

Advanced Quantitative Methods

Homework 4

Johan A. Elkind
jos.elkind@ucd.ie

Due April 21, 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 `ees.dta` (European Election Survey) data set.¹

- Note that the data in this file is not "cleaned", such that missing values are included as regular values (e.g. year of birth 7777) or value labels (e.g. "Don't know").
- 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.²
- 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"))`

¹It might happen that the variable names will have their punctuation mark replaced with an underscore.

²<http://rprogramming.net/recode-data-in-r/>

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

Questions

1. We will perform regression analyses to explain turnout in European elections (ep.turnout), investigating in particular whether political interest is a good explanatory variable.
 - (a) (5%) Recode the interest variable such that it is a dummy variable, grouping the categories "Very" and "Somewhat" as 1.
 - (b) Run a probit regression regressing turnout on interest in politics, controlling for gender, social class, and age.
 - (c) (10%) Select three additional control variable and provide a reason for each why this is a suitable control.
 - (d) Repeat the regression adding these control variables.
 - (e) (10%) Calculate predicted values for the probability of turnout in European elections by the two categories of the political interest variable, keeping all other variables at reasonable values.
2. We will continue with the same dependent variable, but focusing on religiosity as the key independent variable, the hypothesis being that more religious voters are more likely to turn out in elections.
 - (a) Run a logistic regression regressing turnout on religiosity.
 - (b) Repeat, controlling for gender, age, and social class.
 - (c) (10%) Produce a plot of predicted probabilities to turn out in European elections as a function of religiosity, keeping the other variables at reasonable values.
 - (d) (10%) Produce a receiver-operating characteristic (ROC) curve to evaluate the predictive performance of the model.³
3. Write a 600 word essay (30%), interpreting the full results of above exercises,⁴ including the results from all four regressions presented in a table of publishable standard (15%).

³<http://www.unc.edu/courses/2010fall/eco1/563/001/docs/lectures/lecture22.htm#ROC>

⁴Note 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