

## **P2: Critique a Research Design**

Dear Colleague,

Thank you for reaching out and trust me to have a look at this project. I went over each section extensively and made some comments. I think overall, you did a very good job in outlining the strategy for the project. There are a few points in the research that I think could benefit from a second visit. Three areas that I would pay the most attention to are Experimental Design, Experimental Treatment, and Statistical Method. I think the level of details in these three areas are too broad and could use a little more clarification. Nevertheless, it is an interesting research project, and I can see the potential for interesting findings. I'll look forward to that.

### **1) Overview**

The goal of this analysis is to determine if the TurnOutForWhat! (Rock the Vote) letter-writing campaign affected participation in the upcoming General Election in California (Districts 14 and 15). Our consulting firm will design the project, and will provide you with the list of voters to contact (based on the carefully constructed treatment / control condition). TurnOutForWhat! Will distribute the letters. The rest of the document outlines the proposed design.

#### **- Review Statements:**

- The goal of the project “to determine if the letter writing campaign affected participation in the upcoming General Election in California (District 14 & 15)” is well formulated and stated immediately to grab the reader attention.
- Something that I feel unclear here is the format of the letter. I think a high-level clarification of the format here will give your client a better idea of what they will be preparing.
- Otherwise, I think you did a great job in carefully break down the owners of each task. This is well put together in a way that reader can quickly refers to it throughout the research.

### **2) Research Question**

Do letters sent via the United States Postal Service increase one's propensity to vote in the upcoming General Election in CA Districts 14 and 15?

#### **- Review Statements:**

- The research question is clear in term of articulating what is it that the research is aiming to achieve and execute it.

- I think the “propensity” terminology is a little unclear and could use a definition or a trailing justification for the usage of the term. For example, you can sufficiently state propensity as a way of measuring the increase of voter turnout.
- The letter format could be made a little clearer here. We should include the fact that we are writing letters that appeal to potential voters in the mentioned regions in California.
- I proposed an alternative question that could be use here: Do letters that are written to appeal to voters sent via the United States Postal Service increase one’s propensity, measured by a positive change between the number of voters between this coming election and in previous years, to vote in the upcoming General Election in CA Districts 14 and 15?

### 3) Data

We will use voter registration data from the California Secretary of State on voters in Districts 14 and 15 to target likely voters. Our conceptualization of likely voters is those who plan to vote for the first time in the upcoming election, and voters who have voted in 2016 or any election after. The voter file contains voters’ name, address, political party affiliation, date of registration, and past turnout information (what elections they turned out for). Remember we do not know **who** someone voted for.

#### - Review Statements:

- The way of acquiring desired data from the California Secretary of State on voters in District 14 and 15 to target likely voter is clearly stated as a topic sentence. I think is well stated in a way that your client can understand and quickly reference this section as needed.
- You ended this section with what is and is not contained in the voter file, which serve as a good summary to the dataset of interest.
- Here, you conceptualized likely voters as those who plan to vote for the first time in the upcoming election, and voters who have voted in 2016 or earlier election. This statement gives your client a good sense of what is it from the data that you will be using.
- I agree with conceptualizing likely voters as those who plan to vote for the first time because it is well in line with the research question.
- I can see that conceptualizing likely voter as those who have voted in 2016 or earlier election because those are people who have previously voted and would probably vote again. But I think you should keep in mind how exactly would your letter affect these group since this group will most likely vote regardless of having receive the letter.

#### 4) Experimental Design

The experimental method is ideal to assess the causal impact of receiving a letter on one's decision to turnout. Critical to the experimental design is random assignment to treatment. Group assignment to the treatment and control will be determined in the following way: voters whose last name begins with two randomly generated letters will be assigned to the control group, while voters whose last name does not begin with the two letters will be assigned to the treatment group. (e.g. all last names that start with "P" or "R" would be in the control). Those in the treatment group will be sent a letter that encourages them to vote. Those in the control group will not receive a letter.

##### - Review Statements:

- Your suggestion of using a quantitative approach is a good idea since you are assessing the causal impact of receiving letter on one's decision to turnout. This should allow you to perform hypothesis testing to report the level of significance in your findings.
- I think random assignment is a good way to create sample in this experience since we don't want to skew our results.
- I disagree with last name selection method you mentioned here. I think that by just using two randomly generated letter, you could potentially create an uneven stratification within your sample group. Essentially, you might be leaving out some demographic of people who last name does not start with either of the two randomly chosen letter.

#### 5) Sample Size

The sample will include an appropriate number of voters that reflects the budget of your organization. We will also ensure that we have a large enough sample size to have sufficient statistical power to detect an effect given this design.

##### - Review Statements:

- It is great that you include the fact that the gathering of data depends on the budget your client has since this give us a sense of how much data we will be able to get.
- Although, I think you are leaving out an approximation for the amount of data that you will be collected. I think providing this information could give your client a false sense that you are either collecting too few or too many data point. My suggestion is to include a simple percentage of each stratified group you are expecting to study such as men, women, age, demographics, etc.
- You ended this section with ensuring that you will have a large enough sample side to have sufficient statistical power to detect an effect in the research is a good way of clarifying what you will be doing with your dataset.

## 6) Experimental Treatment

As you requested, there are two goals of this project: 1) to increase turnout in the upcoming General Election, and 2) to increase political engagement among the volunteers in your organization. To achieve both goals, volunteers will be provided with a letter template (to follow) and encouraged to write personalized letters in a way that might appeal to specific demographic characteristics of the voter (e.g., race, age, education).

We do not think it is necessary to include a return address on the letters as this may result in some returned mail (undeliverable). If your organization processes these returned messages, the cost of the intervention may increase as a result.

One thing to be mindful of is that the exact timing that the voters receive the letters cannot be precisely determined. We will consult with the United States Postal Service to estimate the best time to send the letter. We want the letter to be sent far in advance of the election so that the letter can have the greatest possible chance to affect the voter.

### - Review Statements:

- After reading this section, I think the two goals of increasing voter turnout in the upcoming general election and increase in political engagement among volunteers is clearly established.
- Your method of providing the volunteers with a letter template and encouraged them to write a personalized letter in a way to appeal to specific demographic does indeed meet the two goals of your research.
- By having the volunteers write the letter, they are more engaged in the campaign process. And by having the written letter be in a way that appeal to specific demographic, you are indeed trying to establish a relationship between the letter and the voter turnout of that demographic.
- I think approaching the problem this way would better equip you to answer the research question at hand. However, I think there is a slight risk in not standardizing the letter since some volunteer might be better at writing than others. To remedy this, I suggest you should have a group of 1-3 people take part in writing batches of letters.
- While I think that mailing bulk of letter through the us postal service without a returning address would minimize our cost of intervention, it might pose a problem for the postal service when they have to deal with people returning the letter to them. This could risk our company future interaction with the postal service. For that reason, I don't think we want to perceive the role of the postal service as a disposable garbage can for our research.
- Furthermore, I think that there is a potential risk of the recipient perceive the letter as suspicious if you do not include a return address.
- I don't think you include the cost of mailing in bulk here. But perhaps a rough estimate could help your client keep track of the budget for this project.

- Finally, I think it is great that you mentioned that you cannot know the exact timing of when the voters will receive the letter but will consult with the postal service to get an estimate of the best time to send the letter. This information is crucial because we don't want to wait a long time between when the recipient receive the letter and when they cast their vote in the election since your audience might forget about the letter.
- Overall, I think you had put a lot of thoughts in the experimental treatment. You sufficiently convey the message of what you are trying to do, how the letter will be written, and how you want to send these letters in order to generate the necessary data.

## 7) Statistical Methods

We will use conventional political science statistical methods to compare participation rates of voters who received letters (treatment condition) and voters who did not receive letters (control condition).

In order to more precisely assess the effect of the treatment, we will account for demographic factors that have been shown in the political science literature to influence turnout—age, previous voting history, and race.

### - Review Statements:

- From what I gathered in earlier section, I think that using the conventional methods to compare participation rates between voters who received the letter and those who did not receive the letter is a great direct approach. Doing so would give you a direct way of measuring the impact of the letter on voter turnout.
- Some questions I have is what kind of hypothesis testing will you be performing here? What is the Null Hypothesis and the Alternate Hypothesis that you are interested in testing?
- I think you should also provide a high-level summary of the level of significant of your test as well as the rejection region for the Null Hypothesis. Doing so would guide your client around the technical complexity of the project.
- You mentioned that you will account for the demographic factors that have been shown in political science literature. I think it would be ideal to cite your sources here so that the reader can refer to that piece of research at their convenience.
- Otherwise, you have a good outline of what statistical methods you are exploring here that could benefit for more clarity.

## 8) Deliverables Post-Election

After the election, we will acquire the post-election voter file to obtain turnout information in Districts 14 and 15. We will conduct the analysis as described above and will provide a report that outlines our findings.

### - Review Statements:

- I think you did a good job by mentioning that the analysis will be conducted on post-election data as the topic sentence here. This gives your client a good sense of where the statistics will come from.
- In term of deliverable, what type of report will you be providing? Is it a slide deck or a scientific paper? Since you mentioned in your email that your client has little technical expertise, I would make the report more concise to get your research message across.
- How long after the election do you intended to collect the data? Perhaps providing your client with an expected time of deliver would be ideal here so that they don't have to wait too long for the report.
- Overall, I think you did a really good job in summarizing the deliverables.

Sincerely,

Ken Trinh

(REAL) Data Scientist

201 Consulting