Regression III: Lab 2

Dave Armstrong
University of Wisconsin – Milwaukee
Department of Political Science

e: armstrod@uwm.edu
w: www.quantoid.net/ICPSR.php

This lab exercise is meant to get you thinking about the nature of multilevel data and how explanations of phenomena in situations like this can be of both the "between" and "within" forms. We're using data from the most recent wave of the World Values Survey (2005-2008) for 38 countries. The following variables are in the dataset:

Table 1: Variables in `wvs_lab3.dta`

nation	Country Name
satisfied	Overall life satisfaction (1=low, 10=high)
org.mem	Number of active organizational memberships
married	Coded 1 if respondent is married
ind.democ	Country is governed: (1=non-democratically, 10=democratically)
weekrel	Coded 1 if the respondent attends church weekly
ccode	COW country code
civlibs	Civil Liberties (from Freedom House Data)
gdppc	GDP/capita

In today's exercise, we are going to be especially interested in the extent to which both individual and country-level factors explain overall life satisfaction.

1. In an exploratory way, consider some individual factors that may influence life-satisfaction (so use `satisfied` as the dependent variable). Use the `lmList()` command to estimate different linear models for each country and investigate the heterogeneity of models. Use `married`, `org.mem`, `ind.democ`, and `weekrel` as the independent variables. What does this tell us about the likely need for random parameters in our multilevel regression model?

   HINT: The `lmList()` command is in the `lme` library and will want a data frame that has no missing data. To accomplish this, use `newdata <- na.omit(dat)`, assuming your data frame is called `dat`.

2. Now, estimate the between-unit regression of satisfaction on the variables you used above *and* GDP/capita, and civil liberties. Given the relatively few observations here, it might make sense (as we're simply in exploratory mode here), to estimate a different "between" regression for each variable you're interested in. To do this, you will have to create a country-level dataset of country means of the variables of interest. What does this tell you about the between-effects you might want to include in the multi-level model?

   HINT: To make the country-level dataset, first, you should get all of the variables you want into a matrix (`X` below). Then, you can do (assuming your dataset is called `dat`):

```
by.X <- by(X, list(dat$nation), apply, 2, mean, na.rm=T)
between.dat <- as.data.frame(do.call(rbind, by.X))
```

3. Estimate a multi-level linear model with `satisfaction` as the dependent variable, `married`, `org.mem`, `ind.democ`, and `weekrel` as level-1 variables and `gdppc` and `civlibs` as level 2 variables. How do the between effects differ for `civlibs` and `gdppc` here as opposed to the between regression from the previous step?

4. Now add the mean of `ind.democ` and an interaction between the mean of `ind.democ` and the level-1 `ind.democ` variable to the model above. Compare the two models using `anova()`. What does this tell you about the importance of random slopes on `ind.democ`?

   HINT: To get the country-mean from the between data above back into your dataset, again assuming your dataset is called `dat`, you can do:

   ```
   dat$mean_ind.democ <- between.dat$ind.democ[match(dat$nation, rownames(between.dat))]
   ```

5. Use `amelia` from the `Amelia` package to impute the missing data in the original dataset. Then, use `zelig` from the `Zelig` package to estimate the model using these data. How do these results compare to those from the models above?