

### **LAB PROJECT INSTALLMENT 4: THE FINAL PAPER**

Each group should turn in one collaboratively-produced final paper. The text of this paper should consist of complete sentences assembled into paragraphs that constitute a coherent essay. Your writing should conform to the guidelines in the "Writing Style Guide" posted on Moodle, and you should follow APA style for in-text citation of sources and your reference list. Failure to follow these guidelines will result in a substantial reduction in your grade for the project.

You may incorporate work from the previous installments of this project in your final paper—indeed, those assignments were intended to build a foundation for this final paper. You are allowed to reproduce passages of any length from the previous installments *verbatim* in the final paper, without citing the previous installment or enclosing those passages in quotation marks. However, you should be sure that any passages from earlier assignments that you use in the final paper are in a form that is appropriate and logical for the final paper. At a minimum, be sure to edit material from previous installments in such a way that it fits in logically and stylistically with the rest of the paper. If you have made any changes in your topic or how you approach it since the time you wrote a previous installment, be sure to modify the material you use from the previous installment in any ways necessary to reflect those changes. And be sure that any material you adapt from previous installments reflects any comments or suggestions I have given you. (You will notice that some of the instructions below were taken *verbatim* from the instructions for previous installments of this project.)

As indicated in the assignments for the previous installments, the nature of your projects differs across groups, so it is impossible to give one outline of the final paper that will apply exactly to all groups. The structure suggested below is meant to give you an idea of what the critical components of the paper are, and to present a standard way of organizing them. If you believe that some adjustments to this suggested structure would give you a better way of presenting your project, you should make adjustments accordingly.

#### ***Suggested structure for the text of the paper:***

Introduction. The first paragraph of your paper should explain concisely what the goal of your project was: what were the main questions you set out to investigate? The introduction should also explain the impetus or motivation for your research: was it motivated by a current public policy issue? an ongoing theoretical debate that might be elucidated by some empirical evidence? results of previous empirical research that led you believe require further study? reports about your topic found in the popular press?

The introduction should also mention the kind of data you used and the kind of analysis you did.

The introduction should include a brief overview of the main findings of your study (which you will of course present and discuss in more detail later in the paper). For some kinds of assignments in some subjects, students are told to present their conclusions at the end of the paper, after they have constructed an argument and marshaled the evidence upon which the conclusions are based. Do not take that approach in this paper: give a concise statement of your main results and conclusions in the introduction.

Previous research. You should discuss a selection of several of the most important and relevant previous studies on your topic. Most of the previous research you cite should be primary sources that report the results of original professional research. In most cases, these sources will be articles from academic journals, books, book chapters, or reports issued by government agencies or non-profit organizations that conduct research of high quality. In some cases, it may be useful to refer also to sources other than professional research studies, such as newspaper or magazine articles, or web sites of think tanks and non-profit organizations. Sources such as these cannot substitute for appropriate references to scholarly research, but they may complement the professional studies you cite.

It is almost always better to organize your literature review by issue or topic than by source. Think about the issues raised by the variety of sources you have read, and organize those issues into a logically structured essay. Your references to specific previous sources should then be embedded in this structure: in the discussion of each issue you choose to highlight, mention (using logically organized sentences and paragraphs rather than just a list) the sources you have found that are relevant to the specific topic, and explain what is interesting and relevant about those previous works (e.g., what their main conclusions were, what statistical or analytical methods were used, the data used in the previous studies, etc.) Simply discussing in sequence one article or book at a time is rarely a good way to organize a literature review.

If you mention a previous study in your literature review, you should say enough about it to let the reader know why you have included it and what its significance is. At a bare minimum, that usually means stating what the questions addressed in the research were, what data were used for the study, and what the main findings were. Just saying that certain researchers did a study on a certain topic, without saying what they learned, usually does not add much. If you don't have more than that to say about a previous study, you should probably not include it in your essay.

In addition to saying something substantive about the questions asked and the main results of the studies you discuss in your literature review, you should say something specific about how your awareness of that study will affect your project: what did you learn from the study that will help you define and/or investigate your own topic. Simply noting that a previous study is on a topic that is the same as or similar to the one you plan to work on is not sufficient.

Data overview. In many cases, your data overview should consist of two parts: (i) information about the original data files you obtained to use for your project, and (ii) information about the final data set that you constructed from these original data files and used for your analysis.

*Information about original data files:* You should provide information about every original data file you used for your project.

You should begin simply by naming the source from which you obtained the original data file. If you are using a well-known source (like the Current Population Survey produced by the Bureau of Labor Statistics, or the World Bank's World Development Indicators), just naming the source may be sufficient. If the data source is not well known, it may be useful to provide some additional information about it. If there are different versions of a data source—data collected in different years, or different surveys focusing on special topics, for instance—you should indicate which version you used.

You should be sure to make clear what the unit of observation was for each of your original data files. Indicate the coverage of the data in terms of any relevant factors or categories like certain years, or countries, or counties, or baseball teams, or sectors of the economy. (For instance, “The data in *original-surdatt.xlsx* were collected in three rounds of the survey, conducted in twelve countries in 2002, 2005 and 20012. Those countries, and the number of respondents surveyed in each country in each round, are shown in Table X. Each row of the data in this file represents one respondent. In each country, samples were chosen independently in each of the three years of the survey, so each individual appears in the data only once.”)

For each original data file, you should indicate which variables you took from it. You should name and give a brief definition of every variable from the original file that you used in your study or used to create a variable that you used in your study. You do not need to mention any variables in the original file that you did not use in your study or to generate any other variables you used.

If there is any other special information about an original data file that a reader would need to know about to be able to understand what you did for your project, explain it. To decide what, if any, additional information of this kind you should provide, think about how you would describe your data to someone who has a basic idea of what working with statistical data is about, but who is not familiar with your project or the data you plan to use—maybe a friend who took this course last year. What would you have to tell that person to give her or him a concrete understanding of what your data consist of, where they came from, and what information you took from them for your project?

*Information about your final data set:* I use the term "final dataset" to refer to the data file (or perhaps files) that you created by cleaning, merging and editing your original data files as necessary to prepare them for the analysis you did for your project.

Your description of your final data set should make clear what the unit of observation is. (The unit of observation for your final data set could be different from the unit of observation of your original data files. In fact, if you have more than one original data file, they do not all necessarily have to have the same unit of observation.)

You should also name and define every variable in your final data set. And for every variable in your final data set, you should make clear which variable or variables from your original data files you used to create it, and what calculations and processing were necessary to transform the variable or variables from the original data into the variable in your final data set.

Discuss any other aspects of the variables in your final data set—like the units quantitative variables are measured in, and the definitions and coding of categories of qualitative variables—that would be useful for the reader to understand. You will provide additional technical details in the data appendix; in the data overview section of your paper, the objective is to give the reader enough information to understand the descriptive statistics and figures you will be presenting.

Descriptive statistics. Your data may include variables that readers need to know something about to understand your analysis and interpret your results. For example, if the distribution of a variable is highly skewed the right, you might want to use it in log form in the analysis. In that case, you might want to present histograms of the variable both in its original units and in natural log units, and explain that you will use the variable in log form to remove the skew revealed in the histogram of the data measured in its original units. Or, as another example, if you are looking at a country/year variable that, like GDP per capita, varies much more across countries than it does for any single country over time, it might be helpful to include in your descriptive statistics a figure that illustrates that fact. One more example: suppose you are using data from a survey that asked people to rate their overall satisfaction with their lives on a scale of 1 (least satisfied) to 7 (most satisfied). Since your reader probably doesn't have a sense of how people would respond to this question (are the responses about evenly divided among the 7 categories? are most responses in the range of 1-4, with responses of 5-7 being quite rare?), it would be helpful to present something like a bar graph, showing the proportion of respondents who gave answers in each of the 7 categories. All of the above are examples: what you need to do is to consider the variables you are working with, and think about any aspects of their distributions that need to be highlighted.

After presenting any relevant information about the univariate distributions of your variables, you should present some graphs and/or tables of descriptive statistics that illustrate relationships that exist among two or more variables. It might be revealing to present graphs and/or tables that show the relationship between two variables when you “control for” the value of a third variable. Like the graphical and numerical descriptive statistics you prepared for your preliminary analysis, these graphs and tables should be constructed in such a way that they give preliminary answers to the questions that your project was intended to address. Indeed, you should think ahead to the main results you will be reporting in the analysis section of the paper, and try to find ways to illustrate those results graphically or in tables of descriptive statistics.

Analysis and results. This is the section in which you explain the formal analysis you undertook and present the details of the results you obtained. In the individual meetings I will be having with each group, I will give you some guidance about the kinds of analyses that might be most appropriate and revealing given the data you are working with and the questions you are investigating.

As you describe the results, make sure that, in addition to reporting the numerical results that Stata produced, explain in words what those results mean and what their significance is for your project. Comment on any ways in which your results relate to anything you found in the previous literature on your topic.

Conclusion: Summarize the results that you reported in detail in the analysis section. Highlight those that are most revealing with respect to the questions you investigated, and explain the conclusions you can (or cannot) draw about those questions on the basis of your statistical results. Comment on how your results and conclusions relate to the previous studies you discussed earlier in the paper. Add any thoughts or observations that you think would be noteworthy.