

Lab 02

Download the `lab02.Rmd` file and open it using RStudio. Then, use the R programming language to help you answer the questions below. **Don't forget to fill out the worksheet form before the next class!**

1. Describe an experiment in a field of interest that can be described by a 2x2 contingency table. What are the two categorical inputs and what are the two categorical outputs? Provide a guess of what the two probabilities are and write them down using notation similar to the notes.

2. In pairs, describe an experiment that can be described by a 2x2 contingency table and can be completed without leaving the room. Some examples include:

- Have one person quiz the other on naming country capitals from two continents. Are they better at, say, European capitals compared to African capitals?
- Using two news aggregation sites (such as hackernews and reddit), open the first page of results. Record how many sites have a photograph that you can see without scrolling. Does the proportion of sites with photographs differ by site?
- Look at the first 50 pages of two books and record whether a page ends or breaks across a sentence boundary. Does this proportion differ between the two books.

It does not need to be something amazingly important. Just an idea to practice with.

3. Actually run the experiment and record your results in a contingency table. Compute the two observed probabilities \hat{p}_A and \hat{p}_B and their difference \hat{D} .

4. Using R, copy the code from the notes, simulate what differences you would get if there was not effect between the two categories. How extreme is the value you observed as a result of the simulation.

5. Modify the simulation so that the number of point collected is increased by a factor of 10. How extreme would the difference \hat{D} that

you observed be if the number of points you collected was larger? How does this compare to your answer to the previous question? Does the change make sense to you?

6. Describe why the plot of simulated values are (a) centered around zero, and (b) relatively symmetric around zero.