Paper Assignment

The paper assignment is to describe a hypothesis, and test it using quantitative data and the techniques covered in this class. There are four goals of this assignment: 1) clearly describe a hypothesis, 2) locate and extract data to test the hypothesis with, 3) perform appropriate analysis to test your hypothesis, 4) clearly write up the results of your test. You need not test an original hypothesis. You can test a hypothesis that has already been demonstrated in the literature. In fact, trying to replicate a published piece of research is an excellent way to learn about doing quantitative analysis. Of course testing a hypothesis of your own can be more fun.

FIRST DUE DATE: You must have your topic approved. Please turn in a statement of your hypothesis (anywhere from 1 paragraph to 1 page), and some indication of the data you will be using to test it, by 3pm, Friday, October 24. You are certainly welcome to turn this in earlier.

LENGTH: Your paper should be long enough to meet the four goals described above. It is hard to believe you could do this in 6 pages plus 2 tables. It is almost as hard to believe that you will need 30 pages plus 12 tables. A suggested length would be 12-15 pages, plus 6-10 tables. A page has 250 words on it. If you think you are going to be way over this or under this, you should probably see me.

Here are some suggestions:

1 Style

1.1 Read “Coding Style and Good Computing Practices”

You are going to be coding and computing, so read this article. Turning in a paper where variables are poorly labelled when you have been assigned an article (written by the professor!) that explicitly tells you not to do that is bad form.

1.2 Read APSA Style Manual on Tables

Tables should be clearly labelled. The first thing to look for is to make sure that you label the dependent variable! Then, be sure to include: the number of observations, standard errors or t-statistics for your coefficients, at least one measure of goodness of fit. It is a good idea to have a note to the table that says “Table entries are .....” The rule is: the table should be able to stand alone.
2 Writing - Define Your Terms

If you use the term “gender gap”, you should define it. It could mean: 1) the gap in democratic and republican vote shares between men and women voters; 2) the difference in turnout rates between men and women, or 3) the difference in interest in the election between men and women. You need to pick one. Any term used in a scholarly paper should have one, unambiguous meaning.

3 Theory

3.1 Introduce and/or Justify Your Model

Why are you including the variables you are including on the right-hand side? Is there some theory? Are they just usually there in the literature?

3.2 Be Crystal Clear as to Your Expectations for Each Coefficient

The best way to test your model specification is to see if you have a clear expectation as to the sign of each of your coefficients. If you don’t know whether you expect an estimated coefficient to be positive or negative, then you haven’t thought through the model well enough.

3.3 Assumptions

Discuss what assumptions you are making in your choice of estimation technique. Discuss the justifications for the ones which are most troubling. How bad would it be if the assumptions are wrong?

4 The Model

The model has to fit the theory! This is a key point in grading. You need to be able to explain why the model you write down (i.e., your choice of left-hand side and right-hand side variables, as well as the functional form) is the appropriate model to test your theory with. This does not mean you have to have a perfect model - you can’t. But you should explain the limitations of your model. And you should explain why it will offer a test of your theory.
5 Results

5.1 How Many Tables?

Most students want to know how many tables they should have. There is no single answer to this question. But a suggestion is that you should have at least 4:

- **Table 1**: Summarize the data you are using.
- **Table 2**: Describe some bivariate relationships.
- **Table 3**: Results of the multivariate (regression) analysis.
- **Table 4**: Answers to some ‘what if’ questions. Here you should estimate predicted effects of some interesting changes in the independent variables.

5.2 What Happened to the Coefficients of Primary Interest?

Did the coefficients of primary interest reach statistical significance? Were they the expected sign? What about their magnitude?

5.3 Other Coefficients

Did the other coefficients make sense? Any surprises?

5.4 Fit of the Model

How does the model fit? If this is OLS, what is the standard error of the regression? How does it compare to whatever range the dependent variable takes?

5.5 Be Very Honest

Don’t try to hide anything. And don’t try to “explain” what you don’t have an explanation for. (This is not to say you should not offer conjectures: but don’t offer nonsense explanations for genuinely baffling results.) If you got very anomolous results with minor specification changes: report that. You are not writing the last paper on any subject: your data analysis is one of many pieces of analysis that will get lumped together and let us put together some picture of reality.
6 Things Students Forget!!!

- Label the dependent variable very clearly on tables and figures!
- Report the number of observations for every regression.
- Do not report F-statistics for the whole regression if any of your t-statistics are over 2.
- Present evidence for ALL claims!!
- Use graphs where appropriate to convey information about the data. For instance, histograms or frequency graphs describe a variable better than just a table of mean and standard-deviation. So a histogram or frequency plot for the dependent variable or key independent variable is useful.
- Do not refer to variables by 6 character names in the text; use real, meaningful names.
- Use labels with direction: “conservative ideology”, not “ideology.”
- Compare RMSE to the standard deviation of the dependent variable, not the range of the dependent variables.