PAP Template for Experiment Projects - Project 3

Project 3

April 2021

Contents

Instructions:

This is a template of an experimental plan document, or what is called a Pre-analysis Plan (PAP). In this course, your team will work together on producing a PAP to plan your experiment. It will be one of your final deliverables.[1]

The PAP is a planning tool to help your team 1) structure the research from design to analysis, 2) think through all the necessary steps before implementation, 3) lay out specific hypotheses you want to test, and 4) plan for analysis before the data comes in. This plan lowers the risk of ex-post cherry-picking of causal relations through precisely defining the hypotheses, outcomes, and tests prior to data collection and/or analysis.

This template is produced to help you frame the planning of your experiment. We have filled in the template based on the Pre-analysis Plan for "Optimal Policies to Battle the Coronavirus Infodemic Among Social Media Users in Sub-Saharan Africa" (Offer-Westort, Rosenzweig, and Athey, 2020). Note that we have adapted and simplified some sections.

You should be able to complete a PAP for your experiment using resources such as the Project Packet, discussions with partners and experts, and the experimental design tutorials.

[1] This tutorial was originally developed for the Spring 2021 course, ALP301 Data-Driven Impact.

Motivation

Colorado's unique election system is often referred to as the "gold standard for the nation" and has some of the highest voter turnout rates in the country. Since 2013, all registered voters in the state receive their ballot in the mail and have a variety of ways to return it (via mail, drop box, or in person). In 2017, Colorado became the first state to use a security feature, Risk-Limiting Audit (RLA), to check and verify electoral results.

While many academics, NGOs, and election officials view RLAs as an ideal way to confirm the integrity of the election and final vote count, the RLA process is often challenging to convey to the general public. Communicating the importance and security of RLAs is critical, particularly as other states are looking to implement RLAs in future elections.

...

The partners for this project include the Colorado State government office and a cross-disciplinary group at Stanford, including researchers and policy experts at the Stanford Internet Observatory, that studies abuse of information technologies and develops a curriculum on trust and safety.

The goal of this project is to learn about the effective means to communicate the RLA technology to the public, and improve public trust in the election process. To achieve this goal, the project team will design an experiment with experts on RLAs. This project would directly inform election policy across the country, enabling election officials to rebuild voter confidence in the integrity of elections. The group will test the effectivness of several interventions to increase confidence in using RLAs to verify election results, including different messages to convey why RLAs ensure integrity in the election from different credible messengers (e.g. teacher, judges). Our outcome of interest will be trust in the individual's State election results from learning about RLAs. Individuals do not need to understand or know about RLAs already. ...

We believe that the insights gleaned from this experiment will both contribute to generalized knowledge about how to increase confidence in elections, and RLAs in Colorado, and provide a path for other states besides Colorado to implement, and market, RLAs to their constituents.

Research Questions

- What are effective strategies to educate voters about how RLAs work?
- Specifically, what characteristics of RLAs are most effective in increasing voters' confidence in the integrity of the election results?
- Do different strategies work better for different subgroups?

Experimental Setup

Describe how the research question can be answered with an experiment, comment on the methodology used. For example, argue why a randomized experiment is useful to answer the research question.

This section should be a "how-to" guide for executing your experiment. In theory, someone should be able to execute your experiment after reading this section.

The research question can be answered using an experiment to test the efficacy of different messages on increasing confidence in election systems by providing information about RLAs.

We will use four different messages with a pre- and post-test question to gauge the effect of the message. We will have one control group and three different messages to use on different treatment groups. The control group will receive a message about election integrity, with no mention of RLAs. The three treatment groups will each receive a different message about RLAs of similar length and reading level.

Participants will first answer survey questions about partisanship and views on the integrity of the most recent elections for our pre-test data. Next, participants will answer a few questions where we can collect demographic information including age, gender, education. These will be useful covariates to analyze our data and measure the effect of messages for subgroups.

Next, participants will receive the intervention (one of the four possible messages). Lastly, participants will receive the same set of questions as the pre-test. This design is a between-subject design, where interventions are implemented at the respondent level.

Sample recruitment *ADD POWER CALCULATION

Participants will be recruited through Lucid during the first half of May 2021, which allows us to target a representative sample of US adults (18 years old and up). Our target sample size is X respondents for each treatment group, totaling X respondents across all 50 states.

Apart from Colorado, several other states have recently or are planning to use RLAs in the future and the goal of using a nationally representative sample is to gain generalizable insights about the potential benefits of well-communicated information about RLAs on voter confidence in the election.

Treatment

Please explain your experimental interventions in detail here. Please include a flow chart/diagram that helps explain to readers your experimental setup and procedure.

You should provide details for each treatment arm here. Remember, visuals help! You can provide images or screenshots of what the actual experiment will look like to participants.

You should also provide the rationale for choosing the treatment arms.

Source: You can draw inspirations from your team discussions, discussion with the partner, as well as previous treatment interventions used in the literature.

Example:

Drawing on the literature on experimental interventions to combat misinformation, we include several treatments designed to reduce the spread of misinformation online, which are targeted both at the respondent level and the headline level. This list of treatments also draws on real-world interventions that companies and platforms have instituted to combat misinformation. The treatments are presented in Table 1.

Respondent-level treatments and headline-level treatments are implemented as separate factors, each of which has an empty baseline level that is the control. So respondents may be assigned the pure control condition, one of the respondent-level treatments but no headline level treatment, one of the headline-level treatments but no respondent-level treatment, or one of the respondent-level treatments and one of the headline-level treatments.

** add 3 treatments and control group control respondent - treated with paragraph of similar length to treament effects on importance of elections

UDPATE PROMPTS

Table 1

Shorthand Name	Treatment Level	Treatment
1. Bipartisan Auditors	Respondent	Prompt:
2. Risk Limit as a Percentage	Respondent	Prompt:
3. Emotional - Politicians are Bad Losers	Respondent	Prompt:
4. Control	Respondent	Prompt:

Table 1. Description of interventions included in the experiment

Outcomes

How will you measure and analyze both primary and secondary outcomes of interest? How do they help you address the research questions?

Source: You can discuss with the partners to understand whether the outcomes are meaningful for them.

Primary:

(1) Is communicating about RLAs an effective way to increase trust in election system for the US adult population?

Secondary:

(2) Which messages about RLAs are the most salient?

We are primarily interested in increasing awareness about RLAs as tools to verify election results. Simultaneously, we want to determine which message is the most effective to convey how RLAs work and why they are trustworthy. Specifically, we are interested in the following outcomes:

Primary:

(1) Increasing voters confidence for US adults that ballots are correctly counted in the state and national elections.

Secondary:

(2) Determining the most effective messages to build trust around RLAs - bipartisan, logical, or emotional appeals.

(3) Gauging whether certain messages more effective at increasing trust in elections for certain subgroups.

Primary Outcomes _ Describe the primary outcome. Argue why this is a relevant metric for answering the research questions. If the design of your experiment is more involved or has multiple outcomes of interest, feel free to add additional sections as necessary._

Example:

You should lay out the questions you use to measure your outcomes.

UPDATE - NATE TO SEND

We will measure trust in elections from RLAs through two questions:

measure effectiveness of treament compared to the control group

- Would you like to share this post on your timeline?
- Would you like to send this post to a friend on Messenger?

We use a pre-test / post-test design. Prior to treatment, we show respondents four media posts from their country (two true and two false in random order) randomly sourced from our stimuli set. For each stimulus, we ask the above self-reported sharing intention questions. Respondents are then asked a series of questions about their media consumption, and are then randomly assigned treatment according to the experimental design. If assigned to one of the respondent level treatments, they are administered the relevant treatment.

They are then shown four additional stimuli (two true and two false), selected from the remaining stimuli that they were not shown pre-treatment. If the respondent is assigned a headline-level treatment, this treatment is applied only to the misinformation stimuli, as flags and fact-checking labels are not generally applied to true information from verified sources. For each of the stimuli we again ask the same self-reported sharing intention questions.

Because of random assignment, we expect to see no systematic differences in pre-test interest in sharing either true or untrue stimuli across treatment conditions, conditional on covariates.

Secondary Outcomes *Secondary outcomes are things that you also wish to investigate in addition to the primary outcomes. Sometimes these outcomes can shed light on the mechanisms through which interventions work. Sometimes they help ensure that your intervention does not lead to unintended effects.*

In addition, we measure secondary outcomes by comparing the effectiveness of each treatment against each other to see which message is more effective. We will run a multiple hypothesis test between each treatment group to determine this marginal treatment effect. We will also examine the effect for particular sub-groups.

Covariates

Which covariates will you include in the experiment and why? This selection can for example depend on expected treatment heterogeneity. Some common covariates include age, partisanship, Internet usage, location, etc. Explain why the covariates are important in your context.

The existing literature on election integrity focuses on voter confidence in specific elections (Atkeson et al., 2015), the role of partisanship (Sances et. al, 2015), and factors that can drastically affect voter experiences like poll workers (Hall et al., 2009). Our study contributes to the existing literature by exploring what messages are most effective at explaining RLAs and building trust and confidence in election systems. What messages are most salient with US adults to understand and trust that their vote was accurately counted?

In addition to the demographic covariates such as age, race, gender, and education, we also include specific questions regarding party affiliation, political ideology, and trust in state and federal government. This includes a control for people with strong or unwavering views on election integrity for us to measure the treatment effect.

These variables capture what might be sources of heterogeneity in responses to RLAs and election integrity: age, education, and political ideology, and trust in government.

Hypotheses

Discuss how you would map the research questions into testable hypotheses. For example, if your goal is to find some prototypes for further exploration, which kind of test would you want to conduct? If your goal is to figure out if an intervention works better for one specific group, how would you formulate the hypothesis?

How do you plan to address the problem of multiple hypotheses testing? If you are making a final recommendation rather than selecting some for future exploration, you might want to choose the more conservative correction methods.

Primary:

The primary hypotheses that we want to measure would be whether any of the treatments to communicate about RLAs are effective in increasing voter trust in election systems.

Secondary (Hypotheses to inform industry practice)

We care about the outcome (reducing the spread of COVID-19 misinformation) and understanding which types of people are nudged toward this outcome by particular treatments. Therefore, we plan to examine how a few select treatments interact with particular covariates of interest.

We select the below treatments because these are currently, or were previously, used by social media companies including Facebook and Twitter. The below covariates were selected as those that social media companies directly collect or have access to, and therefore could more easily use for targeting interventions. For our covariates of interest, we will divide these into two groups for any binary variables (i.e. indicator for male) and split on the median value for continuous variables to test two subgroups (i.e. age $\geq$ median and age $<$ median)

Treatments:

- Bipartisan Auditors (respondent)

- Risk Limit as a Percentage (respondent)
- Emotional - Politicians are Bad Losers (respondent)
- Control (respondent)

Covariates: *ADD TO*

- Age
- Gender
- Race
- Education

UPDATE We hypothesize that the three headline-level treatments listed above will perform better among more educated users, older people, and among women, compared to the less educated, younger and male respondents. We expect that the two respondent-level treatments will reduce sharing of misinformation more among less-educated respondents than those with more education.

We hypothesize that learning about RLAs will have at least a small positive effect on trust in election systems.

Analysis

Label and index your data: outcomes, treatments, covariates. Formally describe how you test the hypotheses described earlier.

Source: Analysis should also follow your hypotheses. For each of your hypotheses, what analysis will allow you to test the hypotheses?

For the primary and secondary outcomes discussed above, we will examine and compare the average treatment effect as the movement on the scale between the pre- and post-treatment.

To test our hypotheses, we will conduct a one-sided T-test to examine the average treatment effect as we believe the treatment will have a positive effect in one direction on the outcome variable. We will compare the average effect for each treatment to the control average, and then compare each treatment average to each other to determine which message is most effective.

We will also analyze the effect of the treatments listed above by sub-groups.

UPDATE We expect that...

Given that testing these treatment-covariate combinations will result in a large number of unique tests, we will adjust for multiple hypothesis testing using the Bonferroni correction method to eliminate Type I errors (false positive discoveries). This is a result of our objective to test multiple treatments against the control and each other.

Power Calculations

A crucial function of the PAP is to help you calculate the sample size you would need to detect a hypothesized treatment effect. Depending on your experimental design, please use the relevant tutorial for power calculation code.

Here is a resource containing principles for running power calculations.

Please indicate your choice of the minimum detectable effect and rationale. There is no universal rule of thumb for determining a "good" minimum detectable effect. For researchers, this might be informed by the existing literature: what have previous studies of comparable interventions found? What would be the smallest effect size that would be interesting to be able to reject? For partners, this might be the smallest effect that would still make it worthwhile to run this program (from their own perspective, or from a funder's or policymaker's

perspective), as opposed to dedicating resources elsewhere. This may mean the smallest effect that meets their cost-benefit assessment, the smallest effect that is clinically relevant, or some other benchmark.

Please include the code, result, and figures from your power calculations.

Source: If your experiment involves multiple treatment arms, and you are comparing each of them against the control, you can adapt the power calculation code in the Multiple hypothesis testing tutorial. If you are using a factorial design experiment, you need to adapt the code in Factorial Design Tutorial, following the hypotheses that you outlined in the above section and the correction method that suits your objective.

Load required packages

```
# you can install the packages you need for power calculation and analysis
if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse)
pacman::p_load(randomizr)
pacman::p_load(estimatr)
pacman::p_load(kableExtra)
pacman::p_load(ggthemes)
pacman::p_load(reshape2)
pacman::p_load(bindata)
```

UPDATE POWER CALCULATIONS

```
# Add your code for power calculation here

# If you are using a factorial design,
# remember to use both the lm_interacted() function in 'lm_model' code chunk and the 'power_simulated'

# Please adjust the number of hypotheses, hypotheses type, effect size, and treatment-control split etc

power_calculator <- function(mu_1, # treatment mean
                             mu_0, # control mean
                             sigma, # standard deviation
                             alpha=0.05, # significance level
                             N, # experiment size
                             hypothesis = 'two.tailed' # type of hypothesis
){
  # The tails of our statistic's distribution
  lowertail <- (abs(mu_1 - mu_0)*sqrt(N))/(2*sigma) # note that this is Z inside the power formula
  uppertail <- -1*lowertail # upper tail matters if we are interested in testing if u1 < u0
  if(hypothesis == 'two.tailed'){
    # HA: \mu_1 \neq \mu_0
    # qnorm(1-a) gives z_a; pnorm(x, lower.tail=TRUE) gives prob(X<=x); pnorm(z, lower.tail=FALSE) give
    pwr <- pnorm(lowertail - qnorm(1-alpha/2), lower.tail=TRUE) + 1- pnorm(uppertail- qnorm(1-alpha/2),
  } else if(hypothesis == 'greater'){
    # HA: \mu_1 > \mu_0
    pwr <- pnorm(lowertail - qnorm(1-alpha), lower.tail=TRUE)
  } else if(hypothesis == 'lower'){
    # HA: \mu_1 < \mu_0
    pwr <- 1 - pnorm(uppertail- qnorm(1-alpha), lower.tail=FALSE)
  }
  return(pwr)
}
```

```r
mu_1 = 0 # treatment mean
mu_0 = 0 # control mean
sigma = 0 # standard deviation
hypothesis = 'greater'  # type of hypothesis

# apply power calculator by plugging in the above parameters
pwr <- power_calculator(mu_1 = mu_1,
                        mu_0 = mu_0,
                        sigma = sigma,
                        alpha=0.05,
                        N = 100,
                        hypothesis = hypothesis)
```

Analysis Scripts

Importantly, please include analysis scripts that can be directly applied to your experimental data. The data engineers should work to extract and organize experimental data so that the data scripts can be applied. Remember to always comment the code and add explanation after the code output.

Source: The experimental design tutorials contain code for analyzing a dataset, including coding up the outcome variables, treatment variables, covariates, checking balance in treatment assignment and analyzing using regression models.

Add t-test and code from tutorials - Anderson to add *UPDATE COVARIATES*

```r
tibble::glimpse(df) # overview of variables in the data

# select covariates
covariate_names <- c("age", "race", "educ", "state", "parent",
                     "female", "trust_fed", "trust_state", "ballot_2016", "ballot_2020")

# treatment
treatment_name <- "treat_real"
# outcome of interest
outcome_variable1 <- "rla_state"
outcome_variable2 <- "rla_fed"
outcome_variable2 <- "rla_info"

# create new dataset containing the covariates, treatment and outcome
election_df <- df %>%
  # select all the variables of interest
  select(all_of(c(covariate_names, treatment_name, outcome_variable1, outcome_variable2, outcome_variabl

# Filtered dataframe with observations that have a phone number
election_full_df <- election_df %>%
  # exclude missing `treat_real` observations
  filter(!is.na(treat_real))

# Remove the full data from memory
rm(df)
```

References