faraway-glm.Rmd

Robesyn

25 February 2017

Code Faraway GLM

## Warning: package 'tidyverse' was built under R version 3.3.2

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Warning: package 'tibble' was built under R version 3.3.1

## Warning: package 'tidyr' was built under R version 3.3.1

## Warning: package 'purrr' was built under R version 3.3.1

## Conflicts with tidy packages ----------------------------------------------

## filter(): dplyr, stats  
## lag(): dplyr, stats

## Warning: package 'faraway' was built under R version 3.3.2

head(gavote)

## equip econ perAA rural atlanta gore bush other votes ballots  
## APPLING LEVER poor 0.182 rural notAtlanta 2093 3940 66 6099 6617  
## ATKINSON LEVER poor 0.230 rural notAtlanta 821 1228 22 2071 2149  
## BACON LEVER poor 0.131 rural notAtlanta 956 2010 29 2995 3347  
## BAKER OS-CC poor 0.476 rural notAtlanta 893 615 11 1519 1607  
## BALDWIN LEVER middle 0.359 rural notAtlanta 5893 6041 192 12126 12785  
## BANKS LEVER middle 0.024 rural notAtlanta 1220 3202 111 4533 4773

glimpse(gavote)

## Observations: 159  
## Variables: 10  
## $ equip (fctr) LEVER, LEVER, LEVER, OS-CC, LEVER, LEVER, OS-CC, OS-P...  
## $ econ (fctr) poor, poor, poor, poor, middle, middle, middle, middl...  
## $ perAA (dbl) 0.182, 0.230, 0.131, 0.476, 0.359, 0.024, 0.079, 0.079...  
## $ rural (fctr) rural, rural, rural, rural, rural, rural, urban, urba...  
## $ atlanta (fctr) notAtlanta, notAtlanta, notAtlanta, notAtlanta, notAt...  
## $ gore (int) 2093, 821, 956, 893, 5893, 1220, 3657, 7508, 2234, 164...  
## $ bush (int) 3940, 1228, 2010, 615, 6041, 3202, 7925, 14720, 2381, ...  
## $ other (int) 66, 22, 29, 11, 192, 111, 520, 552, 46, 52, 709, 40, 7...  
## $ votes (int) 6099, 2071, 2995, 1519, 12126, 4533, 12102, 22780, 466...  
## $ ballots (int) 6617, 2149, 3347, 1607, 12785, 4773, 12522, 23735, 574...

summary(gavote)

## equip econ perAA rural atlanta   
## LEVER:74 middle:69 Min. :0.0000 rural:117 Atlanta : 15   
## OS-CC:44 poor :72 1st Qu.:0.1115 urban: 42 notAtlanta:144   
## OS-PC:22 rich :18 Median :0.2330   
## PAPER: 2 Mean :0.2430   
## PUNCH:17 3rd Qu.:0.3480   
## Max. :0.7650   
## gore bush other votes   
## Min. : 249 Min. : 271 Min. : 5.0 Min. : 832   
## 1st Qu.: 1386 1st Qu.: 1804 1st Qu.: 30.0 1st Qu.: 3506   
## Median : 2326 Median : 3597 Median : 86.0 Median : 6299   
## Mean : 7020 Mean : 8929 Mean : 381.7 Mean : 16331   
## 3rd Qu.: 4430 3rd Qu.: 7468 3rd Qu.: 210.0 3rd Qu.: 11846   
## Max. :154509 Max. :140494 Max. :7920.0 Max. :263211   
## ballots   
## Min. : 881   
## 1st Qu.: 3694   
## Median : 6712   
## Mean : 16927   
## 3rd Qu.: 12251   
## Max. :280975

Per county (rows)

gavote$undercount <- (gavote$ballots - gavote$votes)/gavote$ballots  
summary(gavote$undercount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.02779 0.03983 0.04379 0.05647 0.18810

percounty<- gavote %>% mutate(diff = ballots-votes, undercount = diff/ballots)  
summary(percounty$undercount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.02779 0.03983 0.04379 0.05647 0.18810

Note this is different than overall, for all counties together

with(gavote, sum(ballots-votes)/sum(ballots))

## [1] 0.03518021

overall<-gavote %>% summarize(sumdiff = sum(ballots-votes), sum = sum(ballots), proportionUndervotes = round(sumdiff/sum,3) )  
overall$proportionUndervotes

## [1] 0.035