Lab 1: Comparing Means

w203: Statistics for Data Science

Contents

```r
anes <- read_csv("anes_pilot_2018.csv")
```

```r
anes <- anes %>%
  mutate(
    voted_2016 = case_when(
      turnout16 == 1 ~ 'Voted',
      turnout16 == 2 ~ 'Did not Vote',
      turnout16 == 3 & turnout16b == 1 ~ 'Voted',
      turnout16 == 3 & turnout16b == 2 ~ 'Did not Vote'),
    voted_2018 = case_when(
      turnout18 %in% 1:3 ~ 'Voted',
      turnout18 == 4 ~ 'Did not Vote',
      turnout18 == 5 & turnout18ns == 1 ~ 'Voted',
      turnout18 == 5 & turnout18ns == 2 ~ 'Did not Vote')
  )
```

```r
anes <- anes %>%
  filter(
    geangry  >= 0,
    geafraid >= 0,
    !is.na(voted_2016),
    !is.na(voted_2018)
  )
```

1 Question 4: Was anger or fear more effective at driving increases in voter turnout from 2016 to 2018?

The 2016, 2018 (and now 2020) elections have been remarkable for their acrimony. Language has changed from referring to the electoral contests in terms of "horse races" to instead be in terms of "battles", "wars" and "conflict". In the Senate, once referred to as a bastion of collegiality, Senators have removed the filibuster – referred to rather hyperbolically as using the "nuclear option" – have refused to extend senatorial courtesy on judicial appointments. And, in the 2020 election cycle appointed a justice to the supreme court only weeks before the presidential general election.

Needless to say, fear of the future and anger about the present have play a key part in parties' attempts to turn out voters. But, is one more effective at bringing individuals to the polls?

Table 1: Cross Tab of Self Reported Voting in 2016 and 2018

	Did not Vote	Voted
Did not Vote	0.20	0.05
Voted	0.05	0.70

One of the challenges of this question is that there are many ways that one could attempt to answer, given this data, but none are really satisfactory. In its ideal form, answering this question would require us to actually conduct an experiment that might take the following form:

> Among voters who did not turn out in the 2016 election, randomly assign individuals to receive either fear or anger inducing messaging.

Here is why. Even after a complete analysis of this data, it will not be possible to know why people chose *not* to vote in the 2016 election. Perhaps someone chose not to vote in the 2016 election because they were angry. Or, perhaps a citizen chose *to* vote because of their anger in the 2016 election.

There is another *major* confounder in this question – rates of turnout in midterm general elections are typically much lower than turnout in presidential general elections. And so, relatively high (or low) levels of turnout could be generated either by the fear, anger, or general interest in participating in an election for the office of the president. This confound becomes particularly acute because the 45th president pervasibely used both fear and anger inducing language in their campaign. As a result, there is a tight coupling of language and office that makes this an **extraordinarily** difficult question to answer.

Further complicating this issue is the measurement strategy that assesses whether individuals voted in 2016. The ANES here asks individuals to retrospectively evaluate whether they voted. This task is both difficult, and subject to an acute "social desirability bias" – there is a strong norm that voting is a civic duty and so individuals may over-report past voting.

A final limitation is the amount of data available to use in this form of test. After filtering, the test will have only 616. Nevertheless, we soldier on!

As we report in Table 1, 70% of ANES respondents report that they voted in both the 2016 and 2018 elections. While turnout of 75% might be expected in the presidential general, it is highly unlikely to have turnout this high in an off-cycle election. Also notable in this data is that voting (or not voting) seems to be highly durable – only 10% of the respondents report taking a different action in 2016 compared to 2018.

```
summary_table <- anes %$%
  prop.table(table('Voted in 2016' = voted_2016, 'Voted in 2018' = voted_2018))
```

There is a strong positive relationship between survey respondent's answer to questions and fear and anger. In 1 I plot, the feelings of fear and anger among those who did not vote in 2016. This data is further broken out by whether the individual voted in 2018. Clearly, there is a strong positive relationship between these measures. Notably, there seems to be little difference between the relationship whether or not someone voted in 2018.

```
anes %>%
  filter(voted_2016 == 'Did not Vote') %>%
  ggplot(aes(x = geangry, y = geafraid, color = voted_2018)) +
  geom_jitter() + stat_smooth(method = 'lm') +
  labs(
    x = 'Anger', y = 'Fear', color = 'Voted in 2018?',
    title = '2016 Non-Voters Emotions in 2018'
  )
```
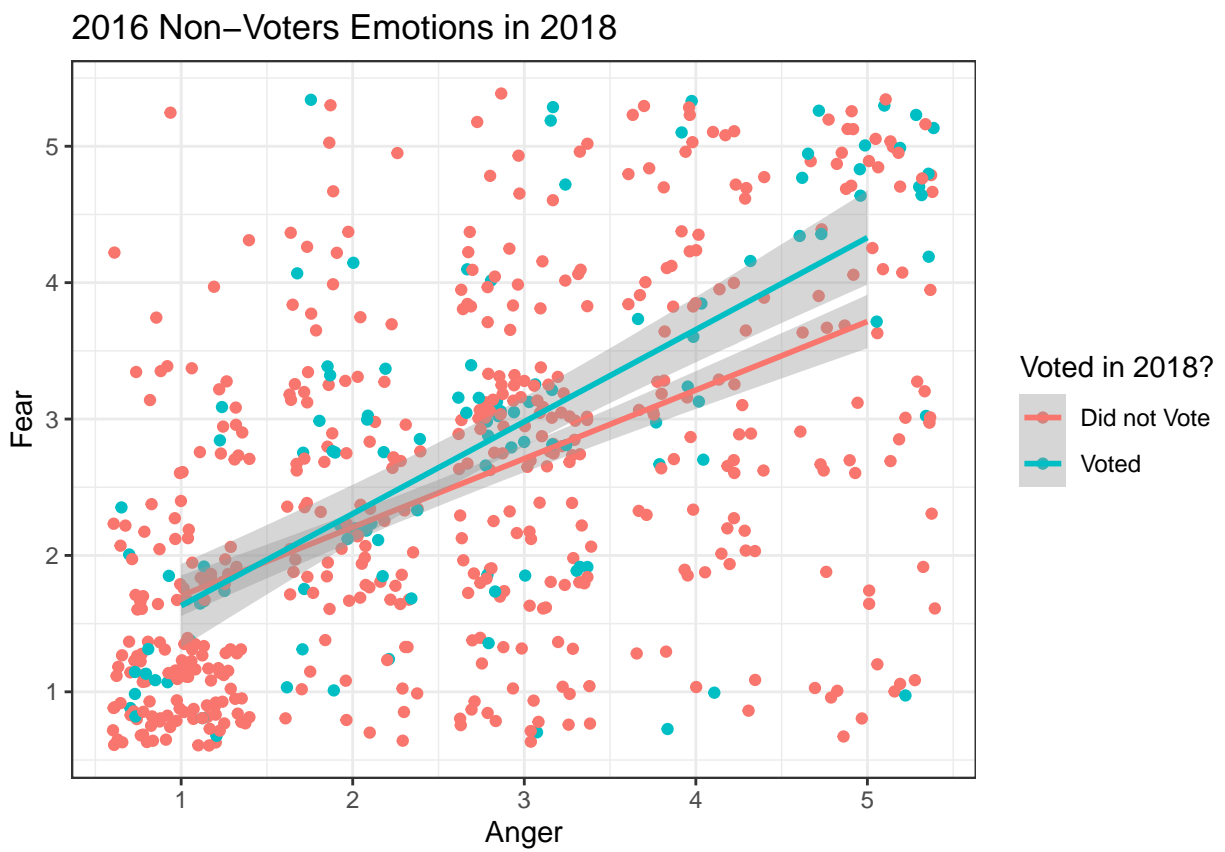
Figure 1: Voter Emotions

The first thing to note in this data is that there are differences in the rank ordering of fear and anger. The next figure plots the relative prevalence of ANES respondent's feelings of fear and anger. While there are about the same proportion of respondents who report being the lowest category in both fear and anger, there are more who report being somewhat afraid than report being somewhat angry. This is offset by fewer individuals reporting high levels of fear.

```r
anes %>%
  filter(voted_2016 == 'Did not Vote') %>%
  select(geangry, geafraid) %>%
  pivot_longer(cols = c(geangry, geafraid)) %>%
  mutate(
    name = case_when(
      name == 'geafraid' ~ 'Fear',
      name == 'geangry' ~ 'Anger')) %>%
  ggplot(aes(x = value, color = name)) +
  geom_density() +
  labs(
    x = "Respondent's Feelings",
    y = 'Density',
    color = 'Fear or Anger'
  )
```
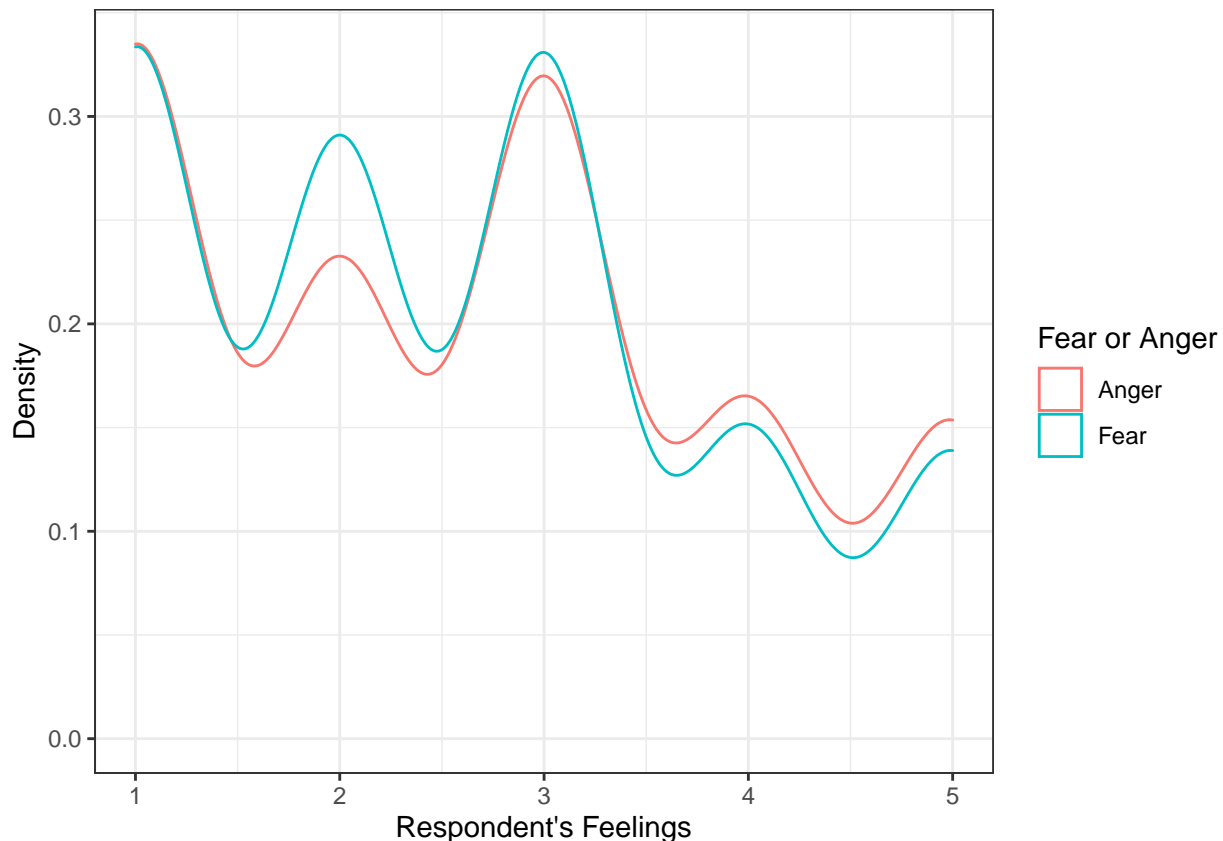


```r
wilcox_test <- anes %>%
  filter(voted_2016 == 'Did not Vote') %$%
  wilcox.test(x = geangry, y = geafraid, paired = TRUE)
```

Despite this pattern, a paired test suggests that these differences could be a product of sampling, with a

p-value of 0.26.

To test whether fear or anger was more effective at motivating people to vote in 2018 I conduct the following test:

- I limit the data to individuals who did not vote in 2016;
- I compute the difference between individual's reported anger and fear
- Then, I test whether this anger/fear differential is measurably different between 2018 voters and non-voters.

Although this feels like the most appropriate test I can construct here, there are several points where the ways the data are measured generate challenges for use.

First, data is reported in a 5 point, ordinal scale. There is no reason to believe that respondents are using the same scale when they think about fear and anger. Another way of noting this is to ask whether voters hold the same anchoring points on these scales. Second, because these are ordinal scales, there are real problems with computing a difference score between two ordinal variables. Third, this data is only to voters who did not vote in the 2016 election; but, the data does not possess any solid information about why they did not vote in 2016.

If this test were to **reject the null hypothesis** I would conclude that among those voters who did not vote in 2016, those who vote in 2018 have a measurably different anger/fear differential compared to those who did not vote in 2018. If the test were to **fail to reject the null hypothesis** then I would conclude that either there is not enough data, the hypothesized effect does not exist, or the test was inappropriate to conduct against data collected in this manor.

```
anes %>%
  filter(voted_2016 == 'Did not Vote') %>%
  mutate(anger_fear_difference = geangry - geafraid) %$%
  t.test(anger_fear_difference  ~ voted_2018)
```

```
##
##  Welch Two Sample t-test
##
## data:  anger_fear_difference by voted_2018
## t = 1.0912, df = 195.49, p-value = 0.2765
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1008062  0.3505291
## sample estimates:
## mean in group Did not Vote        mean in group Voted
##                 0.08946322                -0.03539823
```

This test provides no evidence of a fear/anger differential between voters and non-voters in the 2018 election.