GR5065 Homework 4

Ben Goodrich

Due March 28, 2022 at 4PM

```
# call the set.seed function once here to make the knitting conditionally deterministic
# but remember to also pass a seed to any Stan-related function that draws from the posterior
```

1 Voter Turnout in France

Read Andy Eggers' paper on voter turnout in proportional representation (PR) systems compared to plurality-rule systems. In a PR system, voters vote for a political party from a list and the resulting proportion of legislators from each political party is (roughly, perhaps due to things like winner bonuses) equal to the proportion of votes that the party receives. In a plurality-rule system, voters vote for a named candidate (who typically is a member of a political party) and the candidate with the most votes wins the election. The resulting proportion of legislators each political party is equal to the proportion of winning candidates.

It is widely believed that a somewhat larger percentage of eligible voters turn out to vote in PR systems than in plurality-rule systems, but it is difficult to take into account all of the other factors that could drive voter turnout that may be associated (causally or not causally) with the political system. Eggers' estimates this difference in voter turnout among small municipalities in France where there is a national rule that municipalities with at least $3,500$ people must have a PR system and municipalities with fewer than $3,500$ people must have a plurality-rule system. The idea of "regression discontinuity designs" is that the causal effect of a municipality going from a population of $3,499$ to $3,500$ is attributable to the change from a plurality-rule to a PR system, as opposed to merely having one additional person to represent or any other variable that is correlated with the population. Of course, there are not enough municipalities whose population is exactly $3,499$ in one year and exactly $3,500$ the next, but there are many municipalities whose population is close to the threshold of $3,500$. Eggers then chooses a "bandwidth window" for the population that is centered at $3,500$ to estimate a linear regression model on that subset of the data. In this homework, you do not have to worry about how the "bandwidth window" was chosen.

You can load the dataset used in Eggers' paper with

```
library(dplyr)
Eggers <- readRDS("Eggers.rds") %>%
  mutate(PR = as.integer(rrv >= 0)) # has a PR system
```

1.1 Regression Discontuity Designs and Directed Acylic Graphs

Explain, using a DAG, how regression discontinuity designs identify the average causal effect under the Adjustment Criterion. You can use the regression discontinuity design in Eggers' paper if that helps you explain, but the logic of it would be true for any "forcing variable" (which is population in this case), treatment (system of government in this case), and outcome (voter turnout in this case).

1.2 Drawing from the Prior Predictive Distribution

As explained on page 144, Eggers estimates a basic model (before adding additional covariates, which you do not need to worry about on this homework) where the percentage of voter turnout is conditionally normal, given the population of the municipality. There are five unknown parameters to estimate:

1. β_0, which is the intercept
2. τ, which is the coefficient on the dummy variable for a PR system (due to population being at least $3,500$)
3. β_1, which is the coefficient on the logarithm of the ratio of population to $3,500$ so that `rrv` would be zero in the data if the population were exactly $3,500$
4. β_2, which is the coefficient on the interaction between `PR` and `rrv`, which can be interpreted as how much more sensitive voter turnout is to log population in PR systems than in plurality-rule systems
5. σ, which is the standard deviation of the errors when predicting the percentage of voter turnout with only the previous four variables only

Choose priors for the unknowns and draw 1000 realizations from the prior predictive distribution of voter turnout for each of the $N = 35891$ municipalities in the dataset using your own R code (as opposed to calling `stan_glm` with `prior_PD = TRUE`). In this case, you should *not* center the predictors because `Eggers$rrv` is already constructed so that it is zero when the population is exactly $3,500$. Thus, the intercept, β_0, can essentially be interpreted as the expected percentage of voter turnout for a municipality that has $3,499$ people and thus a plurality-rule system.

1.3 Checking the Prior Predictive Distribution

Show with a plot that (the realizations from) your prior predictive distribution seem reasonable for all municipalities with populations between $1,750$ and $5,250$, which you can achieve by filtering on the `PSDC99` variable, which contains the raw population of each municipality.

Then, show numerically that (the realizations from) your prior predictive distribution seem unreasonable for Toulouse, which is the municipality with the largest population $(390,350)$.

Finally, explain how the differences in the prior predictive distribution between Toulouse and municipalities with populations between $1,750$ and $5,250$ would provide motivation to only condition on the municipalities with populations between $1,750$ and $5,250$ when estimating the causal effect of switching from a plurality-rule system to a PR system at $3,500$.

1.4 Posterior Distribution

Use the `stan_glm` function in the rstanarm package to draw from the posterior distribution of these five unknown parameters conditional on the data from municipalities with populations between $1,750$ and $5,250$ only, where the outcome variable (`to.2008`) is the percentage of voter turnout in the 2008 municipal elections. Note that the `prior_intercept` argument to `stan_glm` is separate from the `prior` argument, which pertains to the coefficients and the latter can take a vector of size 3 for `location` and `scale` whose elements correspond to your beliefs about τ, β_1, and β_2 respectively. The `prior_aux` argument specifies the prior for σ, which you can take to be the exponential distribution with `rate` parameter equal to the reciprocal of your prior mean. Finally, you should pass an integer to `seed` to make Stan deterministic.

1.5 Interpretation

How would you describe your posterior beliefs about τ?

1.6 Prediction

There are some municipalities with populations between $1,750$ and $5,250$ where `to.2008` is missing for some reason (and thus were dropped when you called `stan_glm`). You can subset to those observations with something like

```r
Eggers_missing <- filter(Eggers, PSDC99 >= 1750, PSDC99 <= 5250, is.na(to.2008)) %>%
  select(-to.2008)
```

Call the `posterior_predict` function on the object created by `stan_glm` with `newdata = Eggers_missing` to draw from the posterior predictive distribution of voter turnout in these municipalities, conditional on the municipalities with populations between $1,750$ and $5,250$ where `to.2008` was observed. Do these distributions seem reasonable? Why or why not?

2 Minimum Wage Increases

Read Alan Manning's paper, including the appendices, which addresses the question of why increasing the minimum wage seems to have a noticeable effect on wages but has an "elusive" effect on the level of employment, which is to say that economists have not found the large adverse effects on employment that they theorized decades ago would occur.

Download Manning's dataset from here to your working directory. You will have to sign into OpenICPSR by clicking the box that says sign in using Google, at which point you should enter your Columbia email address and password. You can then load it into R with

```r
library(haven)
Manning <- as_factor(read_dta("ManningElusiveEmployment.dta")) %>%
  mutate(agecat = as.factor(agecat),
         quarterly_date = as.factor(quarterly_date - 75))
glimpse(Manning)
```

```
## Rows: 41,818
## Columns: 15
## $ year           <dbl> 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1979, 1~
## $ region         <fct> east south central division, east south central divisio~
## $ statefips      <fct> alabama, alabama, alabama, alabama, alabama, alaska, al~
## $ agecat         <fct> 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, 2, 3, 4~
## $ qtr            <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ popshare       <dbl> 0.1194241, 0.1388525, 0.1032453, 0.3838268, 0.2546513, ~
## $ ln             <dbl> -1.0879076, -0.4460275, -0.3935704, -0.3444984, -0.5263~
## $ lemp           <dbl> 12.53646, 13.32907, 13.08522, 14.44738, 13.85520, 10.28~
## $ lpop           <dbl> 13.62437, 13.77510, 13.47879, 14.79188, 14.38158, 11.23~
## $ ne             <dbl> 826013.62, 960392.75, 714110.19, 2654790.75, 1761330.62~
## $ ur             <dbl> 0.9393107, 0.9393107, 0.9393107, 0.9393107, 0.9393107, ~
## $ statecensus    <fct> "alabama", "alabama", "alabama", "alabama", "alabama", ~
## $ lw             <dbl> 6.087777, 6.294876, 6.277453, 6.401202, 6.350931, 6.306~
## $ quarterly_date <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ lmin           <dbl> 5.669881, 5.669881, 5.669881, 5.669881, 5.669881, 5.828~
```

In the `Manning` dataset, the unit of observation is a state-quarter, i.e. an aggregate of all the people within an age category in a state whose data are observed during some quarter of the year (but the same people are not observed in multiple quarters):

- `year` is an integer vector with values between 1979 and 2019
- `region` is a factor with 9 genuine categories for what region the person lives in
- `statefips` is factor with 51 genuine categories (including Washington D.C.) for what state the person lives in
- `agecat` is a factor representing 5 age categories that people fall in
- `qtr` is a number among 1, 2, 3, and 4 for what quarter of the year the person was observed
- `popshare` is the percentage of teens in the total population for the state that the person lives in during that time period

- `ln` is the difference between `lemp` and `lpop` (see below)
- `lemp` is the logarithm of the number of people employed in this age category, for this state, at this time, etc.
- `lpop` is the logarithm of the number of people in the population in this age category, for this state, at this time, etc.
- `ne` is a weighting variable used in Frequentist analyses
- `ur` is the prime-age (25 to 54 years old) unemployment rate for that state and time period
- `statecensus` is redundant with `statefips`
- `lw` is the log real hourly wage, which is to say that it is measured in 2019 dollars after adjusting for inflation between 2019 and whenever the nominal wage was observed
- `quarterly_date` is a factor indicating the quarter between 1 and 164 that is basically the intersection of `year` and `qtr`
- `lmin` is the logarithm of maximum value between the federal minimum wage and the minimum wage in the `state` that the person lives in during that `time` period, which is in nominal dollars (i.e. not adjusted for inflation) and presumes the person works 40 hours a week

2.1 Directed Acylic Graph

Start with the "stylized representation of the impact of the minimum wage on employment" in Figure 5 on page 23 and then draw a DAG that corresponds to what you think Manning thinks the data-generating process is for "Specification 1" in Table 2. Is the total causal effect of changing the minimum wage identified using the Adjustment Criterion in that DAG?

2.2 Frequentist Inference

Manning uses Ordinary Least Squares and states with regard to Figures 1 through 4 that "we will summarize results and confidence intervals for the coefficient on the log minimum wage in these regressions". To what extent are the inferences Manning draws Frequentist? For which of the inferences Manning wants to make should he have used Bayesian methods?

2.3 Bayesian Inference

Use the `stan_lm` function in the rstanarm package to estimate the parameters of the model (where `ln` is the outcome variable) listed in "Specification 1" of Table 2 separately for `agecat = 1` (which corresponds to $[16, 19]$) and `agecat = 2` (which corresponds to $[20, 24]$). You might need to read `help(formula)` to understand how R processes the syntax but basically factor (or character) variables in R will be converted into a sequence of dummy variables that is one less than the number of levels of the categorical variable, with no dummy variable for the reference category.

You will need to specify both `prior_intercept`, which can be a call to the `normal` function whose `location` and `scale` arguments should correspond to your beliefs about the distribution of the average log employment rate for people in a particular age bracket. Technically, it should be the distribution of the log employment rate for a person with average values on all predictors, but that number is not too different from the unconditional average. You will also need to specify `prior = R2(...)` to express your prior beliefs about the proportion of variance in log employment explained by all the predictors in "Specification 1" of Table 2 under a linear model. Finally, you should pass an integer to the `seed` argument to make Stan conditionally deterministic.

2.4 Interpretation

How would you describe your posterior beliefs about the coefficient(s) on the logarithm of the minimum wage (in the two models)? To what extent are your conclusions similar to those of Manning?

2.5 Prediction

You may have seen in the news that Democrats like Bernie Sanders have been pushing to raise the federal minimum wage to $15 per hour, which would be much greater than its current value and larger than in many states and cities. However, such a change in the law is opposed by all Republicans in the Senate, and several of them are needed to at least allow a vote to happen on the minimum wage (as opposed to preventing the vote using a filibuster) and it is not clear that all 50 Democratic Senators would support an increase in the federal minimum wage to $15 per hour.

You can select the most recent observations on teenagers with something like

```
recent <- filter(Manning, agecat == 1, year == 2019, qtr == 4)
```

Call the `posterior_predict` function on the object created by `stan_lm` with `newdata = recent` in order to obtain posterior predictions for log employment rate, conditional on all of the past data. Then, use the `exp` function to convert these predictions to employment rates.

Next, create a counterfactual dataset where the federal minimum wage is raised to $15 if the state's minimum wage was less than $15 via

```
recent_ <- mutate(recent, lmin = pmax(exp(lmin), 15 * 40))
```

and call the `posterior_predict` function on the object created by `stan_lm` with `newdata = recent_` to obtain predictions of log employment rates that can be mapped to employment rates via `exp`.

How would you describe your posterior beliefs about the effect on teenage employment rates of increasing the federal minimum wage to $15 per hour if such an increase were to have gone into effect in 2019?

2.6 Prior Predictive Distribution

On page 8, Manning says

> One problem with the first specification is that it implies that the elasticity of teen wages with respect to the minimum wage is a constant, whatever the level of the minimum wage. This is implausible as a universal model, because it predicts that a rise in the minimum wage from $1 to $1.10 has the same impact as a rise from $10 to $11 dollars. We would expect the marginal effect of changes in the minimum wage on wages to be increasing in the minimum wage - a very low minimum wage relative to the prevailing level of wages (what Lee 1999 terms the effective minimum) will have little impact and a higher minimum wage a larger impact. The simplest way to investigate non-linearity is to include a quadratic term in the effective minimum. Details of the procedure (which follows Autor, Manning, and Smith 2016) and the estimates are in the online Appendix, but for the wages of teenagers, one can detect a non-linear effect.

Appendix B has some additional details. Essentially, Manning uses a linear model (in the "first stage") to predict the logarithm of wages among people who are 30 or over and then (in the "second stage") uses those predictions — rather than the actual log-wages among people who are 30 or over — in a model that is linear in the *coefficients* but includes a quadratic function of the minimum wage relative to the predicted prime-age wage.

If instead Manning had used a Bayesian approach, write code to draw 1000 times from the prior predictive distribution of `lw` for people where `agecat = 1`, under this assumed data-generating process. You will have to choose reasonable priors on the coefficients (including shifts for each state and quarter), intercepts, and error standard deviations in both stages of the process. Then, plot the prior predictive distribution of `lw` for people where `agecat = 1` to show that it is reasonable.