# Quantitative Methods II
# Laboratory Activity III
### December 6, 2011

The purpose for this week is very simple: I want you to have more practice in modeling hypotheses. The problem is that we all have different research focuses and areas of substantive expertise. As a result, any attempt to tailor the hypotheses to a specific research agenda will result in two things: Most of the class being outside their research agenda, and your professor saying stupid things. While the former may indeed be an intractable problem, I would like to minimize the frequency of the latter. As such, this laboratory activity will contain questions and data from various sub-disciplines in Political Science.

There are two activities that we may have time and do tonight. Each activity consists of a research question, a hypothesis, a set of data, and a question or so. Your job will be to just answer the several questions.

The test of the hypothesis includes you creating the correct model and testing that your chosen model is acceptable. That means testing the assumptions of that model.

The data can be downloaded from the course website's data page:

`http://courses.kvasaheim.com/pols6123/data.php`

## ACTIVITY 3.1: PATRICK HENRY COLLEGE

This activity uses the `patrickHenry` data file. It consists of seven variables and 662 records (students). The variables are grade point average (`gpa`), `reading`, `math`, and `composite` SAT scores, `gender`, `level` in college, and what type of `highschool` the student attended. Get a feel for the data.

★ ★ ★

The research question asks what effect the other variables have on the grade point average of the student. Broken down, the research suggests that the composite SAT score is positively related to the GPA, that females will have a higher gpa, and that students from private schools will have higher gpa.

- What are the four hypotheses you need to test?

- As all four variable affect the GPA (according to our assumptions), we need to have just the one research model using all four variables. Write out the research model.

- What type of variable is your dependent variable? What types of variables are your independent variables? (Continuous, discrete, nominal, ordinal, etc.) How do your answers to these questions affect the technique you use?

- What assumptions are you making about the dependent variable in your choice of modeling technique? Are these assumptions violated by the model results? Check.

| Student | SAT Math | SAT Reading | Gender | High School | College Level |
|---------|----------|-------------|--------|-------------|---------------|
| Bud | 800 | 800 | Male | Public | Freshman |
| Kelly | 200 | 200 | Feale | Public | Sophomore |

**Table 1.** *Information for the two students for the Patrick Henry activity.*

- What is the predicted gpa for each of the students described in Table 1?

- Produce a prediction plot of predicted GPA against composite SAT score. Create four curves: one for each combination of gender and type of highschool. (Which college level will you use?)

## Activity 3.2: Fuffing

This activity uses the `fuf` data file. This file contains records for each participant in an international (violent) conflict between 1980 and 2001, inclusive. There are 1624 records and 22 variables. Beside the obvious `state`, `year`, and `population` variables, there are `fuffer` and `fuffed` (first user of violent force in a conflict and target of said violent force), `polity2` (level of democracy in the State), government type indicator variables `democracy` (0/1), `anocracy` (0/1), and `autocracy` (0/1), `durable` (number of years since last major government change), `irst` (iron and steel production), `milex` (military expenditures), `milper` (military personnel), `energy` (energy production), `tpop` and `upop` (total and urban populations), `cinc` (capabilities index), `c_total` (neighbors), `polright` and `civillib` (political rights and civil liberties), `freedom` (average of the previous two), and indicator variable `majpower` (0/1 major power in the world).

⋆ ⋆ ⋆

The research question asks if we can predict the probability that a State will be the first user of violent force in a conflict (a `fuffer`). The research suggests that a State will have a higher probability of being a fuffer if it is less democratic (lower `polity2`), is unstable (lower `durable`), is a major power, and has a higher level of capabilities (higher `cinc`).

- What is the research model? What type of variable is the dependent variable? Which type of model will you be using?

- Which variables will you include? Why those variables? In other words, explore. You may wish to enter in a variable or two as quadratic (squared) as well as linear. You may not. This is where you create a model that logically fits the data without over-fitting the data. Keep the research question in mind!

- Which of the hypotheses were supported by the data and model? Which were not? What conclusions can you reach regarding the research question?

- For the States in Table 2, what is their probability of using violent force first if involved in a conflict? If we were to make Iran a perfect democracy (`polity2`=10) without changing any other variable, what would happen to its probability of using violent force first in a conflict? By how much, and to what level?

- Produce a prediction plot of probability of being the first user of force against the level of democracy in the State (`polity2`). Create three curves, one for each government type. Create an appropriate legend.

- Produce a prediction plot of probability of being the first user of force against the durability of the State (`durable`). Create three curves, one for each government type. Create an appropriate legend.

| state | polity2 | durable | majpower | cinc |
|---|---|---|---|---|
| Iceland | 10 | 55 | 0 | 0.000086 |
| South Africa | 4 | 75 | 0 | 0.007416 |
| Iran | 3 | 2 | 0 | 0.012866 |
| United States | 10 | 192 | 1 | 0.149820 |

**Table 2.** *Information for the four States for the Fuffing Activity.*