Agenda

1. Project groups/questions
2. Randomization test
3. Hypothesis Testing

Inference through Randomization Researchers suspect that the attack on a plant by one organism induces resistance to subsequent attacks by a different organism. Individually potted cotton plants were randomly allocated to two groups: one group that was infested by spider mites; and another group that was not. After two weeks the mites were dutifully removed by a conscientious research assistant, and both groups were inoculated with Verticillium, a fungus that causes wilt disease. The following table shows the number of plants that developed symptoms of wilt disease.

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

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

Big question: Is there a relationship between infestation and wilt disease? Said another way, is the proportion of plants that get wilt disease smaller after being exposed to mites compared to no mite exposure?

Activity: Simulation Design In groups of two, your objective is to answer the big question using a simulation with cards. Here are the steps we outlined.

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!

Results Record your simulated statistics (do it 5 to 10 times!) below, first as a table, then as a dot plot.

1. Describe the center, shape, and spread of this randomization distribution.

2. In roughly what range would you expect to generate statistics under this model? What range would be very unusual?

3. Where does the observed statistic fall in this distribution? What conclusions do you draw concerning the association of mites and wilt disease?

```
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)
```

Hypothesis Testing Our goal for this randomization simulation was to assess the likelihood that 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.

1. What was the *null hypothesis* for your simulation?

2. What was the *test statistic*?

3. Where did the test statistic lie in the *null distribution*?

4. Did this evidence cause you to *reject* or *fail to reject* the null hypothesis?

5. Write *one* sentence to your grandpa summarizing what you've learned about mites and wilt disease through hypothesis testing.

```
library(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
```

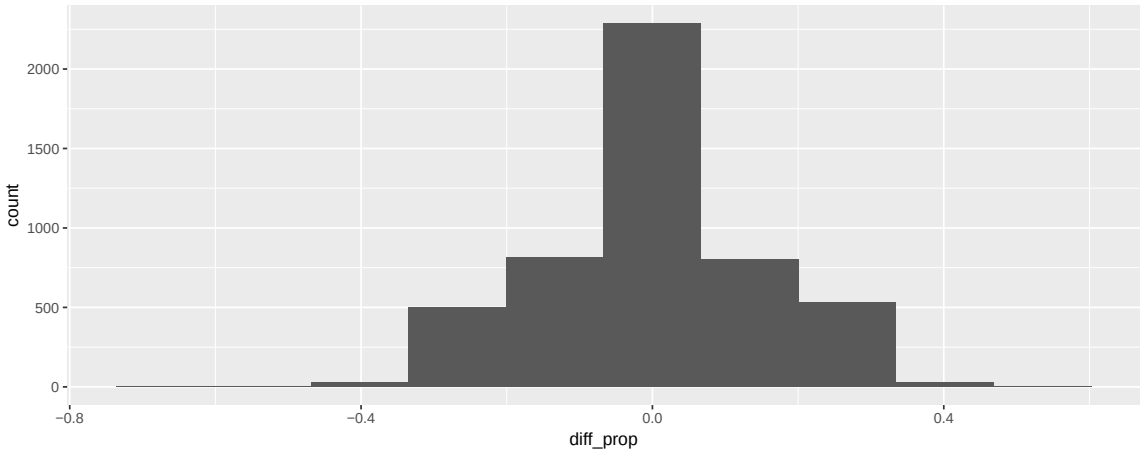
```
tbl <- tally(outcome ~ treatment, data = Mites, format = "proportion")
diff_prop <- tbl[2,2] - tbl[2,1]
diff_prop
```

```
## [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)
```




```
qdata(~diff_prop, p = c(0.025, 0.975), data = null_dist)
```

```
##         quantile     p
## 2.5%  -0.3021978 0.025
## 97.5%  0.3003663 0.975
```

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

```
## [1] 0.0124
```