

## 15.10

Nikhil Gopal

2/15/2022

```
rm(list = ls())
library(rstanarm)
library(rosdata)
```

### 15.10

Logistic regression and choice models: Using the individual-level survey data from the election example described in Section 10.9 (data in the folder NES), fit a logistic regression model for the choice of supporting Democrats or Republicans. Then interpret the output from this regression in terms of a utility/choice model.

```
data <- rosdata::nes
data$martial_status <- as.factor(data$martial_status)
data$south <- as.factor(data$south)
data$income <- as.factor(data$income)

fit <- stan_glm(rvote ~ income+martial_status+south, family = binomial(link = "logit"), data = data, re
print(fit)
```

```
## stan_glm
## family:      binomial [logit]
## formula:     rvote ~ income + martial_status + south
## observations: 13794
## predictors:  11
## -----
##              Median MAD_SD
## (Intercept)   -0.3    0.1
## income2        0.1    0.1
## income3        0.3    0.1
## income4        0.5    0.1
## income5        1.2    0.1
## martial_status2 -0.4    0.1
## martial_status3 -0.1    0.1
## martial_status4 -0.8    0.1
## martial_status5  0.1    0.1
## martial_status7 -1.2    0.2
## south1         -0.1    0.0
##
## -----
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

I fit a logistic regression model to determine whether certain choices like affect one's likelihood of voting for Republicans. Since this is a choice/utility model, I did not include predictors such as age, sex, or race, as these models are meant to be interpreted for an individual's choices effects on utility, and one cannot control their race or their sex. One could argue that one's income is not entirely in their control, but we always have to make assumptions and since people in general can choose how many hours they work or their profession, I will consider it a choice. I also made income a factor since the income variable is classed into 5 categories.

Lets exponentiate the logistic coefficients so that we may interpret them (I rounded to 2 digits):

```
round(exp(coef(fit)), 2)
```

```
##      (Intercept)      income2      income3      income4      income5
##           0.77           1.14           1.37           1.62           3.28
## marital_status2 marital_status3 marital_status4 marital_status5 marital_status7
##           0.66           0.89           0.45           1.10           0.31
##           south1
##           0.94
```

We see that belonging to a higher income category increases of voting Republican by 1.25 times. Being in marital statuses 2-4 and 7 decrease the odds of voting Republican relative to the baseline (marital status = 1), while being in marital status 5 increases the odds of voting by 1.15 times. Finally, being a southerner has a slight decreasing effect on odds of voting Republican.