

# Are two proportions the same?

Stat 145

## Handedness and Foot size

Most people have one foot that is at least a little larger than the other. Let's use the `KidsFeet` dataset to see if there is an association between dominant hand and larger foot. Here is a table summarizing these two variables.

```
tally(biggerfoot ~ domhand, data = KidsFeet)
```

```
##           domhand
## biggerfoot L  R
##           L  2 20
##           R  6 11
```

1. First compute a few numbers “by hand” (you do the arithmetic in R or on a calculator).
  - a. Compute the proportion of left-handers whose left foot is the bigger foot.
  - b. Compute the proportion of right-handers whose left foot is the bigger foot.
  - c. Compute the difference in these proportions
2. Now get R to compute those same numbers for you using functions like `tally()`, `props()`, or `diffprop()`.
3. All of the kids in this data set come from the same grade at the same school. Is that a random sample from a population? How generalizable do you think these results will be?
4. You should have found a difference in proportions of just under 40%. That seems pretty big. Maybe there is an association between dominant hand and bigger foot in the population. Or maybe not. What is the other possible explanation for the difference we observed in our data?
5. Write down the **null and alternative hypotheses** for this situation.
6. Explain how you could use labeled cards to generate the null distribution for the difference in proportions when there is no association between dominant hand and larger foot.
7. Let's create a simulation in R to see whether that alternative explanation can be supported. This situation is very similar to the malaria vaccine study, so you can follow the outline there. Generate the difference in proportions for 1000 or 2000 random simulations where there is no association between dominant hand and larger foot. Build this up step by step.
  - a. Compute the observed difference in proportions in the **original data**. (Use `diffprop()`.)
  - b. Compute the difference in proportions for **one random sample**. (Use `shuffle()`.)
  - c. Use `do()` to **do this 5 times**. You should see 5 differences in proportions.
  - d. Add `set.seed()` to the top of your R chunk so you get the same random results each time you run the chunk. Put your favorite number in the parentheses.
  - e. Now **do it 1000 or 2000 times**, storing the results as `Kids_null` to indicate that this is the null distribution.
  - f. Use `Kids_null` to **make a histogram** of the null distribution for the difference in proportions under the assumption that there isn't an association between dominant hand and bigger foot.

- g. What **proportion of the null distribution** has a difference of proportions at least as large as the observed difference in proportions?
- h. **What should we conclude** based on all this?

These same steps can be used in many situations where we want to see if our data provide enough evidence to conclude that two proportions are not equal in the population.