Homework Set #7
Due: Tuesday, March 26th by 5pm.

Do the following in Kmenta: 8-4 & 8-5.

**Question 1 (Computer)** Download the data set 201hwk07.dta from the class web site. These are data on state interest group populations in 1990. Consider the following basic model:

\[ \text{numgroups}_i = \alpha + \beta_1 \text{init}_i + \beta_2 \text{GSP}_i + \beta_3 \text{ideology}_i + \epsilon_i, \]

which we assume meets the standard OLS assumptions. For the questions that follow you should report the results of the relevant regression model you run as well as provide information about the hypotheses tests that are performed. You should also include the graphs that you generate. Rather than have you merely print out the Stata results, I want you to report them as if you were preparing a research article. Include whatever information you consider appropriate.\(^1\)

1. Estimate this model and report your results. Test the (alternate) hypothesis that the initiative process increases the number of interest groups that register to lobby. Report the \(p\) value of your test. You can run a regression in Stata by typing `regress depvar indvars`.

2. Gray and Lowery’s theory of interest group populations asserts that the effect of GSP (gross state product) is non-linear. Specifically, interest group populations exhibit density dependence whereby the marginal effect of GSP is increasing, but at a decreasing rate. Test for evidence of density dependence by adding a variable GSP\(^2\) to the basic model.

3. Is there evidence of density dependence: is the marginal effect of GSP increasing, but at a decreasing rate? When is it increasing? Is it positive? When?

4. Now do a test for improvement in model fit based on the results from the basic (restricted) model and the one you just estimated. Do these results differ from those you just reported? Why?

5. Test whether the marginal effect of GSP is significant. Is it always significantly different from zero? You can perform this test using the `test` command: `test gsp + 1*gspsquared=0` tests whether the linear combination \(\beta_{\text{GSP}} + 1 \times \beta_{\text{GSP}^2} = 0\) (in Stata using the variable name in the test command tells it which coefficients to use).

\(^1\)You can make your life easier if you learn the `outreg` command — look into the `append` option.
6. A pesky reviewer has suggested that it is necessary to control for whether a state is in the South before the paper can be accepted for publication. How would you respond to this assertion?

7. A good researcher always looks at his residuals. Construct residuals and graph them against the fitted values of your dependent variable. You can generate values of \( \hat{y} \), using the `predict` command: `predict yhat`. Then, once you have predicted residuals, you can graph them against \( \hat{y} \) thusly: `scatter ehat yhat`. Save the graph using the menus (or `scatter ehat yhat, saving(201hwk07-graph01)`). Do you see any problems? What are they?²

8. What could you do about the problems identified in the previous part? Try dropping the offending observation(s) from your regression analysis using the `if` command in your regression statement. Re-run the model and graph the new residuals against the new fitted values (you will have to give them a new name or `drop` the old values). How has the graph changed? How have the regression estimates changed? Has the model fit improved?

²A nice way to figure which residuals are for which observations is to label them in the graph command: `scatter ehat yhat, mlabel(state).`