Quantitative Methods II
Laboratory Activity I
September 27, 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 three activities that we will 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 test the hypothesis and answer the 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. We were explicit with the assumptions for the classical linear model. We were explicit with the assumptions of the other models, too.

African Corruption

This activity uses the `gdpcap` data file. It consists of eight variables: `scode2` and `state` identify the State, `region` provides the region in which the State exists, `democracy` is the State's level of democracy (-10 to +10), `govt` is a trichotomous variable describing the State type, `OPEC` indicates OPEC membership, `hig` is the level of honesty in the government, and `gdpcap` is a measure of wealth in the State. Get a feel for the data by plotting the entire data set (`plot(gdpcap)` produces all of the pairwise plots — a graphical way of seeing correlations).

Your job is to answer the following questions as best you can (this is quite similar to a previous homework assignment, but they differ in your current skillset and a requirement or two).

(1) Appropriately test the null hypothesis:

$$H_0 : \text{The average GDP per capita in Africa is \$5000.}$$

Make sure you state the name of the test you will perform, test the assumption(s) and their relevant statistics, and the appropriate three-part conclusion. Include a boxplot summarizing the GDPs per capita in Africa. This boxplot will be used for *publication* purposes, so make it sparkle.

(2) Appropriately test the null hypothesis:

$$H_0 : \text{The average GDP per capita in Africa is equal to that of Latin America.}$$

Make sure you state the name of the test you will perform, test the assumption(s) and their relevant statistics, and the appropriate three-part conclusion. Include

a boxplot comparing the GDPs per capita for Africa and Latin America. Again, include a publication-worthy boxplot summarizing the GDPs per capita in Africa.

(3) You earlier found that there is a statistically significant relationship between GDP per capita (`gdpcap`) and membership in `OPEC`. Let us perform linear regression to see how robust this finding is. Let our dependent variable be the GDP per capita (`gdpcap`). Let us have two independent variables: the level of honesty in the government (`hig`) and OPEC membership (`OPEC`). Thus, our research formula will be

```
gdpcap ~ hig + OPEC
```

Note. *Observe that `OPEC` is a categorical variable; its estimated effect is* as measured relative to *some base category (OPEN Non-Member).*

Finally, let us perform linear regression (`lm`) to simultaneously estimate the effects of the level of honesty in government and OPEC membership on the GDP per capita.

Create a summary table of the regression results. Make sure you include the variable name, the estimated effect of that variable, its standard error, the test statistic, and the p-value. With that information, test these null hypotheses and write the three-part conclusion for each:

- H_0 : The level of honesty in the government has no effect on the GDP per capita when OPEC membership is taken into consideration.
- H_0 : When taking the level of honesty in the government into account, OPEC membership is independent of (has no effect on) GDP per capita.

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 reading and math SAT scores are 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? 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.
- What is the predicted gpa for each of the four students described in Table 1?

Student	SAT Math	SAT Reading	Gender	High School	College Level
Tom	800	800	Male	Public	Freshman
Tim	800	800	Male	Private	Freshman
Jan	800	800	Female	Public	Freshman
Sam	800	800	Female	Private	Freshman

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

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? What 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!

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 students for the Fuffing Activity.*

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