

Pre-Registration Assignment

Overview

A critical component of conducting replicable research is to develop a clear testable theory *before* analyzing the data. By clearly laying out hypotheses before working with the data, researchers can be confident they are not Hypothesizing After the Results are Known (HARKing). If we wait to hypothesize until after we have seen what are data or models look like, we run the risk of creating post-hoc explanations for results. The data generating process is often noisy, random, and messy. By definition, random samples will frequently produce results that are different than the popular parameter. The hope is that with enough samples and tests, scientists can eventually coalesce around a shared conclusion. However, this process quickly breaks down if theory and expectations are changed to aligned with data, rather than using data to test existing theories.

In response to this, researchers have increasingly proposed using pre-registration as a tool to combat HARKing. Scholars will post not only their expectations, but also their research design and potentially their data generating process before conducting research. Essentially, they are publicly posting their plan of analysis so that once they do begin working with the data others can be confident they are testing their hypotheses, rather than changing or generating them to confirm the results they already found. This approach can help build trust in a researcher's approach and also provide a blueprint for other researchers to use the same approach on new datasets.

For this class, you will be pre-registering your final paper. This means that before you download the data for your final analysis, you must submit your literature review, hypotheses, and research design for review. In addition to being a best practice for replicable research, it also provides us both with an opportunity to check in on the state of your project mid-way through the semester.

Specifics

Check In 1: Research Question

Upload a document that lists what your paper's research question is and who you will be writing the paper with. This needs to be specific (don't say I want to study public opinion, for example. A better topic would be "I want to study how legalization of recreational marijuana has changed attitudes toward marijuana compared to states that have not legalized it"). In addition to your topic/question, include a short writeup (1-2 paragraphs) outlining broadly the motivation for the paper and why you think this question is important).

As part of this also include 1 paragraph describing the dataset you plan on using and include a link to the data. At this stage I'll be evaluating questions for their specificity and feasibility.

Check In 2: Annotated Bibliography

Upload a document that includes at minimum 6 scholarly sources from peer reviewed journals that relate to your topic. You will list a full source citation, then under it write 1-2 paragraphs summarizing the source's findings and how you plan on incorporating it into your paper. This is the start of you developing a theory, so it is vital you think about what sources you are using. After outlining at least 6 sources, include a concluding paragraph that summarizes how these sources collectively will contribute to your paper. Are they pointing in a direction to build a theory? Are you filling a gap you identified in all these studies?

Check in 3: Literature Review and Research Design

As part of this, formalize at least 2 hypotheses for your research question. Some people like to put the hypotheses at the end of a section of the literature review, others like to have a separate section called "expectations" where they have a paragraph summing up their theory with the hypothesis following.

The key to good hypotheses is that they must be specific and related to your research question. A model to follow is "An increase in the [key independent variable] will be associated with an increase/decrease in [key dependent variable]."

You will also be submitting your pre-analysis plan as part of this submission. This will overview the data you will use, including a description of the key variables and controls. I am also looking for a discussion of your modeling approach, including the specific models you will be using. This needs to be as detailed as possible so that you have a clear plan you will follow once you start working with the data.

- Necessary components of your research design:

1. What data will you use to test your hypothesis?
 - a. Discussion of the unit of analysis, time coverage, and geographic coverage.
 - b. Discussion of the source of the data — how representative is it?
 - c. What kind of data is it? (Individual public opinion, aggregate policy, etc.)
2. What is your key dependent variable?
 - a. How is it measured? How does it relate to the concept you discuss in your theory?
3. What is your key independent variable?
 - a. How is it measured? How does it relate to the concept you discuss in your theory?
4. What other factors do you think need to be included in the model?
5. What method are you going to use to test your hypothesis (we will talk about this more in class)?

After you submit your plan of analysis, I will review it and give you feedback. Once you have incorporated any suggested changes, you are then clear to begin cleaning and analyzing the data. However, keep in mind that your hypotheses cannot be changed at this point, and any changes to the research design must provide clear justification for the change in the updated text. For more information on expectations for the pre-analysis plan, see rubric.