

# Stat 245 – HW Internet Search Models

Adham Rishmawi

October 18, 2022

But today, your task is to fit a regression model to understand how the number of internet searches someone does per week relates to demographic information. -> I will do this by making my response variable Searches and my predictors Race\_Ethnicity

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

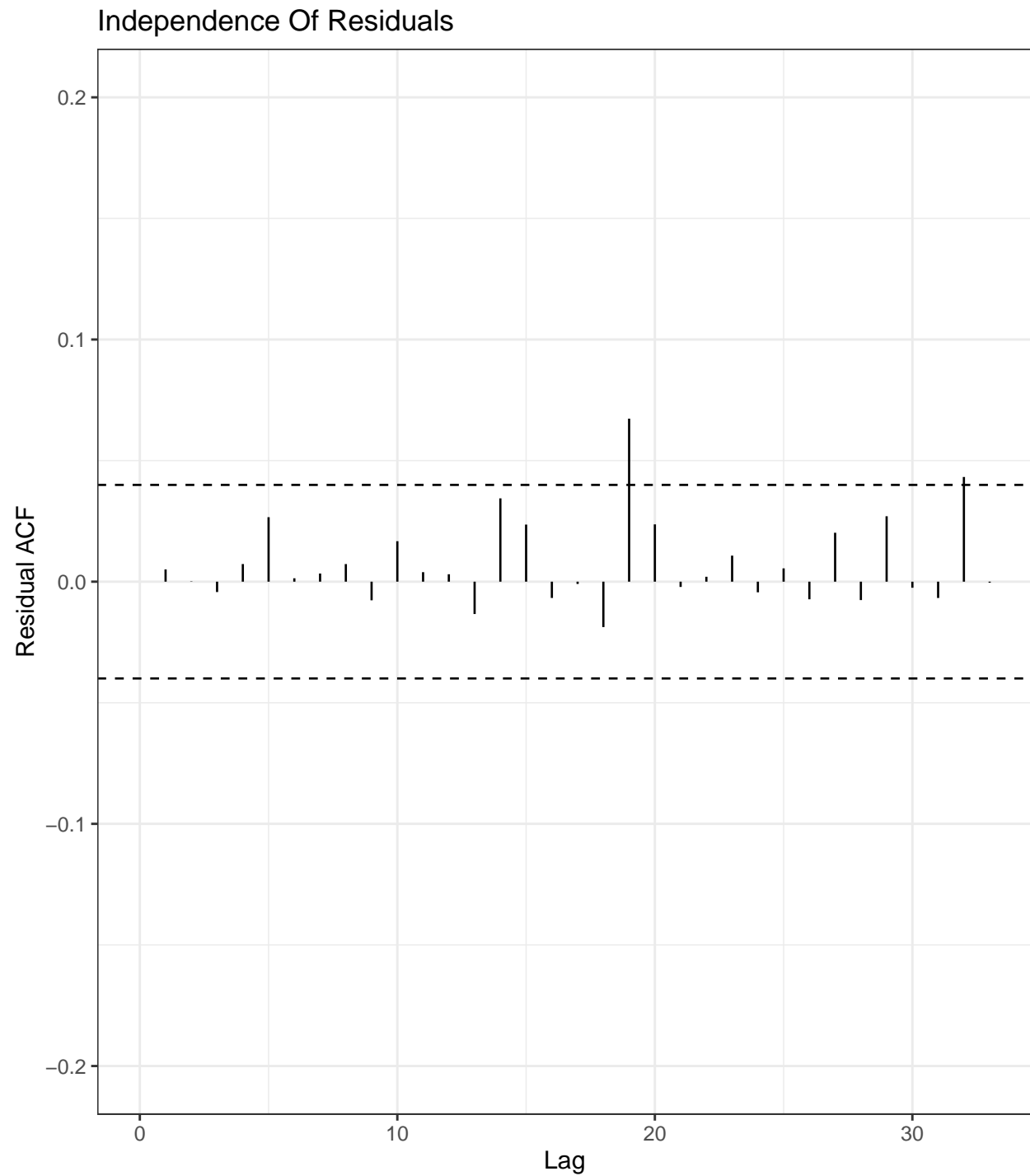
```
slr <- lm(Searches ~ Race_Ethnicity + Vote_Sway,
          data = search)
summary(slr)
```

```
##
## Call:
## lm(formula = Searches ~ Race_Ethnicity + Vote_Sway, data = search)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -120.052   -9.199   -3.990    1.270   236.375
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                        50.02607    17.87480   2.799
## Race_EthnicityNA_Black_Hispanic_NA_NA_NA -37.36810    20.64011  -1.810
## Race_EthnicityNA_Black_Hispanic_NA_NA_OtherRace -44.18251    30.96465  -1.427
## Race_EthnicityNA_Black_NA_Asian_NA_NA -30.07822    30.96066  -0.971
## Race_EthnicityNA_Black_NA_NA_AmerInd_NA -29.92178    30.96066  -0.966
## Race_EthnicityNA_Black_NA_NA_AmerInd_OtherRace -46.86963    30.96226  -1.514
## Race_EthnicityNA_Black_NA_NA_NA_NA -32.46005    17.98396  -1.805
## Race_EthnicityNA_Black_NA_NA_NA_OtherRace -34.97393    25.27866  -1.384
## Race_EthnicityNA_NA_Hispanic_Asian_NA_NA  79.76534    25.28843   3.154
## Race_EthnicityNA_NA_Hispanic_NA_AmerInd_NA -40.13037    30.96226  -1.296
## Race_EthnicityNA_NA_Hispanic_NA_NA_NA -32.62660    18.03770  -1.809
## Race_EthnicityNA_NA_NA_Asian_NA_NA -31.56699    18.01273  -1.752
## Race_EthnicityNA_NA_NA_NA_AmerInd_NA -35.98525    19.20135  -1.874
## Race_EthnicityNA_NA_NA_NA_NA_OtherRace -42.78865    18.74897  -2.282
## Race_EthnicityWhite_Black_Hispanic_NA_AmerInd_NA -46.79141    25.28636  -1.850
## Race_EthnicityWhite_Black_Hispanic_NA_NA_OtherRace -43.13037    30.96226  -1.393
## Race_EthnicityWhite_Black_NA_NA_AmerInd_NA -42.10429    25.28050  -1.665
## Race_EthnicityWhite_Black_NA_NA_NA_NA -34.59777    18.60684  -1.859
## Race_EthnicityWhite_Black_NA_NA_NA_OtherRace -30.02607    30.95987  -0.970
## Race_EthnicityWhite_NA_Hispanic_Asian_NA_NA -35.54754    31.00370  -1.147
## Race_EthnicityWhite_NA_Hispanic_NA_AmerInd_NA -33.85736    21.15025  -1.601
## Race_EthnicityWhite_NA_Hispanic_NA_NA_NA -39.48094    18.29547  -2.158
## Race_EthnicityWhite_NA_NA_Asian_NA_NA -41.40224    18.79367  -2.203
```

```
## Race_EthnicityWhite_NA_NA_Asian_NA_OtherRace      -40.97393    30.95987   -1.323
## Race_EthnicityWhite_NA_NA_NA_AmerInd_NA           -37.27061    18.57588   -2.006
## Race_EthnicityWhite_NA_NA_NA_AmerInd_OtherRace     -37.65644    25.28294   -1.489
## Race_EthnicityWhite_NA_NA_NA_NA_NA               -36.34863    17.88489   -2.032
## Race_EthnicityWhite_NA_NA_NA_NA_OtherRace         -46.35941    23.07618   -2.009
## Vote_Sway                                           0.05215     0.15714    0.332
##                                                     Pr(>|t|)
## (Intercept)                                       0.00517 **
## Race_EthnicityNA_Black_Hispanic_NA_NA_NA          0.07035 .
## Race_EthnicityNA_Black_Hispanic_NA_NA_OtherRace   0.15375
## Race_EthnicityNA_Black_NA_Asian_NA_NA            0.33140
## Race_EthnicityNA_Black_NA_NA_AmerInd_NA          0.33392
## Race_EthnicityNA_Black_NA_NA_AmerInd_OtherRace    0.13022
## Race_EthnicityNA_Black_NA_NA_NA_NA              0.07121 .
## Race_EthnicityNA_Black_NA_NA_NA_OtherRace        0.16663
## Race_EthnicityNA_NA_Hispanic_Asian_NA_NA         0.00163 **
## Race_EthnicityNA_NA_Hispanic_NA_AmerInd_NA       0.19506
## Race_EthnicityNA_NA_Hispanic_NA_NA_NA            0.07061 .
## Race_EthnicityNA_NA_NA_Asian_NA_NA               0.07982 .
## Race_EthnicityNA_NA_NA_NA_AmerInd_NA             0.06104 .
## Race_EthnicityNA_NA_NA_NA_NA_OtherRace           0.02257 *
## Race_EthnicityWhite_Black_Hispanic_NA_AmerInd_NA 0.06437 .
## Race_EthnicityWhite_Black_Hispanic_NA_NA_OtherRace 0.16375
## Race_EthnicityWhite_Black_NA_NA_AmerInd_NA       0.09595 .
## Race_EthnicityWhite_Black_NA_NA_NA_NA            0.06309 .
## Race_EthnicityWhite_Black_NA_NA_NA_OtherRace     0.33223
## Race_EthnicityWhite_NA_Hispanic_Asian_NA_NA      0.25168
## Race_EthnicityWhite_NA_Hispanic_NA_AmerInd_NA    0.10955
## Race_EthnicityWhite_NA_Hispanic_NA_NA_NA         0.03103 *
## Race_EthnicityWhite_NA_NA_Asian_NA_NA            0.02769 *
## Race_EthnicityWhite_NA_NA_Asian_NA_OtherRace      0.18581
## Race_EthnicityWhite_NA_NA_NA_AmerInd_NA          0.04493 *
## Race_EthnicityWhite_NA_NA_NA_AmerInd_OtherRace    0.13651
## Race_EthnicityWhite_NA_NA_NA_NA_NA              0.04223 *
## Race_EthnicityWhite_NA_NA_NA_NA_OtherRace        0.04465 *
## Vote_Sway                                           0.74003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.28 on 2377 degrees of freedom
## Multiple R-squared:  0.02478,    Adjusted R-squared:  0.01329
## F-statistic: 2.157 on 28 and 2377 DF,  p-value: 0.0004115
```

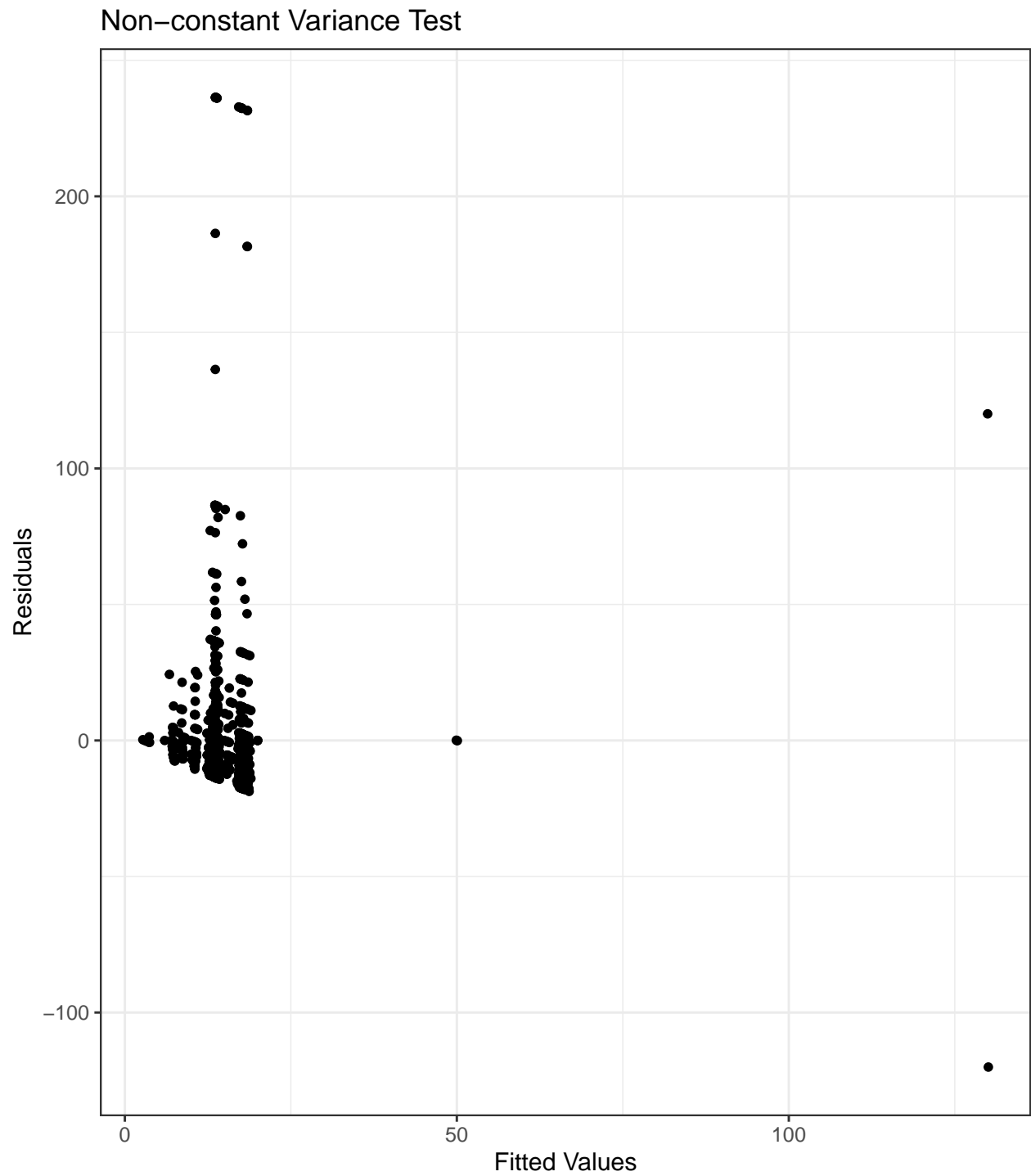
By Looking at the r squared value we see that the predictors are not at all manipulating our response so we already have a negative sign

```
s245::gf_acf(~slr)|>
gf_lims(y=c(-0.2,0.2))|>
gf_labs( title = "Independence Of Residuals", )
```



FAILED(two lines surpass the dashed areas meaning this test failed. Under lenient conditions, we could say this passed but for this I would still stick to failed!)

```
gf_point(resid(slr)~fitted(slr))|>  
gf_labs(x = 'Fitted Values', y = 'Residuals')|>  
gf_labs(title = "Non-constant Variance Test", )
```



FAILED

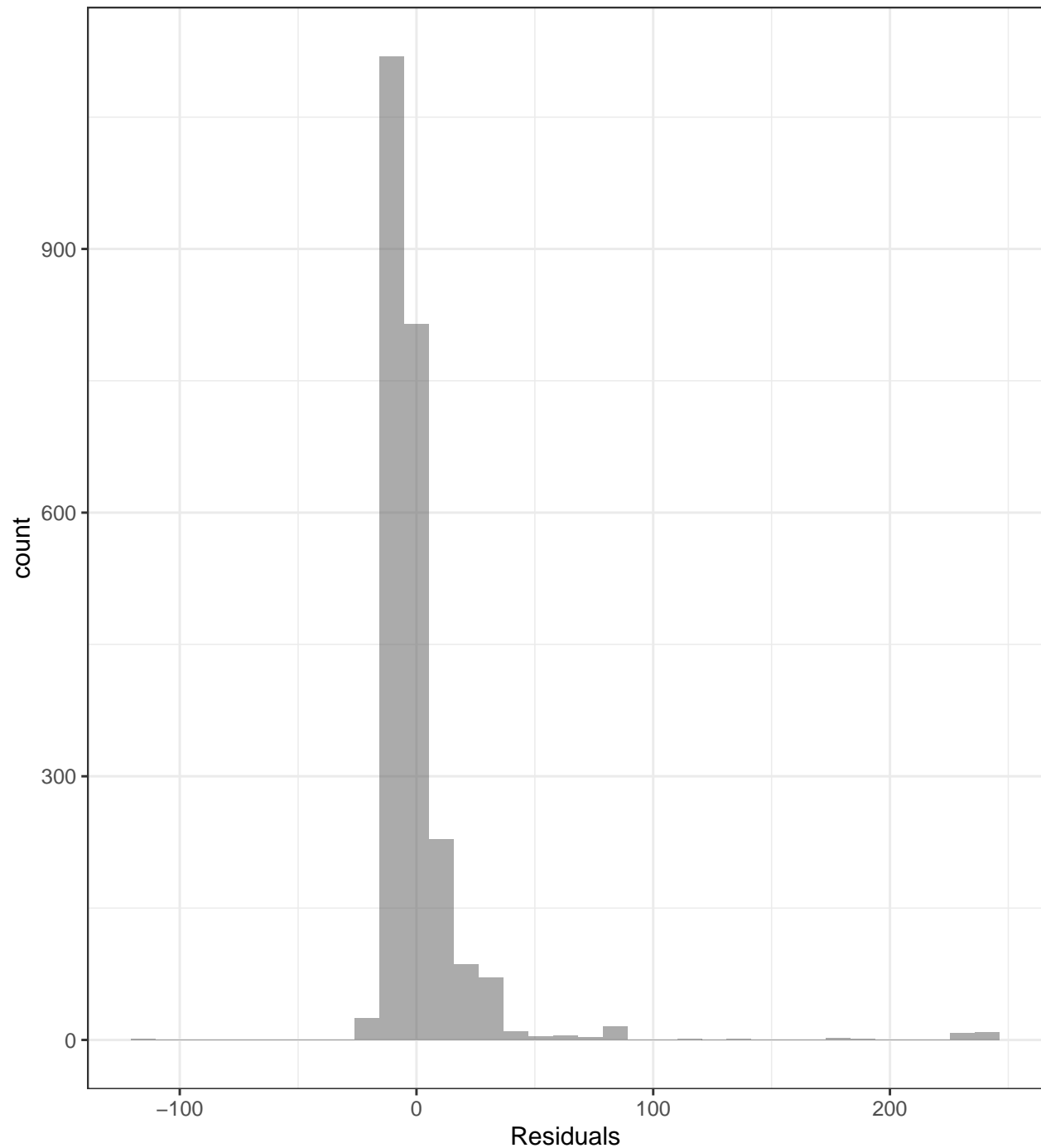
```
gf_point(log(Searches) ~ Race_Ethnicity,  
data = search)|>  
gf_labs(title = "Lack of non-linearity Test", )
```

The scatter plot displays a distribution of points on a 2D grid. The points are most densely packed in a vertical column on the right side of the plot, with a few scattered points elsewhere. The grid lines are light gray, and the plot is enclosed in a black frame.

```
search <- search |>
  mutate(preds = predict(slr),
         resids = resid(slr))

gf_histogram(~resids, data = search, bins = 35)|>
  gf_labs(title = "Histogram: Normality Of Residuals", x = "Residuals")
```

Histogram: Normality Of Residuals

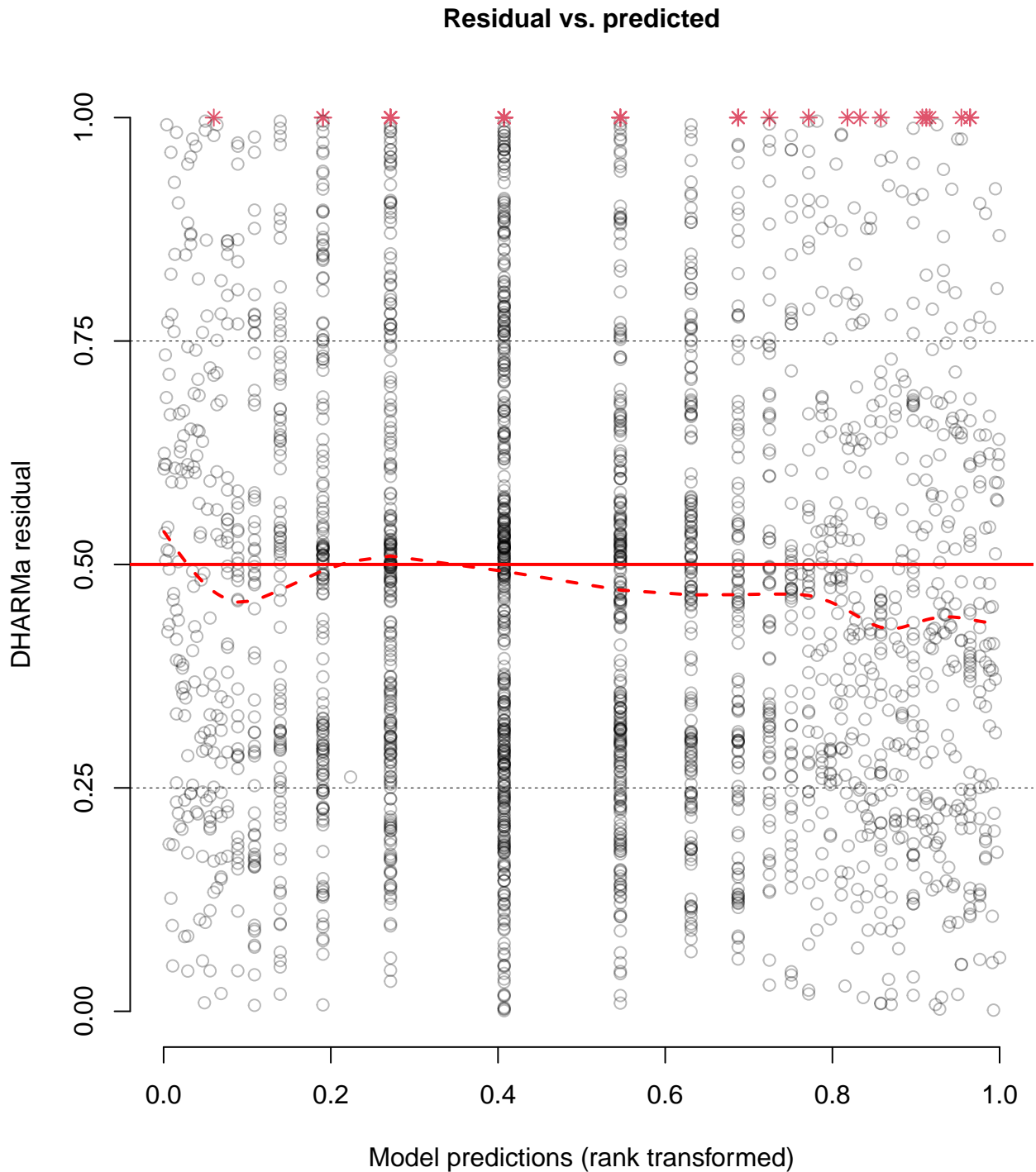


FAILED (there is a huge skew right which makes this histogram fail badly)

```
library(glmTMB)
search_nb2 <- glmTMB(Searches ~ Race_Ethnicity + Vote_Sway,
  data = search,
  family = nbinom2(link = 'log'))
```

```
library(DHARMA)
nb2_sim <- simulateResiduals(search_nb2)
```

```
plotResiduals(nb2_sim,
              quantreg = FALSE)
```



This Graph looks problematic because the dashed line is jumping from above and below the 0.50 point which is something to take into consideration!

Overall i would say the Assessments failed and that we would need to start from the beginning because we need different predictors and response variable.