



PPHA 34600: Program Evaluation
Spring 2020
Problem Set 2

Due: Thursday, May 14, at 9PM to Canvas

Instructions:

This problem set consists of two files: (1) this document with instructions and questions; and (2) a dataset which you will use to answer the questions below.

You can work in groups of up to three. Please identify your group members. Groups can share code, but each group member must turn in their own problem set, and must have separate written answers to the questions. You should submit both written answers -- which should be parsimonious -- and a file which contains your code and results for the data analysis. You must use R. I recommend that you use RMarkdown or knitr, which will allow you to intersperse your code and written answers. Note that you are primarily being graded on your written answers. Problem sets must be submitted in PDF format. Problem sets must be turned in via Canvas; no late submissions will be considered.

Questions:

A well-meaning NGO, Harvest Aid and Rainfall Risk Insurance Services (HARRIS) is interested in improving agricultural yields for farmers in Bangladesh. They are particularly excited about the idea of providing farmers with rainfall-index insurance. While most insurance suffers from adverse selection problems (people who already have low yields are likely to select in) and moral hazard issues (when you have insurance, you may be less likely to work hard on your fields), rainfall-index insurance avoids these issues by tying insurance payouts to how much it rains, rather than individual crop yields. Their idea is that, if farmers have insurance, they can invest in riskier -- but more profitable -- up-front inputs like fertilizer, with less worry of losing money during a bad rainfall year (I'm not just making this up; this is a [real thing](#)), thereby raising overall profitability.

They are working on putting together proposal documents to fundraise to pilot this type of insurance product, and have discovered that the Indian government has already implemented a program to do exactly this, called the Farmer's Insurance On Non-rainy Annus (FIONA) scheme. This scheme ran from 2014 to 2016. HARRIS has asked you to help them evaluate the effectiveness of the FIONA scheme in order to understand whether this is a good idea for Bangladesh. (Assume for the purposes of this problem set that farmers do not change crops in response to FIONA.)

1. HARRIS are interested in answering the following question: *What was the effect of FIONA on profits for the average farmer?* To make sure everybody is on the same page, explain to them what the ideal experiment would be for answering this question. Describe the dataset that you'd like to have to carry out this ideal experiment, and use math, words, and the potential outcomes framework to explain what you would estimate and how you would do so. Make sure to be clear about the unit of analysis (ie, what is "i" here?).



2. HARRIS like what you're suggesting, but think it's answering the wrong question. They aren't going to be able to get every single farmer to participate. They'd instead like to know: *What was the effect of FIONA on profits among farmers who took up insurance?* Describe in math and words, using the potential outcomes framework, what they'd like to estimate. Explain how this differs from what you described in (1), and describe what component of this estimand you will be fundamentally unable to observe.
3. HARRIS are on board with your explanation. Because FIONA already exists in the real world, they can't run an RCT to study it. However, they do know that not all farmers were offered insurance through FIONA. It turns out that FIONA only impacted certain districts. Non-FIONA districts were not offered any insurance products. Explain what you would recover if you simply compared FIONA farms to non-FIONA farms on average. Describe three concrete examples of why this might be problematic.
4. HARRIS hears your concerns, but still wants an estimate of the impacts of FIONA. Given that you're unable to implement your ideal experiment, and you are worried about simple comparisons of FIONA-aided farmers and those without insurance, you'll need to do something a little more sophisticated. Luckily for you and for HARRIS, India makes data on farmers available to the public, in the form of *ps2_data.csv*. Read the data into R and, as always, make sure everything makes sense. Document and fix any errors. Use the variables contained in the dataset to describe, using math and words, two (related) potential approaches to estimating the effect of FIONA on profits. Make sure to be clear about your unit of analysis, and be explicit about how these designs apply to FIONA (ie, describe things in terms of "profits," not just "outcome"). Hint: HARRIS wants you to describe two selection-on-observables designs.
5. Produce a balance table which displays the differences between FIONA and non-FIONA farmers on observable characteristics. Interpret this table. Does this table make you feel better or worse about your concerns in (3)?
6. HARRIS are interested in your approach in (4), but would like to know a bit more about how much they should believe your proposal. Describe the assumptions required for these designs to be valid in math and in words. To the extent possible, assess the validity of these assumptions using the provided data. Discuss whether you think you will be able to obtain a credible estimate of the answer to the questions described in (1) and (2) based on the data, and use concrete examples to explain why or why not.
7. Use a regression-based approach to estimate the effect of FIONA on farmer profits. Describe which variables you chose to include in your regression, and explain why you chose these. Did you leave any variables out? If yes, explain why. Interpret your results. What are the strengths and weaknesses of this approach? How do your results differ from what you find if you instead use the naive estimator?
8. Use an exact matching approach to estimate the effect of FIONA on farmer profits. What variables should you include in the matching procedure? Begin by estimating the answer to the question in (1). Then, estimate the answer to the question in (2). Are these meaningfully



different? Would you have expected these results to be the same? Why or why not? What are the strengths and weaknesses of this approach? How do your results differ from what you find if you instead use the naive estimator? From what you found in (8)? Did you run into the Curse of Dimensionality with this analysis? If yes, describe how it affected your approach. If not, describe how the Curse could have generated problems in this setting.

9. Based on your results in (8), explain to HARRIS whether or not they should implement a FIONA-like program in Bangladesh. Be sure to tell them the reasoning behind your recommendation.