Lab 3: Building regression models

Data: n/a

Practical objective: To practice building, writing, and interpreting coefficients from regression models.

Background on saturated models: In class, we have alluded to the idea of *saturated* models. A saturated model is one that does not impose or encode within it any structure besides positing the existence of subgroup-specific means within all groups defined by the model. For instance, a simple linear regression model with a binary exposure (analogous to a *t*-test) is an example of a saturated model; no linear structure is being "assumed," and the coefficient estimates are going to be the same as what you would get by computing means/mean differences by-hand if you use ordinary least squares.

Exercises: Below is a set of scenarios and accompanying exercises that we will go through individually, in small groups, and/or together as appropriate and as time permits.

Scenario A: We conduct a three-arm randomized controlled trial to evaluate a new antihypertensive drug (X=0: control, X=1: standard dose of experimental drug, X=2: High dose of experimental drug). The outcome, Y, is systolic blood pressure (SBP) after some period of time on the treatment. While baseline SBP is not directly recorded, it is categorized as follows: (Z=0: normal, Z=1: elevated).

Exercise 1: Write down the simplest *saturated* model possible that would allow you to compare the mean SBP across the three groups, and interpret each of its coefficients.

Exercise 2: Describe the major purpose and advantage of a model that adjusts for Z (i.e., by inclusion Z in the model as a single covariate). Interpret each of the coefficients of this model.

Exercise 3: Write down a model that expands upon the model of Exercise 2 but allows an interaction between treatment and baseline SBP. Interpret each of the coefficients of this model.

Scenario B: We conduct a cross-sectional observational study in order to investigate the association between the total number of hours of sleep the night prior to an exam (X) and mean final exam score in an organic chemistry class (Y). Course average prior to the exam (Z) was also recorded.

Exercise 4: Write down a simple linear regression model that would allow you estimate the association of interest, and interpret each of its coefficients. Is this a saturated model?

Exercise 5: Consider a model that adjusts for Z. Discuss the circumstances under which this form of adjustment would be most helpful in allowing you to answer the scientific question of interest.

Exercise 6: Consider a model that expands upon that of Exercise 5 but allows an interaction between X and Z. Interpret each of its coefficients, and discuss the circumstances under which this allowance would be most helpful in allowing you to answer the scientific question of interest.