Stat 245 – HW Int Search Selection/Prediction

Adham Rishmawi

October 19, 2022

```
search1 <- read_csv('https://sldr.netlify.app/data/election_searches.csv')
```

```
search_nb2 <- glmmTMB(Searches ~ Race_Ethnicity + Vote_Sway,
                      data = search1,
                      family = nbinom2(link = 'log'))
summary(search_nb2)
```

```
##  Family: nbinom2  ( log )
## Formula:          Searches ~ Race_Ethnicity + Vote_Sway
## Data: search1
##
##      AIC      BIC   logLik deviance df.resid
##  17763.3  17936.9  -8851.7  17703.3     2376
##
##
## Dispersion parameter for nbinom2 family (): 1.12
##
## Conditional model:
##                                                    Estimate Std. Error z value
## (Intercept)                                        3.914193   0.674819   5.800
## Race_EthnicityNA_Black_Hispanic_NA_NA_NA          -1.376982   0.785505  -1.753
## Race_EthnicityNA_Black_Hispanic_NA_NA_OtherRace   -2.135361   1.230145  -1.736
## Race_EthnicityNA_Black_NA_Asian_NA_NA             -0.922758   1.181611  -0.781
## Race_EthnicityNA_Black_NA_NA_AmerInd_NA           -0.909854   1.181610  -0.770
## Race_EthnicityNA_Black_NA_NA_AmerInd_OtherRace    -2.802710   1.296022  -2.163
## Race_EthnicityNA_Black_NA_NA_NA_NA                -1.048063   0.679096  -1.543
## Race_EthnicityNA_Black_NA_NA_NA_OtherRace         -1.204623   0.966449  -1.246
## Race_EthnicityNA_NA_Hispanic_Asian_NA_NA           0.940086   0.951374   0.988
## Race_EthnicityNA_NA_Hispanic_NA_AmerInd_NA        -1.620220   1.202645  -1.347
## Race_EthnicityNA_NA_Hispanic_NA_NA_NA             -1.062275   0.681305  -1.559
## Race_EthnicityNA_NA_NA_Asian_NA_NA                -0.998524   0.680203  -1.468
## Race_EthnicityNA_NA_NA_NA_AmerInd_NA              -1.271328   0.727588  -1.747
## Race_EthnicityNA_NA_NA_NA_NA_OtherRace            -1.922535   0.711874  -2.701
## Race_EthnicityWhite_Black_Hispanic_NA_AmerInd_NA  -2.796174   1.033496  -2.706
## Race_EthnicityWhite_Black_Hispanic_NA_NA_OtherRace -1.976905   1.220330  -1.620
## Race_EthnicityWhite_Black_NA_NA_AmerInd_NA        -1.843036   0.981630  -1.878
## Race_EthnicityWhite_Black_NA_NA_NA_NA             -1.175916   0.703717  -1.671
## Race_EthnicityWhite_Black_NA_NA_NA_OtherRace      -0.918457   1.181576  -0.777
## Race_EthnicityWhite_NA_Hispanic_Asian_NA_NA       -1.249158   1.190476  -1.049
## Race_EthnicityWhite_NA_Hispanic_NA_AmerInd_NA     -1.134007   0.803722  -1.411
## Race_EthnicityWhite_NA_Hispanic_NA_NA_NA          -1.562990   0.692052  -2.258
## Race_EthnicityWhite_NA_NA_Asian_NA_NA             -1.756141   0.713008  -2.463
```

```
## Race_EthnicityWhite_NA_NA_Asian_NA_OtherRace       -1.712679   1.207154  -1.419
## Race_EthnicityWhite_NA_NA_NA_AmerInd_NA            -1.371960   0.702968  -1.952
## Race_EthnicityWhite_NA_NA_NA_AmerInd_OtherRace     -1.397205   0.970051  -1.440
## Race_EthnicityWhite_NA_NA_NA_NA_NA                 -1.298687   0.675223  -1.923
## Race_EthnicityWhite_NA_NA_NA_NA_OtherRace          -2.614973   0.918253  -2.848
## Vote_Sway                                           0.004301   0.006355   0.677
##                                                    Pr(>|z|)
## (Intercept)                                        6.62e-09 ***
## Race_EthnicityNA_Black_Hispanic_NA_NA_NA            0.07960 .
## Race_EthnicityNA_Black_Hispanic_NA_NA_OtherRace     0.08259 .
## Race_EthnicityNA_Black_NA_Asian_NA_NA               0.43484
## Race_EthnicityNA_Black_NA_NA_AmerInd_NA             0.44129
## Race_EthnicityNA_Black_NA_NA_AmerInd_OtherRace      0.03058 *
## Race_EthnicityNA_Black_NA_NA_NA_NA                  0.12275
## Race_EthnicityNA_Black_NA_NA_NA_OtherRace           0.21260
## Race_EthnicityNA_NA_Hispanic_Asian_NA_NA            0.32309
## Race_EthnicityNA_NA_Hispanic_NA_AmerInd_NA          0.17791
## Race_EthnicityNA_NA_Hispanic_NA_NA_NA               0.11895
## Race_EthnicityNA_NA_NA_Asian_NA_NA                  0.14211
## Race_EthnicityNA_NA_NA_NA_AmerInd_NA                0.08058 .
## Race_EthnicityNA_NA_NA_NA_NA_OtherRace              0.00692 **
## Race_EthnicityWhite_Black_Hispanic_NA_AmerInd_NA    0.00682 **
## Race_EthnicityWhite_Black_Hispanic_NA_NA_OtherRace  0.10524
## Race_EthnicityWhite_Black_NA_NA_AmerInd_NA          0.06045 .
## Race_EthnicityWhite_Black_NA_NA_NA_NA               0.09472 .
## Race_EthnicityWhite_Black_NA_NA_NA_OtherRace        0.43697
## Race_EthnicityWhite_NA_Hispanic_Asian_NA_NA         0.29404
## Race_EthnicityWhite_NA_Hispanic_NA_AmerInd_NA       0.15826
## Race_EthnicityWhite_NA_Hispanic_NA_NA_NA            0.02392 *
## Race_EthnicityWhite_NA_NA_Asian_NA_NA               0.01378 *
## Race_EthnicityWhite_NA_NA_Asian_NA_OtherRace        0.15596
## Race_EthnicityWhite_NA_NA_NA_AmerInd_NA             0.05098 .
## Race_EthnicityWhite_NA_NA_NA_AmerInd_OtherRace      0.14977
## Race_EthnicityWhite_NA_NA_NA_NA_NA                  0.05444 .
## Race_EthnicityWhite_NA_NA_NA_NA_OtherRace           0.00440 **
## Vote_Sway                                           0.49847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
hyp_data1 <- expand.grid(Race_Ethnicity = "NA_NA_Hispanic_NA_NA_NA",
                         Vote_Sway = seq(from = -20, to = 20, by = 1)
                         )
hyp_data1
```

```
##             Race_Ethnicity Vote_Sway
## 1  NA_NA_Hispanic_NA_NA_NA       -20
## 2  NA_NA_Hispanic_NA_NA_NA       -19
## 3  NA_NA_Hispanic_NA_NA_NA       -18
## 4  NA_NA_Hispanic_NA_NA_NA       -17
## 5  NA_NA_Hispanic_NA_NA_NA       -16
## 6  NA_NA_Hispanic_NA_NA_NA       -15
## 7  NA_NA_Hispanic_NA_NA_NA       -14
## 8  NA_NA_Hispanic_NA_NA_NA       -13
## 9  NA_NA_Hispanic_NA_NA_NA       -12
## 10 NA_NA_Hispanic_NA_NA_NA       -11
```

```
## 11 NA_NA_Hispanic_NA_NA_NA       -10
## 12 NA_NA_Hispanic_NA_NA_NA        -9
## 13 NA_NA_Hispanic_NA_NA_NA        -8
## 14 NA_NA_Hispanic_NA_NA_NA        -7
## 15 NA_NA_Hispanic_NA_NA_NA        -6
## 16 NA_NA_Hispanic_NA_NA_NA        -5
## 17 NA_NA_Hispanic_NA_NA_NA        -4
## 18 NA_NA_Hispanic_NA_NA_NA        -3
## 19 NA_NA_Hispanic_NA_NA_NA        -2
## 20 NA_NA_Hispanic_NA_NA_NA        -1
## 21 NA_NA_Hispanic_NA_NA_NA         0
## 22 NA_NA_Hispanic_NA_NA_NA         1
## 23 NA_NA_Hispanic_NA_NA_NA         2
## 24 NA_NA_Hispanic_NA_NA_NA         3
## 25 NA_NA_Hispanic_NA_NA_NA         4
## 26 NA_NA_Hispanic_NA_NA_NA         5
## 27 NA_NA_Hispanic_NA_NA_NA         6
## 28 NA_NA_Hispanic_NA_NA_NA         7
## 29 NA_NA_Hispanic_NA_NA_NA         8
## 30 NA_NA_Hispanic_NA_NA_NA         9
## 31 NA_NA_Hispanic_NA_NA_NA        10
## 32 NA_NA_Hispanic_NA_NA_NA        11
## 33 NA_NA_Hispanic_NA_NA_NA        12
## 34 NA_NA_Hispanic_NA_NA_NA        13
## 35 NA_NA_Hispanic_NA_NA_NA        14
## 36 NA_NA_Hispanic_NA_NA_NA        15
## 37 NA_NA_Hispanic_NA_NA_NA        16
## 38 NA_NA_Hispanic_NA_NA_NA        17
## 39 NA_NA_Hispanic_NA_NA_NA        18
## 40 NA_NA_Hispanic_NA_NA_NA        19
## 41 NA_NA_Hispanic_NA_NA_NA        20
```

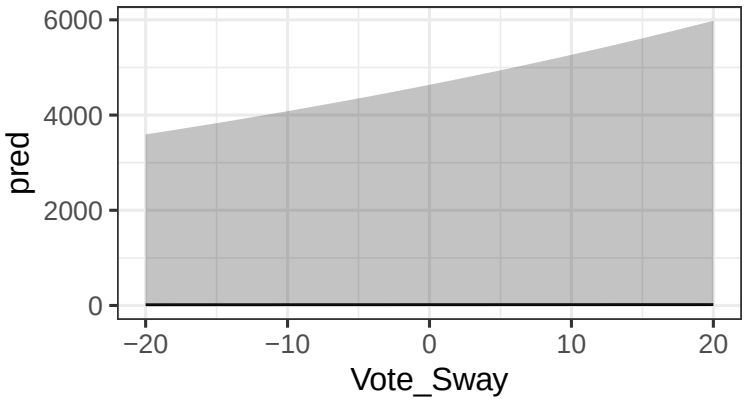
This potential dataset is obtaining data from the two constant variables i used. I used NA_NA_Hispanic_NA_NA_NA and Vote_Sway from -20 and 20.

###Predictions

```
search_preds1 <- predict(search_nb2,
                         newdata = hyp_data1,
                         type = 'link',
                         se.fit = TRUE)
```

```
hyp_data1 <- hyp_data1 |>
  mutate(pred = exp(search_preds1$fit),
         ci_low = exp(search_preds1$fit - 1.96*search_preds1$fit),
         ci_high = exp(search_preds1$fit + 1.96*search_preds1$fit))
```

```
gf_line(pred ~ Vote_Sway,
        data = hyp_data1) |>
  gf_ribbon(ci_low + ci_high ~ Vote_Sway)
```



This prediction plot has completely failed us and shows us a plot with high uncertainty. I would conclude that this variable is probably not the most viable option. I would try something different from Race_Ethnicity.

###Model Selection

```
search_nb2 <- search_nb2 |>
update(na.action = 'na.fail')
search1_dred <- dredge(search_nb2, rank = 'AIC')

search1_dred
```

```
## Global model call: glmmTMB(formula = Searches ~ Race_Ethnicity + Vote_Sway, data = search1,
##     family = nbinom2(link = "log"), ziformula = ~0, dispformula = ~1,
##     na.action = "na.fail")
## ---
## Model selection table
##   cnd((Int)) dsp((Int)) cnd(Rac_Eth) cnd(Vot_Swy) df    logLik     AIC delta
## 2      3.912          +            +              29 -8851.901 17761.8  0.00
## 4      3.914          +            +     0.004301 30 -8851.672 17763.3  1.54
## 1      2.664          +                           2 -8892.808 17789.6 27.81
## 3      2.660          +                 0.007662  3 -8892.088 17790.2 28.37
##   weight
## 2  0.684
## 4  0.316
## 1  0.000
## 3  0.000
## Models ranked by AIC(x)
```

I can conclude that my variable Race_ethnicity was the wrong choice because it failed my prediction plot. I should search for a different variable potentially like Sex or or Group as a better variable of choice. Our AIC isnt much affected from our variable so this even reinforces our decision to change or variables.