Stat 245 – Search Count Model

Gloria Grace

October 13, 2022

Data

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

```
## Rows: 2406 Columns: 18
## -- Column specification --------------------------------------
## Delimiter: ","
```

```
## chr (13): Pre_Choice, Post_Choice, Group, Trust_Search, Party, Poli
## dbl  (5): Searches, Vote_Sway, Pre_Vote, Post_Vote, Age
##
## i Use `spec()` to retrieve the full column specification for this da
## i Specify the column types or set `show_col_types = FALSE` to quiet
```

Plan

For this model, the predictors that I want to include which I think is relevant to the number of internet searches someone does per week (searches) are Age, Education, and Sex. One of the reasons I included both age and education is because I expect that the higher education, the higher the number of searches it will be. For Age, I expect people in the 30s below might have higher search count compares to people in their 60s and above.

Fit

```
search_nb1 <- glmmTMB(Searches ~ Age + Education + Sex, data = search,
                      family = nbinom1(link = 'log'))
```

```
## as(<dgCMatrix>, "dgTMatrix") is deprecated since Matrix 1.5-0; do as
```

```
search_nb2 <- glmmTMB(Searches ~ Age + Education + Sex, data = search,
                      family = nbinom2(link = 'log'))
```

Check which is better:

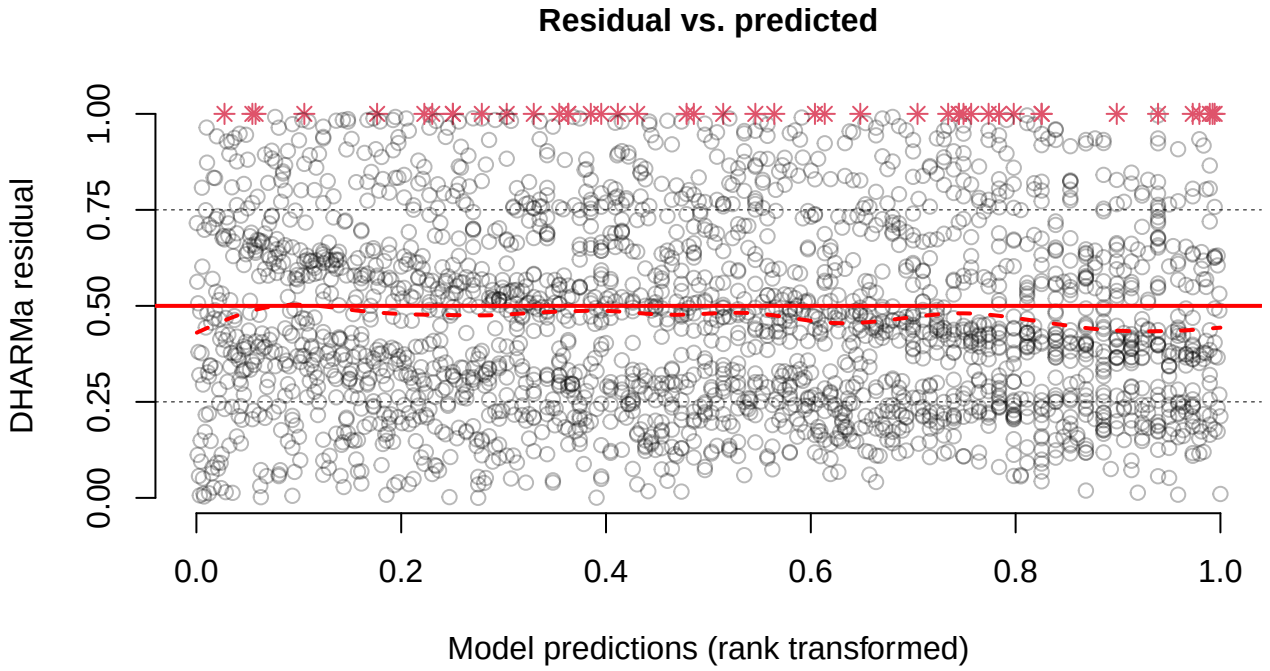
```
AIC(search_nb1, search_nb2)
```

```
##             df      AIC
## search_nb1 10 17705.57
## search_nb2 10 17679.99
```

Model Assessment

Scaled Residuals

```
nb2_sim <- simulateResiduals(search_nb2)
plotResiduals(nb2_sim,
              quantreg = FALSE)
```




```
search <- search |>
  mutate(nb2_pred = rank(predict(search_nb2,
                                 type = 'response')),
         nb2_scaled_resid = nb2_sim$scaledResiduals)
```

```
gf_point(nb2_scaled_resid ~ nb2_pred,
         data = search) |>
  gf_labs(y = 'Scaled Resid.', x = 'Predicted Search')
```