Assignment 1

Corruption and Wealth (Linear model recap)

1. Political and economic corruption is a vexing problem throughout the world. How does the extent of corruption change with the wealth of countries, if at all? Transparency International provides estimates of corruption (based on surveys). The index ranges between 10 (highly clean) and 0 (highly corrupt).

   The data are in the file `Corruption.dta` from the course website (saved in Stata format) and contain the average corruption index between 2000 and 2007 `[ti-cpi]` as well as the (per capita) wealth variables in 2002 `[undp-gdp]`.

2. Run a regression of corruption on GDP per capita. Estimate the model twice: (1) using OLS, (2) using Maximum Likelihood (that is, write a likelihood function and maximize it). Interpret the intercept and the regression coefficient.

3. Interpret the *substantive relevance* of the results. To do so, predict the level of corruption in countries with GDP per capita that corresponds to the 25th percentile and a country with GDP per capita in the 75th percentile.

Maximum likelihood I

1. Why do logit and probit coefficient estimates always differ?

2. What are predicted values from a binary model? How do they differ from expected values?

3. You observe the following distribution of exam results in your class:

Student 1	fail
Student 2	pass
Student 3	fail
Student 4	pass
Student 5	pass
Student 6	fail
Student 7	fail
Student 8	fail
Student 9	pass
Student 10	fail
Student 11	pass
Student 12	pass
Student 13	pass

Choose a suitable probability distribution and write down the likelihood function.

Maximum likelihood II

The file `turnout.dta` contains individual level data on voter turnout in Finland collected by the Comparative Study of Electoral Systems. The following variables are included:

- Did an individual vote in the 2003 parliamentary election? `[voted]`
- Individual background characteristics: age in years `[age]`, education (levels, include it as continuous variable) `[edu]`, and gender `[female]`.
- A measure of political information, created by summing up correct answers to four factual political questions `[informed]`.

Someone claims that political information has a positively monotonous relationship to an individual's propensity to vote. To test this proposition, fit a logit model controlling for basic individual characteristics.

1. Choose *one* coefficient and interpret it.
2. Calculate the predicted probability of turnout for a young woman `[age==25]` with high education `[edu==3]` and average political information.
3. Plot the effect of political information at low `[0]` and high `[3]` education levels.