

Regression III: Lab 1

Dave Armstrong
University of Western Ontario
Department of Political Science
Department of Statistics and Actuarial Science (by courtesy)

e: dave.armstrong@uwo.ca
w: www.quantoid.net/teachicpsr/regression3/

The lab exercises below are intended to get you working with categorical independent variables and interactions in a user-friendly environment. Chris, Nick and I will be walking around to help when issues come up. You should feel free to work through the lab in whatever order you want so as to maximize the extent to which you are practicing the things that are most interesting to you. There are two sets of exercises below, but both ask similar questions. Alternatively, you could work on your own data trying out the same kinds of things we talked about in class.

Before starting, do the following:

```
install.packages("devtools")
library(devtools)
install_github("davidarmstrong/damisc")
install.packages("factorplot")
library(DAMisc)
library(factorplot)
```

1 Question 1

Do the following in R:

```
library(foreign)
dat <- read.dta("http://www.quantoid.net/files/reg3/lab1_nes.dta")
```

These are data from the American National Election Study. The variables included are:

pid	7-point party ID variable (Strong Dem=1, Strong Rep = 7)
pid3	3-point party ID variable (1=Dem, 2=Ind, 3=Rep)
demtherm	Democratic feeling thermometer
reptherm	Republican feeling thermometer
difftherm	Difference between democratic and republican thermometers
age	Survey respondent's age
income	23-category income variable
race	Un-recoded race variable
racerec	Recoded race variable
libcon	Liberal-conservative ideology (smaller = more liberal)

Estimate a model of the democratic party feeling thermometer (demtherm) scores on pid3, racerec, and the interaction between age and libcon (also make sure you specify it so **R** includes the main effects).

1. Is the interaction necessary?
2. Plot the conditional effects of both `age` and `libcon`
3. How do you interpret each of the plots?
4. What would you do differently if, instead, the interaction were between `pid3` and `age` with `libcon` as the control (i.e., not in the interaction)?

2 Question 2

The data you will use below come from the World Bank. Do the following to read in the data:

```
dat <- read.dta("http://www.quantoid.net/files/reg3/lab1_wb.dta")
```

If you want to see the variable descriptions, you could do the following:

```
library(DAMisc)
searchVarLabels(dat, "")
```

In the exercises below, I want you to use life expectancy (`life_exp`) as the dependent variable.

1. Estimate a model of `life_exp` on civilization codes (`civ2`), percentage of the total population living in urban areas (`urban_pct_total`) and the interaction of the natural logarithm of GDP/capita (PPP) (`loggdp_pc_ppp`) and primary school expenditures as a percentage of GDP (`expend_prim`).
 - (a) Is there a significant interaction?
 - (b) If so, what is the nature of that interaction? What sort of inferences make sense from the data?
 - (c) How would you present the results of the `civ2` variable and how would you talk about them?
2. Estimate a model of `life_exp` on primary school expenditures as a percentage of GDP (`expend_prim`), the percentage of the total population living in urban areas (`urban_pct_total`) and the interaction of the natural logarithm of GDP/capita (PPP) (`loggdp_pc_ppp`) and civilization codes (`civ2`).
 - (a) Is there a significant interaction?
 - (b) If so, what is the nature of that interaction? What sort of inferences make sense from the data?