

Agenda

1. Project groups due today
2. Randomization test recap
3. Hypothesis testing

Randomization test recap In the last class, we designed a simulation to help us answer the question of whether exposure to mites was associated, *to a statistically significant* degree, with a decrease in wilt disease after exposure to *Verticillium*, a fungus that causes wilt disease.

Most of us used the ‘no-wilt-mites’ cell in the table as our test statistic. But there are other statistics we might have been interested in. Particularly, we might be interested in the difference in the proportion of plants that wilted between the mites and the no mites groups. To answer this question, we might

1. Count out 19 black cards (no wilt) and 28 red cards (wilt)
2. Shuffle
3. Deal into two piles: 26 (mites) and 21 (no mites)
4. Calculate the proportions of wilt (red cards) for each pile, then finally, the difference in proportions. Record your simulation and do steps 1-4 a bunch more times!

Questions

1. What is the *null hypothesis* for this simulation?
2. What is the *test statistic*?
3. Where does the test statistic lie in the *null distribution*?
4. Does this evidence cause you to *reject* or *fail to reject* the null hypothesis?
5. Write *one* sentence summarizing what you’ve learned about mites and wilt disease.

```
require(mosaic)
tally(~ outcome + treatment, data = Mites)

##           treatment
## outcome  mites no mites
## no wilt   15     4
## wilt     11    17

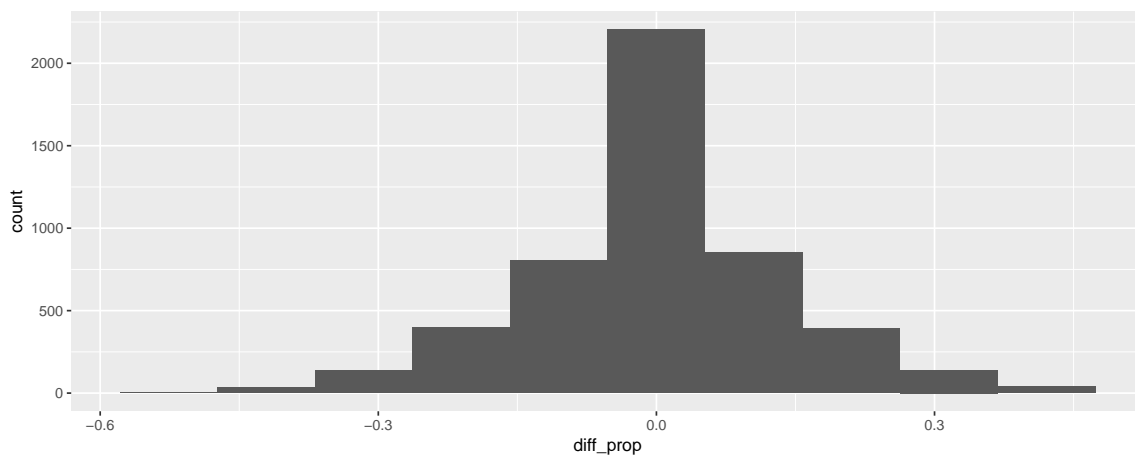
tally(outcome ~ treatment, data = Mites, format = "proportion")

##           treatment
## outcome      mites  no mites
## no wilt 0.5769231 0.1904762
## wilt   0.4230769 0.8095238
```

```
tbl <- tally(outcome ~ treatment, data = Mites, format = "proportion")
diff_prop <- tbl[2,2] - tbl[2,1]
diff_prop

## [1] 0.3864469
```

```
null_dist <- do(5000) * tally(outcome ~ shuffle(treatment), data = Mites)
null_dist <- null_dist %>%
mutate(prop_wilt_nomites = wilt.no.mites/(wilt.no.mites+no.wilt.no.mites)) %>%
mutate(prop_wilt_mites = wilt.mites/(wilt.mites+no.wilt.mites)) %>%
mutate(diff_prop = prop_wilt_nomites - prop_wilt_mites)
ggplot(data = null_dist, aes(diff_prop)) +
geom_histogram(bins = 10)
```



```
2 * pdata(~diff_prop, q=.3864, data=null_dist, lower.tail=FALSE)

## [1] 0.0156
```

What's Wrong? Here are several situations where there is an incorrect application of the ideas presented in this section. Write a short paragraph explaining what is wrong in each situation and why it is wrong.

1. A researcher tests the following null hypothesis: $H_0 : \bar{x} = 23$
2. A study with $\bar{x} = 45$ reports statistical significance for $H_a : \mu > 50$.
3. A researcher tests the hypothesis $H_0 : \mu = 350$ and concludes that the population mean is equal to 350.
4. A test preparation company wants to test that the average score of their students on the ACT is better than the national average score of 21.1. They state their null hypothesis to be $H_0 : \mu > 21.2$.
5. A study summary says that the results are statistically significant and the p-value is 0.98.