

Visualizations of Outreach Effect in Kenyan Elections

Nosa Lawani

4/7/2021

#Written 1

$$y_i = \beta_0 + \beta_1 local + SMS_i + \beta_2 mean_age_i + \epsilon_i$$

Written 2

```
## stan_glm
## family:      gaussian [identity]
## formula:      perc_reg ~ treatment + mean_age
## observations: 100
## predictors:   3
## -----
##              Median  MAD_SD
## (Intercept)    -0.1231  0.0617
## treatmentlocal + SMS  0.0264  0.0076
## mean_age         0.0031  0.0015
##
## Auxiliary parameter(s):
##           Median MAD_SD
## sigma 0.0375 0.0027
##
## -----
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

The median value for the Intercept is the best estimate for β_0 , which is the `perc_reg`, i.e. the percent of registered voters in a polling district, if no one was treated and the `mean_age` was 0. The median value for `local + SMS` is our best estimate for β_1 , which represents the change in percent of registered voters treated counties. The median value for `mean_age` is the best estimate for β_2 , which represents the change in the percent of registered voters for every increase of one year in the polling district's `mean_age`. For all of these estimates, we can discern our confidence in them using the MAD SD; we are 95% confident that the true value lies within two MAD SDs of the median.

Written 3

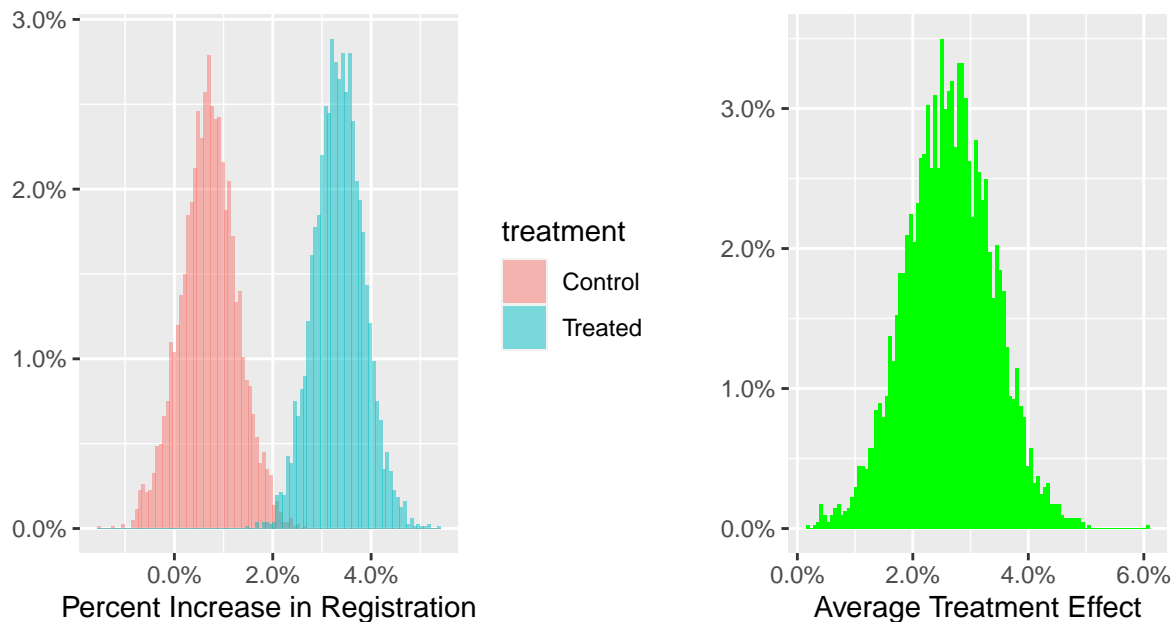
```
##           elpd_diff se_diff
## fit_2    0.0         0.0
## fit_1   -3.6         5.4
```

From the results, our second model, `fit_2`, has a slightly lower elpd and is therefore more accurate. However, the median value for the difference between `fit_1` and `fit_2` is less than 4, a value so small it becomes hard to

distinguish it from noise and so to determine which of the two values is truly better. This median estimate is really only our best guess of the difference between the two. We, however, can be 95% confident that the actual value of the difference is within two standard errors of the mean. With that taken into account, `fit_1` could either be more appreciably worse of an estimate than `fit_2` or appreciably greater than `fit_2`. This only gives more reason to not consider one of these models better than the other.

Posterior Distribution of the Expected Value of the Change in Proportion of Registered Voters

Survey of a 2013 Kenyan Voter Registration Experiment
"local+ SMS" treatment shown



Source: Kenyan Voter Registration Experiment

Written 4

The first plot above shows the posterior for the expected value of the change in the number of voters in a district whose mean age is 42 under both treatment and control. While this same value was negative when age was 0, now that the age has been given a much more reasonable value, draws from the same fitted object show much more reasonable answers. The second graph is a posterior for the estimated average treatment effect. It was obtained by subtracting each of the 4000 draws which comprise the first plot's control posterior from one of the 4000 draws which comprise the first plot's treatment posterior. In doing so, we have calculated an estimate for the Average Treatment Effect, which we cannot know exactly due to the Fundamental Problem of Causal Inference.

```
## Warning: 'as.tibble()' was deprecated in tibble 2.0.0.
## Please use 'as_tibble()' instead.
## The signature and semantics have changed, see '?as_tibble'.
```

Predicted New Voter Registration by Station Average Age

Treated Districts Polled before the Upcoming Election

Tails show 95% Confidence Interval



Source: Kenya Voter Experiment

Written 5

For each age on the y axis, a posterior distribution of newly registered voters for a treated districts of that average age is plotted. As average age increases, the amount of individuals who registered increase. The same is true in the U.S. that old people are more likely to vote. For districts with an average age of 60, the average proportion of newly registered voters is approximately 9%. We would be surprised if any district of an average age of 60 reported an increase greater than 19% or fewer than 0%, since these are outside of our 95% confidence interval.