

# Advanced Quantitative Methods

## Homework 5

Johan A. Elkind  
jos.elkind@ucd.ie

Due May 5, 2014, 5pm

*Please submit by email in PDF format. Add R code in a separate .R file, or SPSS code in a separate .sps file, or Stata code in a separate .do file, or the code for any other package you use separately. Note that if you do not use Latex, there is an "Equation" entry in Microsoft Word under "Insert" that will allow you to include matrices and other mathematical equations.*

(10%) of the grade is used for an overall evaluation of the clarity and presentation of your code.

### Data and instructions

For all questions, we will use the `asiabaro.dta` (AsiaBarometer) data set. Note that this file has been updated when this homework was written, so if you downloaded it earlier, make sure you get the latest version.

- Make sure you produce tables and/or plots for yourself to see the distribution of each variable, to make sure they are suitable for use in regression.
- Also make what you deem to be reasonable decisions whether variables should be treated as categorical or continuous and verify that it has been treated as such in your regression output.
- Note that the `recode()` function in the `car` library can be very useful when transforming variables.<sup>1</sup>
- Note that adding `factor()` inside the regression equation can help to inform R that it concerns a categorical independent variable, e.g. `m <- glm(factor(eu.good) ~ birthyear + factor(gender), data = ees, family = binomial(link = "logit"))`
- Code for any transformations to variables should be included in the R file. When running your R file with the supplied `ees.dta` file, I should obtain the same results you produce in the PDF submission.

---

<sup>1</sup><http://rprogramming.net/recode-data-in-r/>

- Note that all plots and tables should have full captions and should be of publishable standard.

## Questions

1. We will explain the level of `pref.democr` by interest in politics, satisfaction with democracy, evaluation of corruption levels, the level of inequality (`gini`) and the actual level of democracy (`polity`).
  - (a) Estimate an OLS model with the above variables.
  - (b) (5%) Estimate an OLS model with the individual level variables and country fixed effects (use `lm()`).
  - (c) (5%) Estimate a multilevel (random effects) model, without explanatory variables and variances at individual and country level (use `lmer()`).
  - (d) (5%) Add individual-level variables to the model from 1c.
  - (e) Add country-level variables to the model from 1d.
  - (f) (5%) Produce a regression table of publishable standard including all five regressions.
2. We will explain the level of turnout by interest in politics, satisfaction with democracy, sex, age, and the actual level of democracy.
  - (a) Estimate a probit regression model with the above variables, including country fixed effects (use `glm()`).
  - (b) (10%) Using simulated parameters, calculate a 95% confidence interval for the predicted probabilities of turnout in the last elections for someone with high political interest and for someone with low political interest, keeping all other variables at reasonable values.
  - (c) (10%) Using simulated parameters, create a plot of the probability of turning out in an election by age, with 90% confidence intervals.
  - (d) (5%) Re-estimate the model using country random effects (use `glmer()`).
  - (e) (10%) Using a bootstrap procedure with the random effects model,<sup>2</sup> calculate a 95% confidence interval for the predicted probabilities of turnout in the last elections for someone with high political interest and for someone with low political interest, keeping all other variables at reasonable values.
  - (f) (5%) Produce a regression table of publishable standard including both regressions.
3. Write a 600 word essay (30%), interpreting the full results of above exercises.<sup>3</sup>

---

<sup>2</sup>This is likely to be slow – try with few bootstrap iterations until you are satisfied the code is working.

<sup>3</sup>Note that the usual negative marking will not be applied, but rather a grade for the overall quality of the essay and the interpretation will be converted to the 30 point scale.

### Grade conversion scheme

Score	Grade	Score	Grade	Score	Grade	Score	Grade
97-100%	A+	85-87%	B	74-76%	C-	54-64%	E+
94-96%	A	83-84%	B-	71-73%	D+	44-53%	E
91-93%	A-	80-82%	C+	68-70%	D	33-43%	E-
88-90%	B+	77-79%	C	65-67%	D-	0-32%	F