

# POL 350C, Homework 4

April 21st, 2016

Assigned: 4/27/2017

Due: 5/4/2017

In this problem set we're going to analyze candidate choice and the use of political language using the models we've developed in class.

## 1 Candidate Choice

In this problem we're going to analyze preferences for Republican presidential candidates in May 2015. (They weren't asking about Trump back then, so the included candidates are Bush, Cruz, Huckabee, and Paul). The data are contained in `ChoiceData.RData` and includes the following variables

- 1) `fav`: the most preferred candidate for the respondent
- 2) `gay`: is 1 if the respondent agrees with the statement "Homosexuality should be accepted by society." It's 0 if they agree with the statement "Homosexuality should be discouraged by society."
- 3) `trade`: is 1 if the respondent thinks free trade agreement are a good thing, and trade is 0 if the respondent thinks free trade agreements are a bad thing

Using this data set, answer the following questions

- a) Calculate the rate of support for the candidates. Using a cross-tab, show how candidate choice covaries with the values of `gay`. In a separate cross-tab show how candidate choice covariates with `trade`. What do you notice?
- b) Using `mlogit` estimate a model of candidate choice as a function of respondent's view on `gay` and `trade`. Report the coefficients and standard errors.
- c) Using the maximum likelihood estimates for the coefficients from (b), report the estimated "effect" of favoring free trade agreements across candidate choice, averaging over all respondents and leaving the respondent's value of `gay` unchanged. Report a 95% confidence interval for this estimated
- d) Suppose Huckabee and Paul drop out of the race. What is the predicted vote margin for Cruz and Bush after the candidates drop out?

## 2 Analysis of Political Language

Politicians use carefully crafted language to criticize opponents and to garner support from constituents. This was particularly apparent during the battle to pass the Affordable Care Act. In this problem we're going to analyze the number of time members of the House used the phrases "Obamacare"—a derogatory label for the legislation and "government takeover" a description of the government's larger role in health care after the ACA's passage.

In the data set `SpeechDat.RData` we have the following variables:

- 1) `Year` : year of press release (either 2009 or 2010, be sure to include this as a factor)
  - 2) `ICPSR`: a unique identifier for the legislator
  - 3) `obamacare`: number of times "obamacare" appears in legislators' press releases
  - 4) `govTakeover`: number of times "government takeover" appears in legislators' press releases
  - 5) `dem`: An indicator vector that equals 1 if the legislator is a Democrat
  - 6) `dw_nom`: a one-dimensional measure of legislator ideology. Negative values are more liberal, positive values are conservative.
- a) Suppose that we assume the number of times "obamacare" appears in legislators' press releases is distributed according to a Poisson distribution.  $\text{Obamacare}_i \sim \text{Poisson}(\lambda)$ . (Note, we're assuming there is only one rate parameter)
- i) Calculate the maximum likelihood estimate of the rate `obamacare` is used,  $\lambda$ , and provide it along with a measure of the standard error (using a method described in class).
  - ii) Using the maximum likelihood estimate, compare the observed incidence of "obamacare" to the predictions from a poisson distribution with rate parameter equal to the maximum likelihood estimate. What do you notice?
- b) Now suppose that we want to model the rate "obamacare" appears in legislators' press releases as a function of legislator characteristics. Specifically, we want to use a Poisson regression to model the count of `obamacare` as a function of: `dem`, `dw_nom`, and `Year` (which again, we want to include as a dummy variable for each year).
- i) Write out the likelihood and log-likelihood for the  $\beta$  vector.
  - ii) In R write a function to evaluate the log-likelihood for a vector of regression coefficients  $\beta$  given a set of counts and covariates.
  - iii) Using your function from (ii), use `optim` to obtain maximum likelihood estimate of the regression coefficients and standard errors for those coefficients. Report the coefficients and standard errors in a table.

- c) Using those maximum likelihood estimates, we are interested in exploring two relationships:
- 1) Differences between Democrats and Republicans
    - i) Obtain the median DW-Nominate score among Democrats and the median DW-Nominate score among Republicans.
    - ii) Using those obtained values, compare the rate the median Democrat uses the phrase Obamacare to the rate the median Republican uses the phrase Obamacare. Calculate a separate difference in 2009 and 2010.
    - iii) Using the multivariate normal based-simulation, calculate uncertainty for this difference. How much larger is the rate for Republicans?
  - 2) Differences across the ideological spectrum
    - i) Using the data plot the count of the number of times Obamacare is used against the DW-Nominate score.
    - ii) Now using the model, plot the expected rate Obamacare is used across the ideological spectrum. Use simulation to obtain 95-percent confidence intervals for the relationship.
    - iii) Finally, suppose we're interested in the probability a legislator uses the word "Obamacare" at least once. Calculate the probability a legislator uses Obamacare at least once across the ideological spectrum and obtain 95-percent confidence intervals for that probability. Hint: recall that if  $Y_i \in \{0, 1, \dots, \}$  then  $P(Y_i \geq 1) = 1 - P(Y_i = 0)$