Objectives: This activity explores *hypothesis testing*, a statistical framework for comparing the likelihood of observing a particular outcome among several theoretical models.

Assignment You will be arranged into groups of 3 or 4. **Work together** to answer the following questions. Write your answers on class's shared google sheet (link will be posted in your Slack classroom channel) and add your data to the data page on class's shared google sheet

Is Yawning Contagious?

If you see someone else yawn, are you more likely to yawn? In an episode of the US show *Mythbusters*, the hosts conducted an experiment to answer this question.

Fifty adult participants who thought they were being considered for an appearance on the show were interviewed by a show recruiter. In the interview, the recruiter either yawned or did not. Participants then sat by themselves in a large van and were asked to wait. While in the van, the Mythbusters team watched the participants using a hidden camera to see if they yawned.

Of the 34 `seed` individuals exposed to a yawn, 10 of them later `yawned` in the van. And of the remaining 16 `control` individuals, 4 later `yawned` in the van. The results are summarized in the following table:

	`yawned`		
`group`	No	Yes	Total
Control	12	4	16
Seed	24	10	34
Total	36	14	50

Preliminary Questions

1. Is this an observational study or experiment?
2. Identity and classify (categorical/quantitative) the explanatory and response variables in this investigation.
3. Calculate the following:
   (a) The proportion $\hat{p}_1$ of yawning within the seeded group.
   (b) The proportion $\hat{p}_2$ of yawning within the unseeded group.
   (c) The overall proportion $\hat{p}$ of yawning.
4. Suppose there is no association between yawning and proximity to another yawner.
   (a) If you repeated this experiment many time, what would you expect to be the **average** difference between $\hat{p}_1$ and $\hat{p}_2$?
   (b) On the other hand, if you repeated this experiment just once, do you think the outcome you described previously would **always** happen?
   (c) What do you think is the most extreme difference you would plausibly see between $\hat{p}_1$ and $\hat{p}_2$?

Activity

In this part, we will study variation in the differences in proportion that is just due to random sampling. In particular, if we assume proximity to yawning has no effect on future yawning, then assigning subjects to the seeded and unseeded groups is superfluous, since a given subject is just as likely to yawn if placed in the seeded group as the unseeded one.

We assume that exactly 14 of the 50 subjects will yawn, since this was the total number who yawned in the original experiment. We then choose 34 of the 50 subjects at random to be in the seeded group and put the remaining subjects in the unseeded group. Finally, we count how many of the yawners were placed in the seeded group and how many were in the unseeded group, and then compute the difference in proportions.

Follow the instructions for sampling in the `Yawning.rmd` file on the schedule page of the course website. Once you have sampled at least 10 times, enter your data in the class's shared google sheets (link in Slack)

Reflection

After you have submitted your data, reflect on the following questions:

(a) How many yawns would you expect in the seeded group (on average)?

(b) What did you observe?

(c) How would we summarize these results? What conclusions can we draw from this experiment?

Conclusions

We have two competing hypotheses to explain the discrepancies in the proportions we observed:

The **Null Hypothesis** is that yawning and seeding are *independent*, and so any observed differences in proportion was just due to chance.

The **Alternative Hypothesis** is that yawning and seeding are *dependent*, and this is why we observed a difference in proportion.

(a) Based on the data collected by the class, if the Null Hypothesis were true, how likely would it be that we observe a difference in proportion as extreme (or more so) as we did in the original experiment?

(b) Based on this, do you think the outcome we observed was a plausible outcome assuming the Null Hypothesis were true?

(c) If we find the observed outcome implausible, we should reject the null hypothesis in favor of the alternative. Otherwise, we must withhold judgment until we collect more data. In this case, which decision seems more appropriate?