participants:

"You will participate in a study in which we will examine the teachers' effectiveness." He does not mention the stereotype (threat absent condition).

After the tests, participants indicate their age. At the school's main office, the psychologist obtains information about the yearly income of the parents of the 24 students.

This procedure yields an experimental design with two between-subject variables: Condition (dichotomous: threat present vs. threat absent) and income (continuous). The dependent variable is the mean performance of the students on the three tests (each of which yields a grade between 0 and 20).

Threat - Codebook for "ThreatData.dat"

Column Name Variable description Values

-----------------------------------------------------------------------------------------------------------------

1 SubID Participant identification number 1-24

2 Income Yearly income of the participant's 00-99

 parents (in thousands of dollars)

3 Cond Experimental condition 0 = threat absent

 1 = threat present

4 Perf1 Performance on test 1 0-20

5 Perf2 Performance on test 2 0-20

6 Perf3 Performance on test 3 0-20

7 Age Participant's age 01-99

Tests 1 to 3: higher values indicate better performances

In Class Exercise with Threat Data

1. Read the data in and inspect them (structure, summary, descriptive stats)
2. Go through the following steps to create a composite score that reflects overall performance on the three performance measures:
	1. Standardize each performance measure, individually
	2. Check to make sure you standardized these variables correctly
	3. Before making a composite score, check the reliability and write a short comment noting whether it makes sense to combine these three scores and use them as a single measure. Does it matter whether you use standardized or raw versions?
	4. Create a new variable (called PerfZ) that represents a student’s average score on these three tests (regardless of answer to part c)
	5. Generate descriptive statistics for this new measure
	6. Be prepared to explain why in the present data set it is important to standardize the performance measures before combining them into a composite score.
3. Generate a scatterplot that shows whether the overall performance varies as a function of family income.
	1. Add a best fitting line.
	2. Make a comment about the relationship. Strong/weak? Positive/negative?

For the following exercises, put yourself in the shoes of the psychologist who ran this study. For now, let’s say that you’re interested in explaining the variance in overall performance (the standardized measure, PerfZ) scores across students. What information allows you to better predict a particular individual’s score?

1. First, create a histogram of the overall performance measure, so you can see what the distribution looks like! You’d like a little more granularity than it provides by default, so figure out how to change the command line to get more bars to appear.
2. Would it make sense to do a model comparison here to see whether the basic model would perform better than the null model in estimating students’ composite performance? Why or why not?
3. Instead, do something much simpler. The bottom of the scale for the “Perf2” measure is 0, and a reviewer wants you to show that people performed significantly better than 0 on this measure overall. Do this using a model comparison (just using lm is fine). Comment on the result. Does the number that comes up as the “estimate” look familiar?