Analysis Draft: 2018 r/feminism Study

Tyler Simko

10/23/2019

```r
library(MASS)
library(ggplot2)
reddit <- read.csv("../data/r-feminism-study-data-merged-10.17.2019.csv",
                   stringsAsFactors = FALSE)

colnames(reddit)
```

```
##  [1] "X"                  "assignment_datetime" "ban_actions"
##  [4] "ban_days_2_weeks"   "ban_days_4_weeks"    "ban_days_8_weeks"
##  [7] "ban_observation_days" "block_id"          "comment_id"
## [10] "comments"           "comments_2_weeks"    "comments_4_weeks"
## [13] "comments_8_weeks"   "community_closeness"  "completed_survey"
## [16] "id"                 "identify_feminist"   "message_status"
## [19] "randomization"      "submission_id"       "timestamp"
## [22] "treatment"          "username"
```

```r
table(reddit$treatment)
```

```
##
##   0   1
## 498 487
```

Overall participation

```r
comments.mod <- glm.nb(comments_2_weeks ~ treatment, data=reddit)
summary(comments.mod)$coefficients
```

```
##              Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -0.8875901  0.1637118 -5.421662 5.904745e-08
## treatment    0.3588082  0.2295748  1.562925 1.180702e-01
```

```r
comments.mod.4 <- glm.nb(comments_4_weeks ~ treatment, data=reddit)
summary(comments.mod.4)$coefficients
```

```
##              Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -0.5068176  0.1538040 -3.2952171 0.0009834567
## treatment    0.1415994  0.2177053  0.6504175 0.5154225599
```

```r
comments.mod.8 <- glm.nb(comments_8_weeks ~ treatment, data=reddit)
summary(comments.mod.8)$coefficients
```

```
##              Estimate Std. Error   z value  Pr(>|z|)
## (Intercept) -0.18473410  0.1485981 -1.2431796 0.2138017
## treatment    0.04147529  0.2110952  0.1964767 0.8442371
```

```r
## results
res <- data.frame(rbind(summary(comments.mod)$coefficients[2,],
                        summary(comments.mod.4)$coefficients[2,],
                        summary(comments.mod.8)$coefficients[2,]))
```

```
res$period <- c("2 weeks", "4 weeks", "8 weeks")
colnames(res) <- c("est", "se", "z", "p", "period")

ggplot(res, aes(x = period, y = est)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = est - 1.96*se, ymax = est + 1.96*se), width = 0.2) +
  geom_hline(yintercept = 0, lty = "dashed", color = "#A36D90") + theme_bw() +
  xlab("Time after Treatment") + ylab("ATE on # of Comments")
```
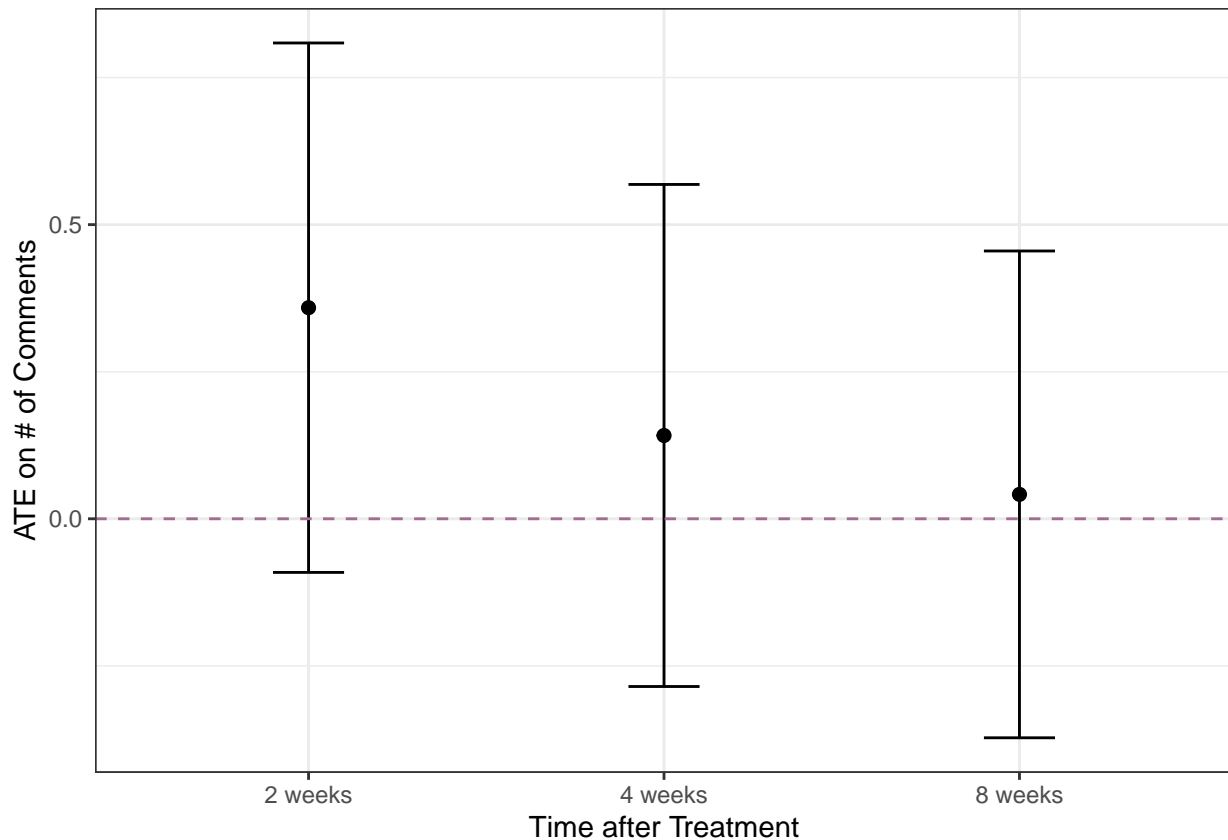

Survey measures

```
inclusion.mod <- lm(community_closeness ~ treatment,
                    data=subset(reddit, completed_survey=="True"))
summary(inclusion.mod)$coefficients

##              Estimate Std. Error    t value      Pr(>|t|)
## (Intercept) 3.9130435  0.1775355 22.0409017 2.145485e-57
## treatment   0.1821946  0.2569816  0.7089793 4.790948e-01

reddit$identify_feminist <- ifelse(reddit$identify_feminist == "Yes", 1,
                            ifelse(reddit$identify_feminist == "No", -1,
                                   ifelse(reddit$identify_feminist == "Undecided", 0, NA)))

identify.mod <- lm(identify_feminist ~ treatment,
                   data=subset(reddit, completed_survey=="True"))
```

```
summary(identify.mod)$coefficients
```

```
##              Estimate Std. Error  t value      Pr(>|t|)
## (Intercept) 0.4782609 0.07475158 6.398003 9.481650e-10
## treatment   0.1122153 0.10820245 1.037087 3.008444e-01
```