BIOSTAT 702: Module 3

One Sample Inference; Part 2: Binary Outcome

Dr. Marissa Ashner

Department of Biostatistics and Bioinformatics

Fall 2025

Duke University School of Medicine

Module Goals

- Be able to run one-sample hypothesis tests for a binary outcome

Resources for this Module

Textbooks

- ADLM: Chapter 3

Let's Change our Running Example

- What if, instead of the average height of Duke students, we want to measure the proportion (π) of Duke students that are at least 6 feet (72 inches)?
- Each person would contribute a binary outcome: Yes (or 1) if they are at least 6 feet, or No (or 0) if not
- Similar to a continuous outcome, the best estimator for the population proportion is simply the *sample proportion* ($\hat{\pi}$)

Sampling Distribution for a Sample Proportion

- The sampling distribution for a sample proportion will follow a **binomial distribution**
- We know that the mean of the binomial distribution is $\boxed{n * \pi}$ and the variance is $\boxed{n * \pi * (1 - \pi)}$
  - Since we are interested in the proportion itself and not the number of subjects, we will divide by n, and the mean and variance of our sampling distribution will be $\boxed{\pi \text{ and } \pi * (1 - \pi)/n}$
  - Why do we divide by n for the variance?
- Although the exact sampling distribution is binomial, the CLT still applies here
  - As n increases, the distribution approaches a normal distribution

One-Sample Test for a Population Proportion π

We think the proportion of Duke students at least 6 feet tall is 30%. We know our sample proportion will likely vary from the true average height, but we want to know if our sample proportion is likely drawn from a distribution where the true proportion of those over 6 feet tall is 30%.

CLT

- *Statistical Hypotheses:* $H_0 : \pi = 0.3$, $H_A : \pi \neq 0.3$
- *Test Statistic:* $Z = \frac{\hat{\pi} - 0.3}{\sqrt{0.3(1-0.3)/n}} \sim N(0, 1)$
  - Why did we choose this?
- *p-value:* $P(Z \geq |z_{obs}| | H_0 \text{ is true})$
- *interpret:*
  - If $p < \alpha$, reject the null, concluding that there is sufficient evidence to support the alternative
  - If $p \geq \alpha$, do not reject the null, concluding there is *insufficient* evidence to support the alternative

Confidence Intervals for a Sample Proportion

- Previously, we talked about confidence intervals in terms of an inverted hypothesis test
- For a sample proportion, we have to be more careful
  - Notice that the standard error is directly calculated using the proportion itself
  - For hypothesis testing, we calculate the distribution of the estimate *under the null* (i.e., $\pi = 0.3$)
  - For CI, however, we use the *estimate* as part of the standard error instead

The Exact Binomial Test

- What if your sample size is too small to approximate the normal distribution?
- You can perform a hypothesis test using the exact sampling distribution (i.e., the binomial)
  - It is more taxing to write out "by hand" but can be done easily by software
- Achieving an exact desired type I error rate may not be possible due to the binomial distribution being discrete

Q & A

Questions?