SOC 504: Writing a replication/extension paper
April 20, 2017

Acknowledgements: This handout draws inspiration from materials by Kosuke Imai and Gary King. Contributors to prior versions: Elisha Cohen.

The most important reference for your paper is Publication, Publication by Gary King.[1] Note in particular points 4-9 in the Elements of the Paper section. The article itself is a great guide to writing a clear and concise paper and you should absolutely read it a second time before beginning to write. A key piece of advice from the supplement to that article is:

> This is probably the first paper you are writing that is not about you: Explaining how hard you worked, that you included everything the professor asked for, and how smart you are are now all irrelevant distractions. Your goal instead is to construct a paper that makes people want to read it. Your title, abstract, and paper each must focus on what others will learn if they devote the time to read it. You must answer "**Whose mind are you going to change about what?**", such as by starting the abstract with "In this paper, we demonstrate that…" (emphasis, ours)

I will add to this that the goal is to make a contribution to the scholarly literature, not to do a replication project. Thus, while replicating the authors' results are necessary to complete your extension the bulk of the paper should be about the nature of your contribution. Importantly this is now *your* paper and not the original authors', so you decide what parts of the problem are interesting. This is the essential **pivot** in writing a replication/extension paper.

Important components to an empirical paper

- Lay out your story.[2]
- Clearly state your estimand.
- Explain your model(s) [3]
  - Mathematically write out your model
  - Describe covariates and model assumptions
- Explicitly show your reader how the paper is organized [4]
  - Give your readers a roadmap of the paper (what's to come), clearly signpost the argument
  - Use headers that make claims, not just provide topics.
  - Use different font styles, bullet points or enumeration when necessary, to distinguish important sections

[1] King, G. (2006). Publication, publication. *PS: Political Science & Politics*, 39(01):119–125

[2] This includes whose mind are you going to change about what, but also establishing simple claims or providing an empirical puzzle.

[3] "For the ith student in class c and school s, we can write

$$(2) \quad y_isc = X_s'\beta + n_{sc}\alpha + \mu_c + \eta_s + \epsilon_{isc}$$

where y_{isc} is pupil i's score, X_s is a vector of school characteristics, sometimes including functions of enrollment, and n_{sc} is the size of class c in school s. The term μ_c is an i.i.d. random class component, and the term η_s is an i.i.d. random school component. The remaining error component ϵ_{isc} is specific to pupils." (Angrist and Lavy, 1999)

[4] "The remainder of this paper describes our approach to interpreting and presenting statistical results. We begin by formalizing the problem of statistical interpretation (Section 2)…" (King et al., 2000)

For this class, we will be loosely following the formatting guidelines of the Proceedings of the National Academy of Sciences (PNAS). This is an important general audience scholarly journal which uses a rather short paper form. You can refer to their website for more information.[5] We have included some of the most salient details below

Article Format

The key advantage of the PNAS format is that it imposes a useful structure on the article format which prioritizes a quick statement of the importance of the problem followed by a more detailed statement of the methods.

- **Title Page:** Include on the title page all contributing authors in alphabetical order.[6] You should also include a significance statement in the title page: a 120-word-maximum statement about the significance of their research paper written at a level understandable to an *undergraduate-educated* scientist *outside* your field of specialty.

- **Main Text:** The PNAS format does not require, and in fact is too short to include a detailed literature review. You should, however, comment on the relevance and significance of your work.

    The main text usually include following sections:

    – Abstract. Provide an abstract of no more than 250 words on page 2 of the manuscript.

    – Introduction: quickly establish the problem and its importance to the field

    – Materials and Methods: in this section you need to describe the data and methods you used.

    – Results: the core results of the paper.

    – Discussion: usually the last page, discuss the implications/limitations of the paper.

    Writing short versions of things is *hard*. If you want good examples look at papers in *Nature* and *Science* which have to establish the stakes of a problem very quickly.[7] There is always a temptation here to write for someone who already knows the literature-essentially confirming that you have also read it by reciting the shibboleth. But you want instead to summarize the broad strokes for a literature who doesn't know it- establishing why we should care and why this is an exciting area.

- The website provides more details on the contents of each section.[8]

- Here is a sample of recent social science publications in PNAS. You can follow this link to find some papers and get a sense of the relative length of each section.[9]

[5] http://www.pnas.org/site/authors/preparation.xhtml

Figure 1: The PNAS format demonstrated with an image from a paper written by Melissa Sands, a graduate student at Harvard, in PNAS.

[6] Haggling over author order is a hard part of academic collaborations. For this everyone should contribute equally and you should go in alphabetical order acknowledging that fact.

[7] Two we like in particular: Salganik et al. (2006) and King et al. (2014).

[8] http://www.pnas.org/site/authors/preparation.xhtml

[9] http://www.pnas.org/content/by/section/Social%20Sciences

Supplemental Information

You may include a document of supplemental information containing additional details which the paper does not have room for. Here you can include details necessary to replicate your results and, if you like, more extended discussions of robustness checks or other methodological details.

You should also include your data file and replication code for all results in the article. These will be helpful for the group reviewing your work.

Stylistic Formatting

- **Word Processing:** You can use wither Word, LATEXor R Markdown to write your paper. If you use LATEXor R Markdown, use BibTeX to compile the bibliography.

- **Paper Length:** Following two ways to make sure that you are within this limit:

  - If you use WORD to write your paper: PNAS put a six-page length cap and total characters: 39,000 on the submitted word document.

    1. Generate a pre-submission copy of your word document following the guidelines here.
    2. Submit your paper to PNAS's page length estimation website. and make sure that you pass their test.

  - If you use LATEXto format you paper: PNAS has their own template. Download their template here or use Overleaf to write your paper online at this website. If you download a local template, put the decompressed files in the template to the folder of your .tex file. A help file is here.

    With the template, you should be able to generate a PNAS-style pdf document. Make sure that your pdf document is within 6 pages (excluding title page).

- **Section Formatting:** Be sure to include the required sections in the main text.

- **Tables and Figures:** Clear and informative tables and figures are vital to communicate your findings. Here are a few guidelines:

  1. Include only the most important tables and figures in the main paper. Any table or figure included should be referenced and discussed. Additional tables and figures can be included in supplementary materials.

2. All tables and figures should be numbered and referenced appropriately in the article.

3. Include detailed captions so that each table and figure can be understood on it's own. The reader should be able to fully understand the tables and figures without needing to refer back to the text. Be sure to state in the caption what you want the reader to take away from the figure. If you find yourself unable to write a clear statement of what you want the reader to take away- you should reconsider whether the table or figure is necessary.

4. Think critically about your tables and figures so that they are interpretable to someone without your expertise. Labels should be understandable (e.g. variable names are clear). Use legends, titles, axis labels judiciously to communicate your results.

Figures and Tables

- All figures and tables need detailed and informative captions so your reader can understand them without referring back to the main text (see figure 2)

- All included figures and tables need to be referenced and interpreted in the document.

- Every figure and table in the appendix needs to be mentioned in the main text so the reader knows what and where they are. [10]

- Graphical defaults are not necessarily best (see figure 3).

   - Legends can take up unnecessary space as well as make figures harder to read. Alternatively, try annotating directly onto the figure.

   - The grey background and gridline defaults of ggplot can be changed.

   - Labels can be rotated to be more legible.

   - Proofread to ensure font size and the plot size as a whole is readable.

   - Don't let anythig be cut-off. This includes titles and labels.

- Make sure positioning of the figure within the document is correct. Don't let things float weirdly to another page or have too much whitespace.

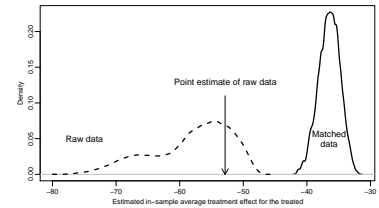- When possible, make your figures decipherable if printed in black and white.

Figure 2: Kernel density plot (a smoothed histogram) of point estimates of the in-sample ATT of the Democratic Senate majority on FDA drug approval time across 262,143 specifications. The solid line presents a density plot of the MLEs of ATT using the matched data set, whereas the dashed line is based on the raw data. The vertical arrow shows the point estimate from Carpenter's Model 1 based on the raw data. The estimate does not match Carpenter's estimate exactly because it is on a different scale and also because of the slightly different set of predictors used, as discussed above. The figure shows that ATT estimates are considerably more sensitive to model specification using the raw data as compared with the preprocessed matched data. (From Ho et al. (2007))
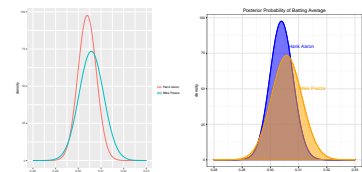

Figure 3: An example using all ggplot defaults (left) and customized aesthetics (right). Robinson (2017) http://varianceexplained.org/r/bayesian_ab_baseball/

[10] Complete summary statistics are available in the Appendix in Table 1.

Writing

- Write for a broad audience and general interest. Limit jargon and write clearly and concisely.

- Your readers may not be familiar with the original replication paper or the literature in general. Give your reader enough background to understand how your analysis fits in.

- Tell the reader what's at stake. Why does the question matter? How do your findings inform what we think about what's at stake? Who's mind are you going to change about what?

- PNAS format still needs some background information (e.g. short, concise, literature review) and proper citations.

- Your paper should first set-up the original paper you replicated, then clearly describe how you build and deviate from this.

- Make sure to weave together your arguments and your supporting analysis on the backend. This is your last chance to highlight your findings and show the reader what's at stake. Leave the reader with a strong conclusion.

What's a poster session?

Poster sessions are a great way to get feedback on an ongoing project.[11] You put up a visual aid of your results (the poster) and then pitch the project repeatedly to the passers-by, typically summarizing the entire work in under 2 minutes. Believe us- nothing cements the key features of a piece of work like being forced to repeatedly explain it.

- **Your Poster is a Visual Aid:** Your poster is not intended to be read, it is a visual aid to help you explain your argument. Thus, it should have way less text than you might imagine. One of Brandon's favorite posters contains very little text (see Figure 4). This also means you are likely to write in short phrases rather than complete sentences.

- **How to Make a Poster:** Some notes on designing a poster with beamer (powerpoint, such as for Ian's poster (5) reviewed in precept). We can think of creating in roughly the following steps:

  1. Begin with a Beamer template that makes a poster of the correct dimensions: A0 (33.1 × 46.8 inches). We recommend overleaf.com as a general source of Beamer poster templates. If you want to have a specific template to use, here is a link to a version of Rebecca's poster (6) with content removed on Overleaf;

[11] This section owes a lot Kosuke Imai's "How to Make a Poster for Social Science Research" handout: `http://imai.princeton.edu/teaching/files/HowToPoster.pdf`.
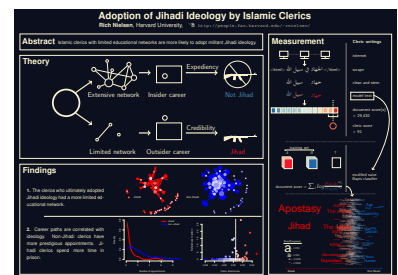

Figure 4: Rich Nielsen's Polmeth 2012 poster on a portion of his dissertation work. Note how little text there is.
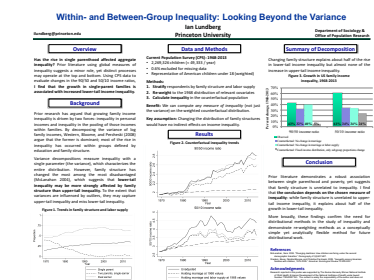

Figure 5: Poster from Ian created using PowerPoint

if you use, make sure to change some of the aesthetics so that each poster looks different!

2. Separate the poster into columns using

   \begin{column} \end{column}

   Most posters have 3-4 columns.

3. Within each column, you can use commands like

   \begin{block} \end{block}

   (for a block without a background) or

   \begin{alertblock} \end{alertblock}

   (for a block with a background/outline) to separate the poster into sections like: "Research question," "Hypotheses," etc.

4. Within each block, you can put different combinations of Figures and text (though not too much text!) as you see fit

- **Getting Feedback:** Use the session to figure out what people are reacting to in your project. Try out different pitches, look for what their questions are or what claims they don't buy, watch whether or not they understand your tables and figures. Finally, be sure to listen to yourself or your partner. What you are saying to communicate the point of a figure or table is what should be included in the paper's caption for that figure or table.

Figure 6: Poster from Rebecca created using Beamer

References

Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236.

King, G. (2006). Publication, publication. *PS: Political Science & Politics*, 39(01):119–125.

King, G., Pan, J., and Roberts, M. E. (2014). Reverse-engineering censorship in china: Randomized experimentation and participant observation. *Science*, 345(6199):1251722.

King, G., Tomz, M., and Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American journal of political science*, pages 347–361.
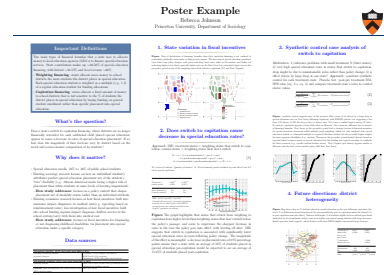
Robinson, D. (2017). *An introduction to empirical bayes: Examples from baseball statistics*.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856.