

GR5065 HW3

Load the data on Presidential vetoes of bills passed by the U.S. Congress by calling

```
vetoos <- readr::read_csv("vetoos.csv", show_col_types = FALSE)
```

These data were assembled at

<https://www.presidency.ucsb.edu/statistics/data/presidential-vetoos>

which also has a brief explanation of the process and the variables.

With few exceptions, the party with the majority of the seats in either House of Representatives or the Senate gets to decide which bills get voted on. If the same bill receives a majority of votes cast separately in both the House and the Senate, then the President has an opportunity to sign it or veto it within 10 days. If the bill is vetoed, the House and the Senate have the opportunity to override a Presidential veto, but it requires another vote and two-thirds of the votes need to be in favor of the bill in both the House and the Senate to make it a law.

In addition, if the bill passes Congress with fewer than 10 days in the legislative session, then the President can essentially do nothing, which is called a “pocket veto”. Congress cannot override a pocket veto because the legislative session ends, but it can start the bill approval process from scratch in a subsequent legislative session.

Legislative sessions index two-year terms, after which there are elections for all of the seats in the House of Representatives and about one-third of the seats in the Senate. Joe Biden has vetoed one bill that was passed last week in the 118th term of Congress.

1 Frequentism

Frequentists attempt to derive the probability distribution of various functions of a sample of size N from a population that is characterized by parameters. Given values of these parameters, it is often possible to derive the probability of such a function, such as an estimator of those parameters.

However, in the above example on Presidential vetoes — and many examples with historical data — the data is not a random sample from any population. In this case, it contains literally every veto that has been issued in the United States since the constitution was ratified in 1789. Nor is it plausible to imagine that there any dataset other than **vetoes** could be collected with the same sequence of values on **House** and **Senate**.

And yet, many social scientists would utilize Frequentist with **vetoes** or similar historical data. What consequences does this have for inference? Why do social scientists do so?

2 Generative Model

1. Draw all the unknown parameters in your model $R = 10000$ times
2. Use each of those R parameter realizations to
 - Draw the number of regular (not pocket) vetoes in each of the $N = 123$ Congressional terms (i.e., excluding the 39th and the 40th because **House** and **Senate** are **NA**; Andrew Johnson became president when Abraham Lincoln was assassinated but Johnson and Lincoln were of different parties, albeit united against the states that were trying to secede)
 - Draw the number of overrides of regular vetoes in each of the N Congressional terms, using the realizations of regular vetoes from the previous part.

Both parts of (2) should — in some way — depend on the percentage of seats the President's party holds in the House and Senate. All of the relevant fields are in the **vetoes** dataset, although you may want to create additional variables that are derived from them (including but not limited to interaction terms).

In particular, in (2A) the number of regular vetoes in Congressional term n should be drawn from a Poisson distribution with expectation $\mu_n \epsilon_n$, where ϵ_n is drawn from a Gamma distribution with shape and rate ϕ and $\mu_n = e^{\eta_n}$ where η_n is the linear predictor. Then, in (2B) the number of regular vetoes overridden in Congressional term n should be drawn from a binomial distribution with **size** the number of regular vetoes and success **probability** $\pi_n = \frac{1}{1+e^{-\lambda_n}}$ where the log-odds, λ_n , is another linear predictor that is not necessarily the same function as for η_n (or at least your priors should differ).

3 Prior Prediction

Use the `geom_bar` function in `ggplot2` to plot the prior predictive distribution of the number of regular vetoes overridden, but also use `facet_wrap` to stratify into four subplots based on whether or not the President's party holds a majority of the seats in the House and Senate.

This may require that you make a copy of `House` and `Senate` from `veto`s each time you draw from the distribution of the outcomes in the previous problem so that these two variables are present in the `data.frame` of draws. In what senses are these four subplots collectively reasonable?

4 Posterior | Regular Vetoes

Use the `stan_glm.nb` function in the `rstanarm` package to draw from the posterior distribution of all of the parameters that are involved in the part of the previous generative model that pertain to the number of regular vetoes; i.e. `Regular` is the outcome. Is the error variance — which drives the overdispersion in the outcomes — large or small?

5 Posterior Prediction

Draw from the posterior predictive distribution of the number of bills that Biden would veto in the (current) 118th Congress if, hypothetically, the Democrats only controlled 49 seats in the Senate (they actually have 51). Plot this distribution using `geom_bar` in `ggplot2`.

6 Posterior | Overrides

Use the `stan_glm` function in the `rstanarm` package to draw from the posterior distribution of all the parameters that are involved in the part of the previous generative model that pertain to the number of overrides; i.e. the outcome syntax is `cbind(Overrides, Regular - Overrides)`. How would you describe the posterior probability of the (current) 118th Congress overriding any bill that Biden vetoes?