MPP/MIA C5 - Statistics I

Final Data Analysis
Unemployment and Turnout
Due 11 December 2015, 5PM

Due by 11 December, 5PM. Please place a hardcopy in the TA mailbox <u>and</u> upload a pdf to the "Final DA" dropbox on the Moodle. File names should start with the student's surname, e.g. "JonesFinalDA.pdf". Details on the location of the TA mailbox will be given in class in Week 12. Please keep your write-up to no more than 8 pages in length and attach your do-file as an appendix. The do-file does not count toward your page limit. Do not include extraneous output irrelevant to your argument. For example, regression tables made with `esttab` *or* `estout` *are a better use of space than pasting in Stata output. This counts as an exam. All work must be done individually. Notes and books can be consulted. Other students may not be consulted.*

1 Introduction

General elections have last been held in Germany in 2013. While the overall distribution of the currently 630 seats in the German *Bundestag* is determined by Proportional Representation, 299 of these seats are allocated to candidates who received the most votes in an electoral district.[1] The dataset `btw2013.dta` contains the official electoral results of the 2013 election by electoral district. The results have been merged with socio-economic and demographic data on the districts. In this data analysis you will analyze the relationship between unemployment and turnout. Turnout will be your dependent variable and unemployment will be your main independent variable.

---

[1] For those of you fascinated by electoral systems: the German system is a so called mixed-member proportional system.
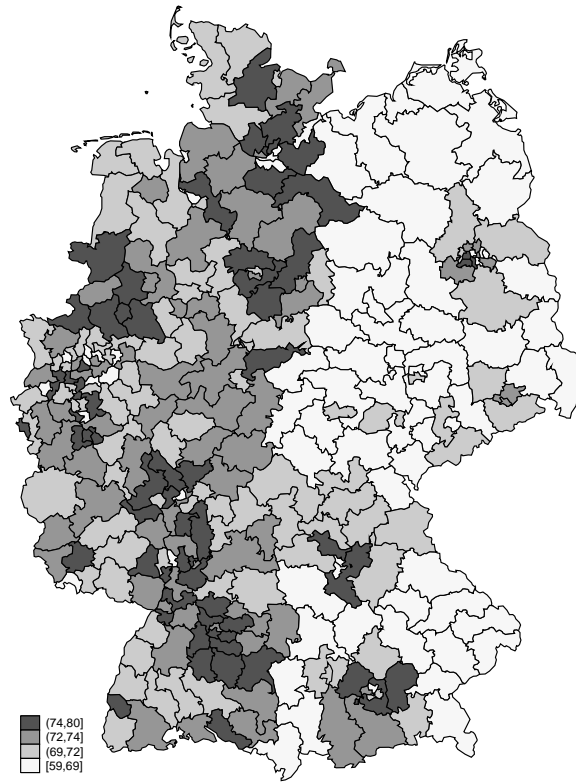
Figure 1: Turnout in electoral districts in the 2013 German federal election

2 Questions

1. **Theory and Model Development.** You are interested in the relationship between turnout in elections and unemployment. Turnout (% of eligible citizens that participated) in an electoral district is your dependent variable and the unemployment rate (% of working age population registered as looking for a job) in an electoral district is your main independent variable. All other variables, if you include them in your model, are control variables.

   (a) Express the relationship between your dependent and your independent variables as a hypothesis.

   (b) Express the appropriate null hypothesis.

   (c) Choose and name two control variables and justify their inclusion in the model.

   (d) Write down the regression equation of your model.

2. **Analysis.** Run your proposed analysis in `Stata`.

   (a) Start with a simple bivariate model regressing turnout on only the unemployment rate. Create a scatter plot that includes the fit line and confidence interval for this simple bivariate model.

   (b) Provide a table detailing your regression results. The table will contain two models, the simple bivariate model and your more elaborate multiple regression model. The `esttab` or `outreg2` command can be used for this purpose. Do not copy and paste `Stata` output. Your regression table must include information on the sample size, R^2 and $adj.R^2$.

   (c) Discuss the results in light of your expectations detailed in Question 1a. Evaluate both substantive and statistical significance of your estimate of the coefficient of unemployment.

   (d) Report and interpret your model's R^2. You are primarily interested in the relationship between unemployment and turnout. Should you worry about the magnitude of the R^2? Why or why not?

   (e) There is a big worry in Germany that once people lose their jobs they also lose a lot of their social ties and in consequence also do not participate in elections anymore. What do your results contribute to answer this question: Does losing their job make a citizen less likely to vote?

3. **Diagnostics.** Based on the model you ran in 2, carry out the following diagnostic tests:

   (a) Heteroskedasticity. Conduct and briefly discuss one (and only one) of the appropriate diagnostics for heteroskedasticity.

   (b) Outliers. Conduct and briefly discuss one (and only one) of the appropriate diagnostics for outliers. What are the potentially influential outliers?

   (c) Name and explain the reason for one (and only one) other diagnostic you should run before you place confidence in your results above. (Multiple answers could be correct here).

4. **Region Effect.** Is there a difference in turnout between East and West Germany?

   (a) Create a dummy variable called *East* which indicates whether an electoral district is in East Germany or not. The East German states are Brandenburg, Mecklenburg-Vorpommern, Thuringia, Saxony and Saxony-Anhalt. Drop all Berlin districts from the dataset – as a formerly divided city Berlin's place in the East/West divide is ambiguous.

(b) Provide box plots that illustrate the difference graphically. How do you interpret them?

(c) Carry out a difference in means test to determine whether turnout differs between East and West Germany. Provide a small table containing the means, standard errors and the details of the test.

(d) Is there a statistically significant difference between the two? Explain.

(e) You know that East and West Germany, even more than 20 years after reunification, still differ in a lot of regards, not just turnout. Estimate a model that tries to take these differences into account. Report the model in a regression table. What do you conclude about the difference in turnout between East and West Germany?

5. **Interaction.** You suspect that the turnout difference between East and West is not only partly determined by unemployment but that the effect of unemployment is also different within the two regions. You ask: Is the correlation between unemployment and turnout conditioned by region?

(a) Model an interaction effect between unemployment and the East dummy in `Stata`. For now, do not include any other predictors. Provide a table of your results.

(b) What is the effect of unemployment on turnout in East Germany? What is it in West Germany?

(c) Visualize your interaction model. Plot two fit lines, one for the relationship between turnout and unemployment in the East and one for the same relationship in the West – in one graph. Include confidence intervals!

6. **Robustness.** Report all models you estimated so far and which include unemployment as independent variable in a regression table. Also include one new specification to test the robustness of your results. What do you conclude about the relationship between unemployment and turnout?