

Final Data Analysis Project

General instructions: Unlike the first project, *you must do this project on your own*. Although you may discuss your project with other students, I expect each student will have a distinctive topic and will do all her or his own work.

The final project is due at the time of the final exam: Wednesday, December 5, 9:10 a.m. There will be absolutely no extensions and I will not accept late projects. I encourage you to try to get it done early.

1. *Basics:* The final data analysis project will be on a topic of your own choosing, applying regression analysis to actual data. The finished product should consist of a brief report on your topic, regression results, and interpretation of findings. You should also prepare a brief presentation based on your project for the last week of class.
2. *Topic:* Try to pick a topic that interests you but can also be examined using regression analysis. A good topic can usually be summarized in terms of a single question. For example, in Part 3 of the first data analysis project we tried to answer the following question: “Controlling for the effects of schooling and work experience, are women’s and men’s earnings different?”

I encourage you to think of your own topic, but please ***discuss your idea with me*** before you settle on it. I can point you toward data and readings. Here are just a few ideas.

- Earnings differences by race and/or ethnicity and/or immigrant status
 - Effects of marital status on earnings
 - Effects of immigration on local labor markets (e.g. on wages and employment)
 - Effects of labor-market policies (such as transfer programs or regulations) on wages and employment
 - Effects of globalization (extent of trade) on labor markets
 - Sources of earnings inequality in the United States
 - Determinants of the labor force participation decision
 - Effect of the wage on individual hours of work
 - Effects of recent changes in the minimum wage
 - Effects of “living wage” ordinances
3. *Data:* Your data could be individual-level data, such as we have used in class, or cross-sectional data for states or countries, or time-series data. An example of the use of cross-sectional data would be to examine the determinants of work hours or labor-force participation rates using data on wages, incomes, hours, etc. from a cross-section of countries. Cross-sections of states or countries can also be useful for examining the effects of alternative labor-market policies.

4. *Data analysis:* Your analysis should definitely include multiple regression analysis of the data. But before running regressions, you may want to provide some simple descriptive statistics of the variables you are using, such as mean/median/s.d. In running the regression(s), make sure to think carefully about what variables should be included. For instance, even if your topic does not directly involve gender differences, you might want to control for gender in your equation.
5. *Topic proposal:* You must submit a written topic proposal by **Thursday, November 8** (note that this is a *week earlier* than what's on the syllabus). Make sure you have checked your topic out with me before submitting the proposal. The proposal should include:
 - Title
 - Description of question / issue (a paragraph or two)
 - Bibliography with at least 2-3 sources on your topic
 - Source of data
6. *Presentation:* Prepare a 5-10 minute presentation on your project, to be given during the last week of classes.
7. *Final Report:* Your final report should consist of two parts: a typed text, and supporting tables and figures. Each part is limited to 5 pages, so the total length of the report is no more than 10 pages.
 - I. *Text of report:* The text portion of the report should be *double-spaced* in at least 10-pt. type, and *no longer than five (5) pages*. The text should include the following sections:
 - **Executive summary:** *one brief paragraph* summarizing the question and the findings.
 - **Introduction:** What is the basic topic? What question(s) are you trying to answer? What does economic theory predict about the answer?
 - **The model:** What equation(s) are you estimating to answer your question(s)? (It may help to write out the equation.) What hypotheses will you test?
 - **Data:** Briefly discuss the data. Are there any measurement problems or other issues?
 - **Regression results:** Discuss the general validity of the regression, in terms of R-squared, residuals, etc. Then discuss the results relative to your model. Be sure to consider statistical and economic significance. Test hypotheses. Also, discuss any potential problems or sources of bias. Refer to tables and figures.
 - **Conclusions:** Implications of the regression results for answering your question(s).
 - **References**
 - II. *Tables and figures:* Attached to the text portion, please include carefully labeled tables and figures, including regression results, etc. *The tables and figures should be no more than five (5) pages*. Make sure to number all tables and figures for easy reference.
8. **Make sure your work is your own.** Discussing the project with other students is OK, but plagiarism is not. I encourage creativity and originality.