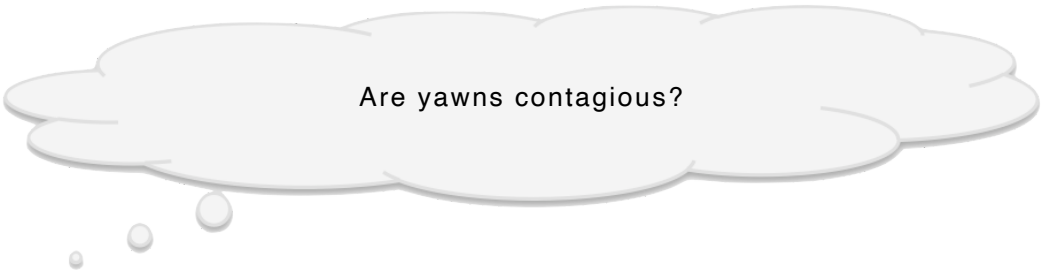
Contagious Yawns

Conventional wisdom says yawns are contagious; when you see someone else yawn, you are prone to feel sleepy and let out a yawn yourself. How many times have you caught yourself in this situation, or noticed it in someone else? But will this hypothesis withstand a scientific test? Will data support this claim?

The folks at MythBusters, a popular television program on the Discovery Channel, investigated this issue by using a two-way mirror and a hidden camera. Fifty subjects sat in a booth, accompanied only by an experimental attendee. For some of the subjects, the attendee yawned (planting a yawn "seed"), while for other subjects the attendee did not yawn. The researchers decided in advance, with a random mechanism, which subjects would receive the yawn seed and which would not. As time passed, the researchers watched to see which subjects yawned. In this activity, you will answer the following research question:



Explore the Observed Data

1. Based on the research question, specify the treatment variable.

2. Based on the research question, specify the response variable. Also, identify whether it is a quantitative or categorical variable.

Observed Data: The researchers found that ten of 34 subjects who had been given a yawn seed actually yawned, compared with four of 16 subjects who had not been given a yawn seed.

3. Organize these data/results (i.e., frequencies) into a 2x2 table. This table is sometimes referred to as a **contingency table**.

	Yawn	No Yawn	Total
Yawn Seed			
No Yawn Seed			
Total			

4. Of the 34 subjects assigned to the yawn seed condition, what proportion yawned?

5. Of the 16 subjects assigned to the no yawn seed condition, what proportion yawned?

6. Find the difference between the proportion of subjects assigned to the yawn seed condition that yawned and the proportion of subjects assigned to the no yawn seed condition that yawned.

7. Write a few sentences summarizing the results in the sample. This should include a summary of what the data suggest about: (1) the *overall percentage of participants who yawned;* (2) the differences between the two treatment groups; and (3) whether or not the data appear to support the claim that yawns are contagious.

Dummy Coding the Outcome

One trick that statisticians use to deal with categorical outcomes, is to *dummy code* the variable. Dummy coding is a way to turn the categories into numbers, so that the outcome becomes "quantitative". Once the outcome is quantitative, we can compute means, etc.

The idea of dummy coding is that each category of the outcome gets a numerical value of either "1" or "0". For example, consider the 16 subjects who did not receive the yawn seed. In the data, four of them yawned; give those folks a "1". Twelve of them did not yawn; assign them a "0". Their data looks like this:

Condition	Outcome	Yawn?
No yawn seed	Yawn	1
No yawn seed	Yawn	1
No yawn seed	Yawn	1
No yawn seed	Yawn	1
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0
No yawn seed	No yawn	0

In this table the column “Yawn?” is the dummy coded outcome. You can think of the the variable name, “Yawn?”, like the question, “Did the participant yawn?” The two responses are “1” (Yes), or “0” (No).

Why choose “0” and “1” as the numbers for our categories? Why not choose “5” and “10”; or “2” and “3”? The reason we choose “0” and “1” is that these values give meaning to interpretations of the mean and other summaries.

8. Compute the mean of the Yawn? dummy coded values for the 16 participants in the “No Yawn Seed” condition. Compare that value to your response to Question 5.

9. Re-code the outcome for the 16 participants to create a variable called Didn’t yawn?. Use dummy coding to do this. This time participants who didn’t yawn would get a “1” and the participants who waned would get a “0”. Now compute the mean of the Didn’t yawn? dummy coded values for the 16 participants in the “No Yawn Seed” condition. What does this value tell you?

10. Explain how to interpret the mean of a dummy coded variable generally.

Modeling the Experimental Variation

You will answer the research question by using TinkerPlots™ to conduct a **randomization test** in order to account for experimental variation (variation in the difference of means just because of random assignment) *under the assumption that there is no effect of the yawn seed*.

- Set up a model that will produce the **fixed dummy coded responses** for all 50 of the subjects (use Yawn = 1 and No yawn = 0). You can use a Mixer, but since there are only two values for the response variable, it can be quicker to use a Stacks device. (If you have forgotten how to do this, refer back to the instructions in the *Sleep Deprivation* course activity.)
- Add a linked Stacks device that includes the **condition labels**. (If you have forgotten how to do this, refer back to the instructions in the *Sleep Deprivation* course activity.)
- Run the model.

11. Should the sampling device containing the outcome values be sampled *with* or *without* replacement? What about the sampling device containing the condition labels? Explain why.

Plotting and Collecting the Results

- Use TinkerPlots™ to plot the results for the trial.
- Collect the results from the trial.

Simulate and Evaluate the Results

- Carry out 500 randomized trials of the simulation in TinkerPlots™.
- Plot the differences in means for the 500 simulated trials.

12. Sketch the plot of the distribution of simulated differences.

13. Based on the hypothesized model, what is the *expected* difference in means? Explain.

14. Compute and report the standard deviation of the differences in means.

15. Using the expected difference in means and the standard deviation, provide a range of likely results under the hypothesized model.

16. How compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.

Quantifying the Likelihood of the Observed Result Given the Model: p-Value

17. Compute and report the *p*-value for the observed difference in the *Yawn* study. (If you have forgotten how to do this, consult the *p*-value reading.)

18. Interpret the *p*-value you computed.

    The *p*-value of ____ is the probability of ...

19. Based on the *p*-value you computed, how compatible is the observed difference in means with the results produced by the model specified in the null hypothesis? What does this suggest about the answer to the research question? Explain.