

# ‘Lab 1: Question 1’

w203: Statistics for Data Science

## Contents

<b>1</b>	<b>Are survey respondents who have had someone in their home infected by COVID-19 more likely to disapprove of the way their governor is handling the pandemic?</b>	<b>6</b>
1.1	Description of Data . . . . .	6
1.2	Importance and Context . . . . .	6
1.3	Most appropriate test . . . . .	6
1.4	Test Limitations . . . . .	7

```
df%>%
  group_by(V201624, V201145, V201146)%>%
  summarise(Counts=n())
```

```
## # A tibble: 16 x 4
## # Groups:   V201624, V201145 [8]
##   V201624 V201145 V201146 Counts
##   <dbl>   <dbl>   <dbl> <int>
## 1      0      0      0    1780
## 2      0      0      1    3087
## 3      0      0     NA      9
## 4      0      1      0     843
## 5      0      1      1    2134
## 6      0      1     NA      1
## 7      0     NA     NA     41
## 8      1      0      0     64
## 9      1      0      1     95
## 10     1      1      0     40
## 11     1      1      1     85
## 12     1     NA     NA      3
## 13     NA      0      0     18
## 14     NA      0      1     30
## 15     NA      1      0     10
## 16     NA      1      1     32
```

```
0 0 0 1780
0 0 1 3087
0 1 0 843 0 1 1 2134
1 0 0 64
1 0 1 95
1 1 0 40
1 1 1 85
```

```
library("gplots")
```

```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##   lowess
```

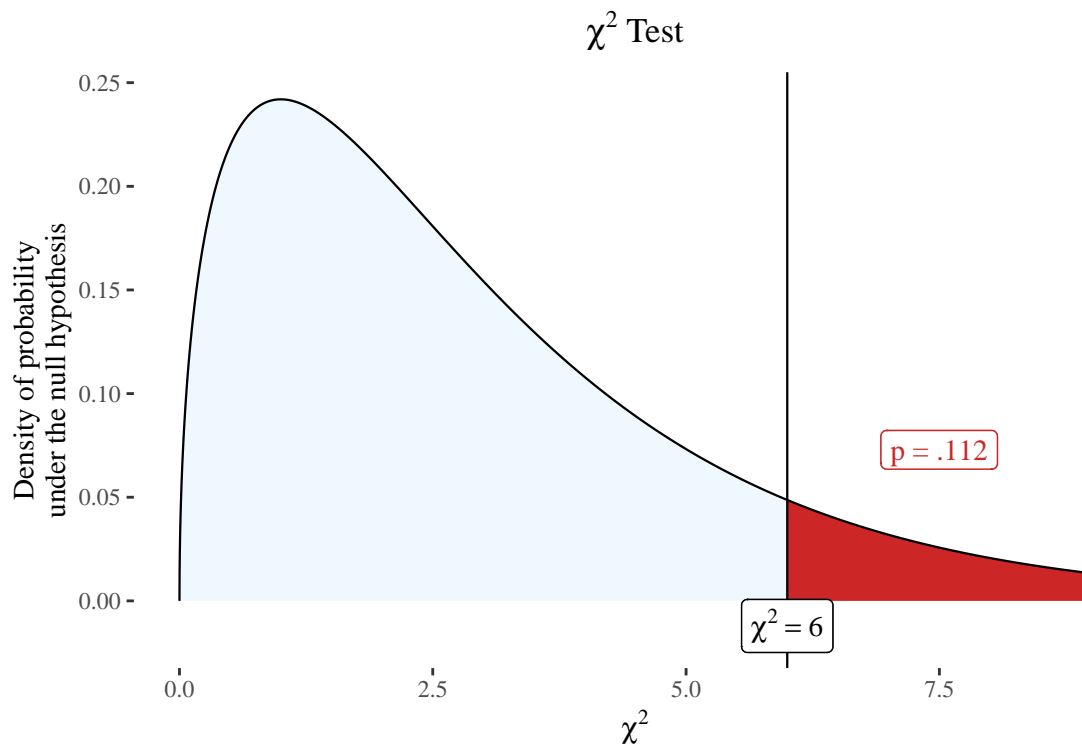
```
testor01 = rbind(c(1780,3087,843,2134),+ c(64,95,40,85)); testor01
```

```
##      [,1] [,2] [,3] [,4]
## [1,] 1780 3087 843 2134
## [2,]  64  95  40  85
```

```
chi01 = chisq.test(testor01,correct=F); chi01
```

```
##
## Pearson's Chi-squared test
##
## data:  testor01
## X-squared = 5.9986, df = 3, p-value = 0.1117
```

```
plotchisqtest(chi01)
```



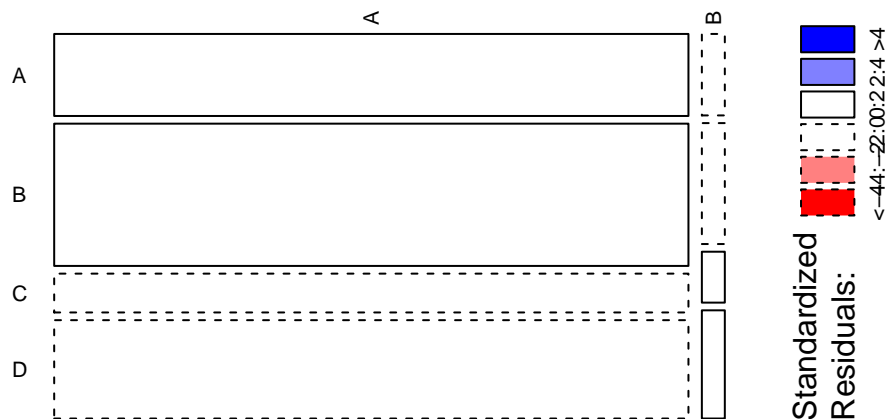
```
chi01$expected
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1779.5689 3070.8179 852.14715 2141.46604
## [2,]   64.4311  111.1821  30.85285   77.53396
```

```
#testor01 = rbind(c(1780,3087),+ c(64,95)); testor01
#chi01 = chisq.test(testor01,correct=F); chi01
#chi01$expected
```

```
dt <- as.table(as.matrix(testor01))
```

## housetasks



```
## NULL
```

```
testor02 = rbind(c(159,4876),+ c(125,2978)); testor02
```

```
##      [,1] [,2]
## [1,] 159 4876
## [2,] 125 2978
```

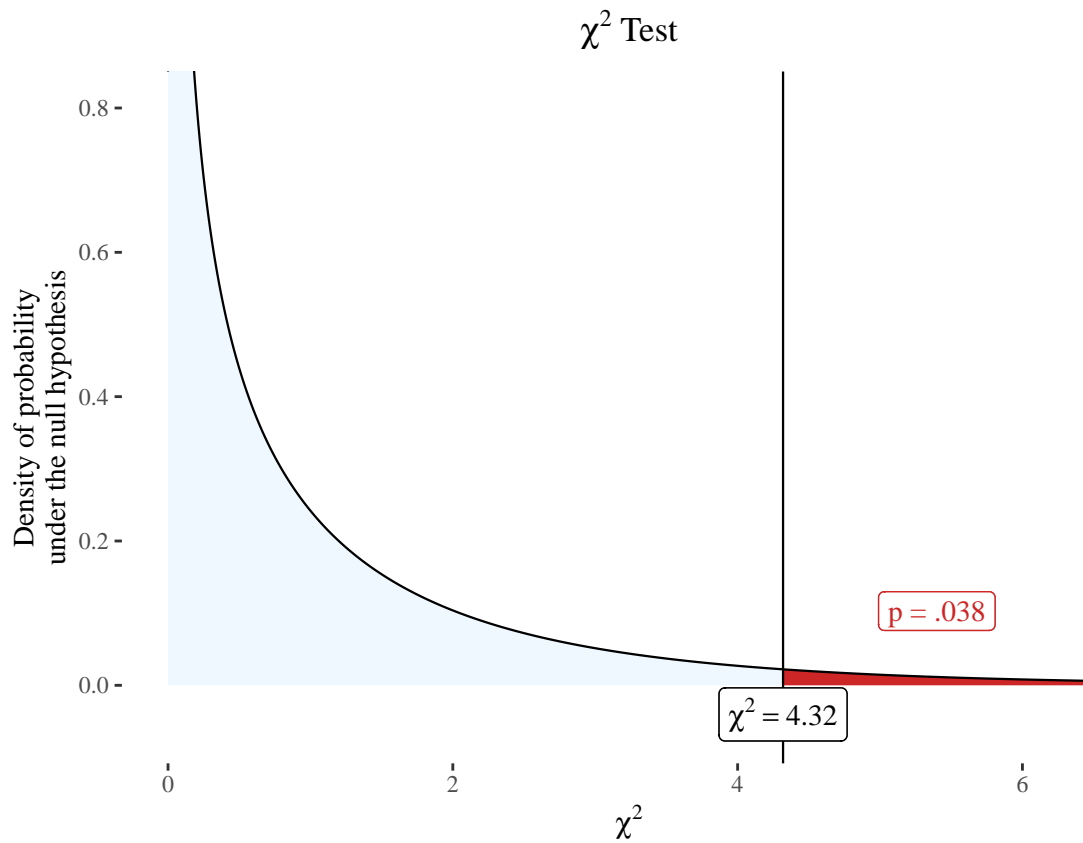
```
chi02 = chisq.test(testor02,correct=F); chi02
```

```
##
## Pearson's Chi-squared test
##
## data: testor02
## X-squared = 4.3191, df = 1, p-value = 0.03769
```

```
chi02$expected
```

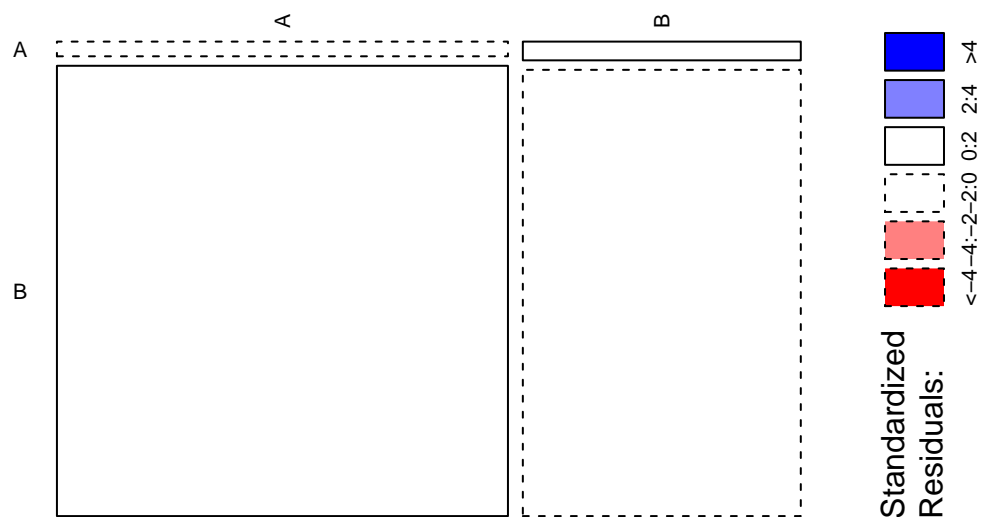
```
##      [,1]      [,2]
## [1,] 175.7115 4859.289
## [2,] 108.2885 2994.711
```

```
plotchisqtest(chi02)
```



```
dt02 <- as.table(as.matrix(testor02))
```

## housetasks



```
## NULL
```

# 1 Are survey respondents who have had someone in their home infected by COVID-19 more likely to disapprove of the way their governor is handling the pandemic?

## 1.1 Description of Data

We will address this question using data from the 2018 American National Election Studies (ANES). This is an observational dataset, based on a sample of respondents drawn from the YouGov platform. As we report in Table ??, 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. To examine factors behind voter turnout, I limit the data to individuals who did not vote in 2016, then switched to voting in 2018. This group most closely represents a drive in turnout, and may hold the best clues about changing voter behavior. After subsetting, we have observations.

## 1.2 Importance and Context

According to a Fox News report on February 2 local time, the approval rating of California Governor Gavin Newsom has plummeted, and this situation may exacerbate the possibility of his removal. A poll conducted by the Berkeley Institute of Government on more than 10,000 registered voters in California found that Newsom’s work performance was only approved by 46%, a sharp drop from his 64% approval rate in September last year. In addition, Newsom has been widely criticized for the previous strict stay-at-home order. Up to now, 1.3 million people have signed a petition calling for his removal as governor. The report pointed out that 36% of California voters currently support the removal of Newsom, and most of them support the Republican Party. They are dissatisfied with Newsom’s performance in handling the epidemic. In addition, less than one-third of the interviewees said that the governor had “excelled” in responding to the epidemic. In recent weeks, the number of new confirmed cases of new coronary pneumonia in a single day in California has declined. According to statistics from the state, only 12,000 new confirmed cases were reported on the 2nd, while in December last year, the highest number of new confirmed cases in a single day exceeded 50,000. Was the infection of Covid-19 impacts people approve or disapprove their governor handling Covid-19? The answer to this question could provide guidance to future political campaigns hoping to increase voter turnout. It could also provide useful background for governments that are interested civic participation. A better understanding of the factors that polarize society, and how they express themselves through the voting process, may also help those hoping to counteract that polarization. Since the beginning of the pandemic, and the resulting economic slump, Governors have been directing their states’ resources where they are needed most and where they will achieve the greatest impact. As practical problem solvers, Governors are effectively delivering state services to protect public health and to stimulate economic recovery and growth.

## 1.3 Most appropriate test

I then test whether this subgroup reports being more angry or more afraid. Because these variables are measured on ordinal scales, a non-parametric test is appropriate. Furthermore, the data is paired, since each individual has a pair of measurements. Some researchers may use a signed-rank test in these circumstances, but that test, while technically non-parametric, requires data on a metric scale. Instead, I will use a sign test, implemented in R using `binom.test`.

The sign test requires the following assumptions to be true:

- i.i.d. data. The ANES 2018 pilot uses a panel of individuals that use YouGov. This is an online system that rewards individuals for filling out surveys. There is a possibility that this introduces dependencies.

For example, participants may tell friends or family members about YouGov, resulting in a cluster of individuals that give similar responses. Nevertheless, YouGov claims to have millions of users, which suggests that links between individuals should be rare.

- Ordinal scale. The data levels showing an increase in intensity from “not at all” to “extremely.” There is some question, however, as to whether participants use the same scale when they think about anger and fear. Another way of noting this is to ask whether voters hold the same anchoring points on these scales.

## 1.4 Test Limitations

We have conducted this test based on the data available in the ANES. However, even after an appropriate analysis of the data on hand, it would not be possible know why people chose to **not** vote in the 2016 election. Perhaps someone chose not to vote in the 2016 election because they were angry. If this were to be the case, then our study design cannot find an effect of anger-inducing language in the 2018 data. And so, while the data that test on does fails to reject the null hypothesis that there is a difference in anger and fear as a motivation for increased voting in 2018, it is possible that the very nature of this observational data has precluded our ability to produce a satisfactory answer to the question.