Agenda

1. Inference through Randomization

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.

```
require(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?

Activity: Simulation Design   Your objective is to answer the big question using a simulation of your own design. You may want to refer back to the yawning activity that we did on the first day of class. You're welcome to use the cards in whatever manner is most useful. *Outline your procedure in detail* below, and be sure to touch on the following questions.

1. What does each card represent?

2. Upon what assumptions does your simulation rely?

3. What role does randomization play in the simulation?

4. What statistic will you be recording from each simulation?

Results   Record your simulated statistics 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?

Randomization test summary   Our goal here is 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 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
```

```
null_dist <- do(5000) * tally(~ outcome + shuffle(treatment), data = Mites)
ds <- tidyr::gather(null_dist, key = type, value = N)
qplot(data = ds, x = N, facets = ~type, binwidth = 1)
2 * pdata(~N, q = 11, data = filter(ds, type == "wilt.mites"))
2 * pdata(~N, q = 17, data = filter(ds, type == "wilt.no.mites"), lower.tail = FALSE)
2 * pdata(~N, q = 15, data = filter(ds, type == "no.wilt.mites"), lower.tail = FALSE)
2 * pdata(~N, q = 4, data = filter(ds, type == "no.wilt.no.mites"))
```