Absentee ballot rejection rates in North Carolina

Merlin Heidemanns

10/5/2020

```
data {
  int N;
  int A;
  int G;
  int g[N];
  int a[N];
  int rejected[N];
  int submitted[N];
  real<lower = 0> prior_mu;
  real<lower = 0> prior_sigma;
}
parameters {
  matrix[A, G] raw_theta;
  row_vector[G] mu;
  row_vector<lower = 0>[G] sigma;
}
transformed parameters {
  matrix[A, G] theta;
  theta[1] = mu + raw_theta[1] .* sigma;
  for (i in 2:A) theta[i] = theta[i - 1] + raw_theta[i] .* sigma;
}
model {
  vector[N] pred_val;
  for (n in 1:N){
    pred_val[n] = theta[a[n], g[n]];
  }
  target += normal_lpdf(mu | 0, prior_mu);
  target += normal_lpdf(sigma | 0, prior_sigma);
  target += std_normal_lpdf(to_vector(raw_theta));
  target += binomial_logit_lpmf(rejected | submitted, pred_val);
}
```

```r
df_collapse <- df %>%
  filter(age <= 90,
         ethn %in% c("white", "black", "hispanic", "asian", "other"),
         gender != "U") %>%
  mutate(outcome = abs(status - 1),
         ethn_race = ifelse(ethn == "white", 1,
                            ifelse(ethn == "black", 2,
                                   ifelse(ethn == "hispanic", 3,
                                          ifelse(ethn == "asian", 4,
                                                 ifelse(ethn == "other", 5, NA))))),
         gender = ifelse(gender == "M", 1, 0)) %>%
```

```
  arrange(gender, ethn_race) %>%
  mutate(ethn_gender = group_indices(., paste(gender, ethn_race))) %>%
  group_by(age, ethn_gender) %>%
  summarize(ballots = n(),
            rejected = sum(outcome)) %>%
  mutate(age = as.integer(age)) %>%
  arrange(ethn_gender, age)
```

```
## Warning: Problem with `mutate()` input `ethn_gender`.
## i The `...` argument of `group_keys()` is deprecated as of dplyr 1.0.0.
## Please `group_by()` first
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
## i Input `ethn_gender` is `group_indices(., paste(gender, ethn_race))`.
```

```
## Warning: The `...` argument of `group_keys()` is deprecated as of dplyr 1.0.0.
## Please `group_by()` first
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
## `summarise()` regrouping output by 'age' (override with `.groups` argument)
```

```
data <- list(
  N = nrow(df_collapse),
  A = length(unique(df_collapse$age)),
  G = 10,
  submitted = df_collapse %>% pull(ballots),
  rejected = df_collapse %>% pull(rejected),
  g = df_collapse %>% pull(ethn_gender),
  a = df_collapse %>% pull(age) - 17,
  prior_mu = 1,
  prior_sigma = 0.5
)
fit <- rstan::sampling(m1, data = data, chains = 4, cores = 4, warmup = 2000, iter = 4000)
# plot
theta <- inv.logit(rstan::extract(fit, pars = "theta")[[1]]) * 100
medians <- apply(theta, MARGIN = c(2,3), median)
q25   <- apply(theta, MARGIN = c(2,3), function(x) quantile(x, c(0.25)))
q75   <- apply(theta, MARGIN = c(2,3), function(x) quantile(x, c(0.75)))
q10   <- apply(theta, MARGIN = c(2,3), function(x) quantile(x, c(0.10)))
q90   <- apply(theta, MARGIN = c(2,3), function(x) quantile(x, c(0.9)))
df_plot <- data.frame(age = rep(seq(18, 90),5),
                      ethn =  rep(c(rep("white", dim(medians)[1]),
                                    rep("black", dim(medians)[1]),
                                    rep("hispanic", dim(medians)[1]),
                                    rep("asian", dim(medians)[1]),
                                    rep("other", dim(medians)[1])), 2),
                      income =  c(rep("Female", dim(medians)[1] * 5),
                                  rep("Male", dim(medians)[1] * 5)),
                      median = c(medians[,1], medians[,2], medians[,3], medians[,4], medians[,5],
                                 medians[,6], medians[,7], medians[,8], medians[,9], medians[,10]),
                      q25 = c(q25[,1], q25[,2],q25[,3], q25[,4], q25[,5],
                              q25[,6], q25[,7],q25[,8], q25[,9], q25[,10]),
                      q75 = c(q75[,1], q75[,2],q75[,3], q75[,4], q75[,5],
                              q75[,6], q75[,7],q75[,8], q75[,9], q75[,10]),
```

```
                      q10 = c(q10[,1], q10[,2],q10[,3], q10[,4], q10[,5],
                              q10[,6], q10[,7],q10[,8], q10[,9], q10[,10]),
                      q90 = c(q90[,1], q90[,2],q90[,3], q90[,4], q90[,5],
                              q90[,6], q90[,7],q90[,8], q90[,9], q90[,10]))
ggplot(data = df_plot %>%
         mutate(ethn = factor(ethn, levels = c("white", "black", "hispanic", "asian", "other")),
                gender = factor(income, levels = c("Female", "Male"))),
       aes(x = age, y = median, color = income)) +
  geom_point(size = 0.8) +
  geom_errorbar(aes(x = age, ymin = q25, ymax =  q75, color = gender), width = 0, size = 0.75) +
  geom_errorbar(aes(x = age, ymin = q10, ymax = q90, color = gender), width = 0, size = 0.25) +
  theme_bw() +
  scale_x_continuous(breaks = seq(20, 90, 10)) +
  labs(x = "Age",
       y = "Rejection probability (in %)",
       caption = "Thick line: 50%
                  Thin line: 80%",
       title = "Estimated rejection rates by income, race, and age in NC so far",
       color = "Gender") +
  theme(legend.position = "bottom") +
  facet_wrap(.~ethn)
```