Three Demonstrations that Build Intuition for Hypothesis Test Properties

Julie Deeke
Department of Statistics, University of Illinois Urbana-Champaign

Hypothesis Test Properties

Students struggle to conceptualize models that represent hypothesis testing.[1,2] Some of the properties that are challenging include that:

- parameters have true but unobserved values in the Frequentist framework,
- p-values are calculated with a hypothetical proof by contradiction approach, and
- evidence is evaluated based on the null hypothesis claim.

Recognizing these challenges, I aimed to design and implement activities to help students build intuition for these hypothesis test properties through non-mathematical and contextually-relevant examples.

Context

These activities were used during a second course in data science. All students had taken the prerequisite algebra-based first data science course or passed its proficiency exam, which included procedural hypothesis testing. Student background varied widely, including those with no other statistics training, those with AP Statistics experience, and Statistics majors.

The activities were used during the 24-25 academic year in 3 lecture sections for a total of 599 students. The previous unit focused on sampling distribution properties. The following lecture introduces formal simulation-based hypothesis testing.

Candy in a Paper Sack

Students pick up a paper sack that contains candy (mini Snickers, Milky Way, Reese's, Twix, Jolly Ranchers). Using any sense but sight, students answer: do they think they have a Snickers in their bag?

They reveal their candy and share the outcome of their prediction + bag. This helps to reinforce the idea that there is a true, hidden value for a parameter. It also demonstrates how students use available data to make an inference, which is sometimes correct.

When Should We Do Dishes?

How can we demonstrate to students that a p-value is part of a constructive argument in a way that they understand? Can I also demonstrate the "proof by contradiction" approach to the p-value calculation at the same time?

I make two opposing arguments for why dishes don't have to be completed immediately. One is forceful and less compelling to students. The other starts with supposing that dishes are completed immediately and then logically describes the many negative consequences that result from the one decision.

Students observe the effectiveness of a logical argument. They also see how a proof by contradiction can be applied.

Reddit's changemyview

Reddit is a social media platform and forum with conversations organized in subchannels, including a University of Illinois one. We focus on a couple of posts from the changemyview (CMV) subreddit, where posters provide their original opinion that they accept might be flawed. Other users then make arguments that support and/or persuade the author.

My goal is for students to observe starting with a potentially flawed premise (hypothesis) that we evaluate using available data to determine if we have compelling evidence to update our original premise.
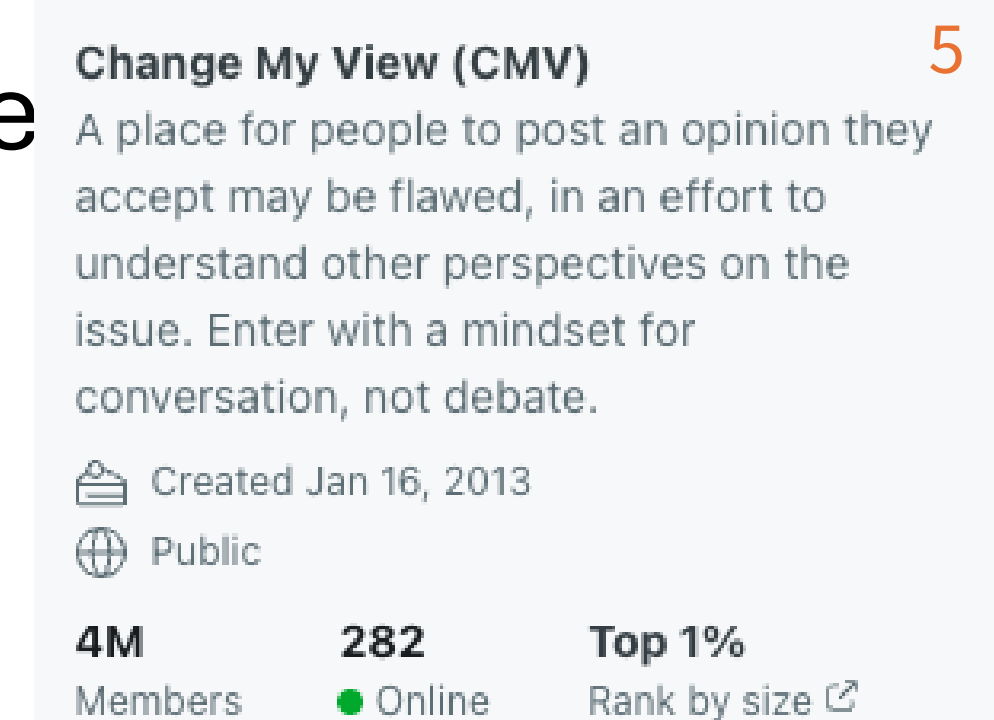
Next Steps

- Upcoming adjustments to the activities include making them more interactive and helping students to construct the connection to the upcoming course content.
- A more systematic evaluation of the effectiveness of these activities remains to be performed.
- An ethical extension to the Reddit activity will be incorporated, as AI research was recently performed within the CMV subreddit.[6]

References & Sources

1 Park, R. (2018). Practical Teaching Strategies for Hypothesis Testing. *The American Statistician*, 73(3), 282-287.
2 Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on *p*-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129-133.
3 Image from https://www.marketplace.org/story/2024/08/02/why-do-paper-bags-have-a-zig-zag-cut-at-the-top
4 Image from https://pngtree.com/freepng/dishes-soaking-in-the-kitchen-sink-multicolored-modern-colorful_15695140.html
5 Screenshot from https://www.reddit.com/r/changemyview/
6 O'Grady, C. (2025). 'Unethical' AI research on Reddit under fire. *Science*.