AC 4

For this analysis challenge, you’re going to pre-register an analysis plan. You’ll then implement this analysis plan next time using data that I will make available to you later.

**Why a pre-registered analysis plan (PAP) is beneficial**: We’ve been talking about the problem of p-hacking and p-screening. PAPs provide a nice mechanism around these problems. By pre-registering with a journal or some other public hub, scholars have a harder time “torturing” the data until it “confesses.” Also, journals have a vested interest in publishing pre-registered studies because they want to appear like responsible scientific outlets. The first means less p-hacking. The second means less p-screening.

**What you’ll do**: You’re going to put together a short PAP for this analysis challenge. In it, you’ll tell me the names of the variables you’ll control for when estimating the relationship between an outcome and explanatory variable of interest. You’ll then write example code for how you will do your analysis. *I’ve provided a template on the next page*.

## Pre-registered Analysis Plan for Montana Wildfires

### Background

The state of Montana has a problem. Forest fires are a regular seasonal occurrence. To limit the spread and damage caused, the Forest Service offers to inspect the properties of individuals to ensure they are “fire proofed.” If more people take advantage of this service, forest fires wouldn’t be eliminated, but they would spread less intensely and pose a lower risk of damage and loss of life. Unfortunately, this service is voluntary and estimates suggest that less than 1% of the population take advantage of it. So, in conjunction with the Forest Service, we’ve conducted a randomized controlled trial (RCT) in the form of behaviorally informed letters intended to highlight the risks individuals pose to their neighbors by not ensuring that their property is fire proofed. We want to know, did the letters increase applications for property inspections?

### Design of the Experiment and Expected Data

Households in two counties were randomly assigned to either treatment or control. Theoretically, each household, regardless of characteristics, had an equal chance of getting a letter.

For each household, we have the following data:

* county: a county identifier
* hhincome: household income
* acres: the size of the property in acres
* urban: an indicator for whether the property is urban or rural

We additionally will have information about which households received a letter and which submitted an application for an inspection. These variables will be letter and app respectively.

### Analysis

*This is where you come in. Given what you know about the data, how do you want to analyze it? Your answer will need to address the following:*

1. Will you use a regression model? If you will, how will it be specified (you’ll need to write out the formula you’ll use in R, e.g., y ~ x).
2. If you do use regression, will you include control variables? Which ones? And why?
3. What will be your criteria for rejecting the null hypothesis?