# Statistical Inference and the Scientific Method

Ryan Miller

# Outline

1. The scientific method
   - framework and investigative steps
2. Falsifying a hypothesis
   - a conceptual framework for statistical testing

# The scientific method

1. Propose a hypothesis
2. Collect data pertaining to the hypothesis
3. Assess the *strength of evidence* provided by the data and reach a conclusion
4. Repeat steps #2 and #3 until a consensus is reached

# Falsifiable hypotheses

In step #1, we focus on hypotheses that are *testable* and **falsifiable**, meaning you could observe evidence that *disproves the hypothesis*.

1) $H_1$ : There once was life on Mars
2) $H_2$ : There's never been life on Mars

Consider $H_1$ and $H_2$, which of these is a *falsifiable hypothesis*? Briefly explain.

# Statistical hypotheses

As statisticians, we focus on **statistical hypotheses** related to *population parameters*:

1) $H_1 : \mu_1 - \mu_2 = 0$
2) $H_2 : \mu_1 - \mu_2 \neq 0$

Consider $H_1$ and $H_2$, which of these is a *falsifiable hypothesis*? Briefly explain. (Hint: think about using the sample difference in means $\bar{x}_1 - \bar{x}_2$ as evidence)

# Statistical hypotheses

1) $H_1 : \mu_1 - \mu_2 = 0$ is *falsifiable*
2) $H_2 : \mu_1 - \mu_2 \neq 0$ is *not falsifiable*

A falsifiable statistical hypothesis must imply a specific value (ie: zero) for the population parameter. Otherwise, it could never be disproven by sample data (because even a sample difference of exactly zero doesn't disprove $H_2$ due to *sampling variability*)

# The scientific method and statistical inference

Using statistical methods to establish a scientific relationship requires the following steps:

1) Evaluate the possibility of *bias* and *confounding variables* (ie: study design)
2) Propose a falsifiable hypothesis stating that the desired relationship *doesn't exist* and then use statistical methods (ie: confidence intervals) to establish sufficient evidence against that hypothesis.
3) Have others independently replicate our conclusions.

We'll focus on #2 for the remainder of the semester, but #1 and #3 are just as important to keep in mind.

# Practice

A/B testing is a method used by market researchers to optimize the engagement or satisfaction of product users/consumers. Consider an A/B testing experiment that randomly assigns visitors to a website to one of two landing pages (page A or page B) and records their click-through rate (ie: whether they clicked anything on the page):

| | Click | No | Total |
|---|---|---|---|
| A | 19 | 66 | 85 |
| B | 16 | 65 | 81 |
| Total | 35 | 131 | 166 |

1) How concerned, if at all, should we be about the possibility of bias or confounding variables in this study?
2) Is the outcome variable categorical or quantitative? With this in mind, what is the falsifiable hypothesis these researchers should evaluate?

A/B testing is a method used by market researchers to optimize the engagement or satisfaction of product users/consumers. Consider an A/B testing experiment that randomly assigns visitors to a website to one of two landing pages (page A or page B) and records their click-through rate (ie: whether they clicked anything on the page):

| | Click | No | Total |
|---|---|---|---|
| A | 19 | 66 | 85 |
| B | 16 | 65 | 81 |
| Total | 35 | 131 | 166 |

3) Use statistical methods to construct a 95% CI estimate for the difference in the proportions. Does this confidence interval suggest these data are sufficient to statistically disprove the hypothesis (from Question #2)?
4) If we'd used a lower confidence level, is it possible that our interval might support a different conclusion?

# Practice (solution)

1) This experiment was randomized, so confounding variables should not be a concern. Other biases are also unlikely to be present.
2) We should propose the hypothesis the click rate (categorical) is the same for each page design. Statistically, this is $H : p_1 - p_2 = 0$ (where $p_1$ and $p_2$ represent the population-level click rate for each design).
3) The 95% CI is found by:
$19/85 - 16/81 \pm 1.96 * \sqrt{\frac{19/85*(1-19/85)}{85} + \frac{16/81*(1-16/81)}{81}} =$ (-0.099, 0.151)
4) Yes, notice when $c = 0.4$ the interval is entirely positive. However, this corresponds to a 31% confidence level.

# Next Steps

- Confidence intervals can be used to evaluate statistical hypotheses, but they aren't the best tool for doing so
- We'll spend the remainder of the semester covering *hypothesis testing*, a broad area of statistics aimed at more precisely quantifying the degree of compatibility the sample data has with a null hypothesis