SentimentRegressions

Aaron Dantzler

2023-07-26

afinn

```
data <- read.csv(file = "D:\\Princeton\\BSPL\\norms_sent_afinn.csv")
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
new_df <- select(data, prolific, control, treatment, frq_topic_t1,
                 frq_topic_t2, frq_topic_t3, combined_sentimentt1,
                 combined_sentimentt2, combined_sentimentt3)

new_df <- rename(new_df, sentiment_t1 = combined_sentimentt1, sentiment_t2 = combined_sentimentt2,
                 sentiment_t3 = combined_sentimentt3)
```

```
frq_topic <- c(new_df$frq_topic_t1, new_df$frq_topic_t2, new_df$frq_topic_t3)

stacked_df <- data.frame(frq_topic)
```

```
time1 <- c(1)
num_repetitions <- 616
time1 <- rep(time1, times = num_repetitions)

time2 <- c(2)
time2 <- rep(time2, times = num_repetitions)

time3 <- c(3)
time3 <- rep(time3, times = num_repetitions)
```

```r
time <- c(time1, time2, time3)

stacked_df$time <- time
```

```r
stacked_df$prolific <- c(new_df$prolific, new_df$prolific, new_df$prolific)
stacked_df$control <- c(new_df$control, new_df$control, new_df$control)
stacked_df$treatment <- c(new_df$treatment, new_df$treatment, new_df$treatment)

stacked_df$sentiment <- c(new_df$sentiment_t1, new_df$sentiment_t2,
                          new_df$sentiment_t3)
```

```r
stacked_df$treated <- ifelse(((data$control == "climate") &
                              (data$frq_topic_t1 == 1 | data$frq_topic_t1 == 2)),
                             0,
                       ifelse(((data$control == "health") &
                              (data$frq_topic_t1 == 4 | data$frq_topic_t1 == 5)),
                             0,
                        ifelse(((data$control == "politics") &
                              (data$frq_topic_t1 == 5 | data$frq_topic_t1 == 6)),
                             0, 1)))
```

```r
stacked_df$evidence <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "evidence"), 1, 0)

stacked_df$normevidence <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "normevidence"), 1, 0)

stacked_df$norm <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "norm"), 1, 0)
```

```r
model <- lm(sentiment ~ treated + treated:time + time, data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ treated + treated:time + time, data = stacked_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.1989  -3.5287   0.4713   3.7688  21.1946
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.13441    0.63063   1.799   0.0722 .
## treated      -1.88247    0.75480  -2.494   0.0127 *
## time          0.03226    0.29193   0.111   0.9120
## treated:time  0.24449    0.34941   0.700   0.4842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.63 on 1844 degrees of freedom
## Multiple R-squared:  0.01388,    Adjusted R-squared:  0.01227
## F-statistic: 8.649 on 3 and 1844 DF,  p-value: 1.063e-05
```

```
model <- lm(sentiment ~ norm + evidence + normevidence +
             normevidence:time + evidence:time + norm:time + time,
            data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ norm + evidence + normevidence + normevidence:time +
##     evidence:time + norm:time + time, data = stacked_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.1989  -3.2962   0.5025   3.7565  21.5025
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.13441    0.63067   1.799   0.0722 .
## norm              -1.28281    0.95103  -1.349   0.1775
## evidence          -2.24026    0.94742  -2.365   0.0182 *
## normevidence      -2.13686    0.97042  -2.202   0.0278 *
## time               0.03226    0.29194   0.110   0.9120
## normevidence:time  0.21774    0.44922   0.485   0.6279
## evidence:time      0.36977    0.43857   0.843   0.3993
## norm:time          0.14240    0.44024   0.323   0.7464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.631 on 1840 degrees of freedom
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.01216
## F-statistic: 4.248 on 7 and 1840 DF,  p-value: 0.0001154
```

```
model <- lm(sentiment ~ norm + evidence + normevidence,
            data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ norm + evidence + normevidence, data = stacked_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.1989  -3.4975   0.5025   3.7991  21.5025
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1989     0.2383   5.032 5.33e-07 ***
## norm          -0.9980     0.3593  -2.778  0.00553 **
```

```
## evidence      -1.5007     0.3579  -4.193 2.89e-05 ***
## normevidence  -1.7014     0.3666  -4.641 3.72e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.628 on 1844 degrees of freedom
## Multiple R-squared:  0.01465,    Adjusted R-squared:  0.01305
## F-statistic: 9.141 on 3 and 1844 DF,  p-value: 5.275e-06
```

bing

```
data <- read.csv(file = "D:\\Princeton\\BSPL\\norms_sent_bing.csv")
```

```
library(dplyr)

new_df <- select(data, prolific, control, treatment, frq_topic_t1,
                 frq_topic_t2, frq_topic_t3, sentimentt1,
                 sentimentt2, sentimentt3)

new_df <- rename(new_df, sentiment_t1 = sentimentt1, sentiment_t2 = sentimentt2,
                 sentiment_t3 = sentimentt3)
```

```
frq_topic <- c(new_df$frq_topic_t1, new_df$frq_topic_t2, new_df$frq_topic_t3)

stacked_df <- data.frame(frq_topic)
```

```
time1 <- c(1)
num_repetitions <- 616
time1 <- rep(time1, times = num_repetitions)

time2 <- c(2)
time2 <- rep(time2, times = num_repetitions)

time3 <- c(3)
time3 <- rep(time3, times = num_repetitions)

time <- c(time1, time2, time3)

stacked_df$time <- time
```

```
stacked_df$prolific <- c(new_df$prolific, new_df$prolific, new_df$prolific)
stacked_df$control <- c(new_df$control, new_df$control, new_df$control)
stacked_df$treatment <- c(new_df$treatment, new_df$treatment, new_df$treatment)

stacked_df$sentiment <- c(new_df$sentiment_t1, new_df$sentiment_t2,
                          new_df$sentiment_t3)
```

```
stacked_df$treated <- ifelse(((data$control == "climate") &
                              (data$frq_topic_t1 == 1 | data$frq_topic_t1 == 2)),
                             0,
```

```
                      ifelse(((data$control == "health") &
                              (data$frq_topic_t1 == 4 | data$frq_topic_t1 == 5)),
                             0,
                       ifelse(((data$control == "politics") &
                              (data$frq_topic_t1 == 5 | data$frq_topic_t1 == 6)),
                             0, 1)))
```

```
stacked_df$evidence <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "evidence"), 1, 0)

stacked_df$normevidence <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "normevidence"), 1, 0)

stacked_df$norm <- ifelse((stacked_df$treated == 1) &
                                (stacked_df$treatment == "norm"), 1, 0)
```

```
model <- lm(sentiment ~ treated + treated:time + time, data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ treated + treated:time + time, data = stacked_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.8226  -1.9977   0.1674   2.1747  12.0023
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.172043   0.371684  -0.463  0.64351
## treated      -1.160515   0.444867  -2.609  0.00916 **
## time         -0.002688   0.172056  -0.016  0.98754
## treated:time  0.167804   0.205933   0.815  0.41526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.319 on 1844 degrees of freedom
## Multiple R-squared:  0.01401,    Adjusted R-squared:  0.0124
## F-statistic: 8.733 on 3 and 1844 DF,  p-value: 9.444e-06
```

```
model <- lm(sentiment ~ norm + evidence + normevidence +
             normevidence:time + evidence:time + norm:time + time,
            data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ norm + evidence + normevidence + normevidence:time +
##     evidence:time + norm:time + time, data = stacked_df)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.8226  -2.0274   0.1385   2.1747  11.9726
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.172043   0.371860  -0.463  0.64367
## norm              -1.458094   0.560754  -2.600  0.00939 **
## evidence          -1.159038   0.558627  -2.075  0.03814 *
## normevidence      -0.842663   0.572187  -1.473  0.14100
## time              -0.002688   0.172138  -0.016  0.98754
## normevidence:time -0.037753   0.264871  -0.143  0.88667
## evidence:time      0.195256   0.258594   0.755  0.45031
## norm:time          0.331455   0.259578   1.277  0.20180
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.32 on 1840 degrees of freedom
## Multiple R-squared:  0.01522,    Adjusted R-squared:  0.01147
## F-statistic: 4.061 on 7 and 1840 DF,  p-value: 0.000198
```

```
model <- lm(sentiment ~ norm + evidence + normevidence,
            data = stacked_df)

summary(model)
```

```
##
## Call:
## lm(formula = sentiment ~ norm + evidence + normevidence, data = stacked_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.8226  -2.0274   0.0956   2.1774  11.9726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.1774     0.1405  -1.262 0.206980
## norm          -0.7952     0.2119  -3.752 0.000181 ***
## evidence      -0.7685     0.2111  -3.640 0.000280 ***
## normevidence  -0.9182     0.2163  -4.246 2.29e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.32 on 1844 degrees of freedom
## Multiple R-squared:  0.01313,    Adjusted R-squared:  0.01152
## F-statistic: 8.177 on 3 and 1844 DF,  p-value: 2.084e-05
```