PS699: Problem Set 4

In this assignment, in any statistics package you like, and using our data set, you will regress the lower-house proportionality index (LPROP) on the (country's average) lower-house electoral-district magnitude (LMAG) for all available data points.

Background info: The proportionality index provides a summary measure for the entire set of parties in a country of the discrepancy between the percentage of total votes that each party receives and that party's percentage won of the total seats in the (lower house of the) legislature. If these percentages match perfectly for all parties, this proportionality index is at its maximum of 100. An electoral district's magnitude is the number of lower house representatives (MP's in most cases) elected in that district. One representative per House district in the US. The Netherlands, at the other extreme, has only one electoral district: all Dutch MP's are elected from a single, national, electoral district. LMAG is the average district-magnitude for the country's entire set of electoral districts. Germany is a strange case which elects to the lower house from a two-ballot system with separate effective electoral districts for each ballot. Each voter gets two votes. The first vote is cast in 1-MP-per-district plurality (most votes wins) contests and there are about 260 such districts. The remaining roughly 260 seats are filled from second votes cast in one 260-magnitude national district (in "strict party list" balloting) in a manner expressly intended to bring the overall seat proportions in the parliament into line with parties' shares of the national vote.

1. Your hypothesis is, obviously, that district magnitude and vote-seat proportionality are related somehow. State specifically a hypothesis of this sort empirically evaluable by the linear regression suggested above.

2. Given your theory and the goal of evaluating this hypothesis empirically, decide how to code LMAG for Germany. Briefly (very briefly, a paragraph should do for our purposes) describe and defend that decision.

3. Regress LPROP on LMAG and present your results. Print your stats-package results, but also arrange those results appealingly as you would in a table intended for professional presentation and show that appealing table.

4. Do you believe that LMAG and LPROP are related in levels (i.e., each 1-unit change in LMAG produces a β-unit change in LPROP? Or do you think perhaps that percentage changes LMAG produce level changes in LPROP (i.e., each 1% change in LMAG produces a β-unit change in LPROP)? Or that absolute changes in LMAG produce percentage changes in LPROP (i.e., each 1-unit change in LMAG produces a β% change in LPROP)? Or that percentage changes in LMAG produce percentage changes in LPROP (i.e., each 1% change in LMAG produces a β% change in LPROP)? Pick one (defend it in at most a paragraph).

5. Estimate that regression and present your results (print your stats-package results and an appealing table).

6. Using these results (from 5), give a **substantive description** (i.e., what does it mean in terms of your theory and the substance in this instance) of each of the following and give a formula and/or describe formally how each of the following is calculated **mathematically**.

a) The estimated coefficient, *b*, on LMAG or *ln*(LMAG) (whichever you used).

b) The standard error (i.e., estimated standard deviation of the sampling distribution) of that coefficient estimate.

c) The t-statistic for the Wald test of the null hypothesis that this coefficient is zero, the associated p-level, and the related confidence interval.

d) R^2 and the estimate of σ_ε^2 (the "Standard Error of the Estimate" or "...of the Regression," or whatever your stats package calls it)

e) "The F-statistic for the regression" and its associated p-level. State the null hypothesis that this statistic tests. What is the relationship between the t-stat in c) and the F-stat here in e), in this bivariate-regression case? (Give answers to this last question both in terms of the null hypotheses to which they are addressed in this case and in terms of the relationship between the distributions of the test statistics.)