# Title of Your Report
## Your subtitle

Junming Zhang, Hairong Sun, Xiaoxi Bai, Yangyang Liu

Wednesday, October 28, 2020

**Abstract**

First sentence. Second sentence. Third Sentence. Fourth Sentence.

**key words:** USA 2020 election

## Title of your Report

**Name(s) of Author(s)**

**Date**

## Model

Here we are interested in predicting the popular vote outcome of the 2020 American federal election (include citation). To do this we are employing a post-stratification technique. In the following sub-sections I will describe the model specifics and the post-stratification calculation.

### Model Specifics

I will (incorrectly) be using a linear regression model to model the proportion of voters who will vote for Donald Trump. This is a naive model. I will only be using age, which is recorded as a numeric variable, to model the probability of voting for Donald Trump. The simple linear regression model I am using is:

$$y = \beta_0 + \beta_1 x_{age} + \epsilon$$

Where $y$ represents the proportion of voters who will vote for Donald Trump. Similarly, $\beta_0$ represents the intercept of the model, and is the probability of voting for Donald Trump at age 0. Additionally, $\beta_1$ represents the slope of the model. So, for everyone one unit increase in age, we expect a $\beta_1$ increase in the probability of voting for Donald Trump.

### Post-Stratification

In order to estimate the proportion of voters who will vote for Donald Trump I need to perform a post-stratification analysis. Here I create cells based off different ages. Using the model described in the previous sub-section I will estimate the proportion of voters in each age bin. I will then weight each proportion

estimate (within each bin) by the respective population size of that bin and sum those values and divide that by the entire population size.

# Results

Here you will include all results. This includes descriptive statistics, graphs, figures, tables, and model results. Please ensure that everything is well formatted and in a report style. You must also provide an explanation of the results in this section.

Please ensure that everything is well labelled. So if you have multiple histograms and plots, calling them Figure 1, 2, 3, etc. and referencing them as Figure 1, Figure 2, etc. in your report will be expected. The reader should not get lost in a sea of information. Make sure to have the results be clean, well formatted and digestible.

# Discussion

Here you will summarize the previous sections and discuss conclusions drawn from the results. Make sure to elaborate and connect your analysis to the goal of the study.

## Weaknesses

Here we discuss weaknesses of the study, data, analysis, etc. You can also discuss areas for improvement.

## Next Steps

Here you discuss subsequent work to be done after this report. This can include next steps in terms of statistical analysis (perhaps there is a more efficient algorithm available, or perhaps there is a caveat in the data that would allow for some new technique). Future steps should also be specified in terms of the study setting (eg. including a follow-up survey on something, or a subsequent study that would complement the conclusions of your report).

# References

1. Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0

2. PUMS USA, University of Minnesota, www.ipums.org.

3. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686