

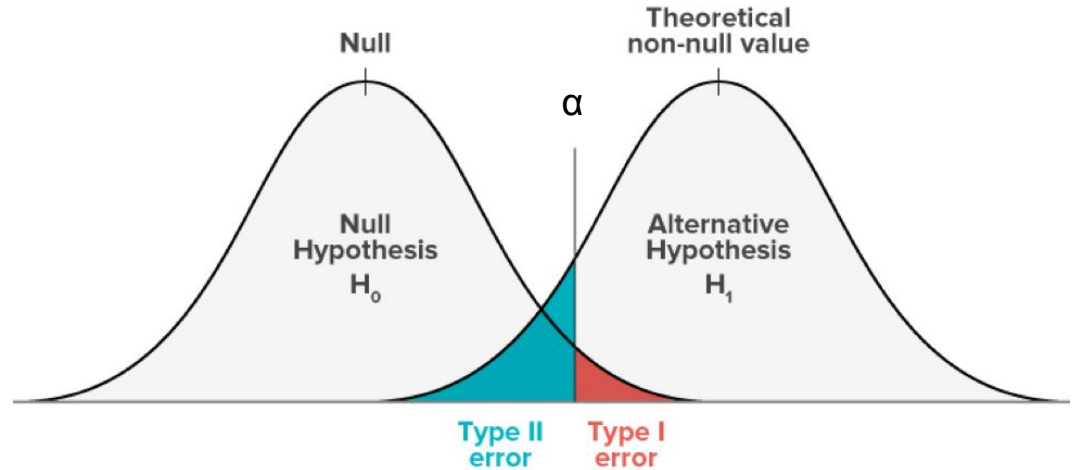
Discussion 2

8/13/2024

Dan Chen

Type I and Type II Error illustrated

	H_0 True	H_0 False
Accept H_0 /Reject H_1	True	False: Type II error
Reject H_0 /Accept H_1	False: Type I error	True



Practice Problem (2.3.17)

Suppose that you are considering whether to publish a weekly alternative newspaper on campus. You decide to survey a random sample of students on your campus to ask if they would be likely to read such a newspaper. Your plan is to proceed with publication only if the sample data provide strong evidence that more than 10% of all students on your campus would be likely to read such a newspaper.

- a. Identify the parameter of interest, in words.
- b. Express the appropriate null and alternative hypotheses for conducting this test.
- c. Specify what Type I error represents in this situation. Also indicate a potential consequence of Type I error.
- d. Specify what Type II error represents in this situation. Also indicate a potential consequence of Type II error.

Practice Problem (2.3.14)

The significance level α determines the probability of making a Type I error. Errors are bad. So, why don't we always set α to be extremely small, such as 0.0001?

Additionally how can we mitigate the problem of setting an extremely small α value?
(Hint: How can you reduce the standard error of the null distribution?)

Complete the code walkthrough in “Discussion 2.Rmd”

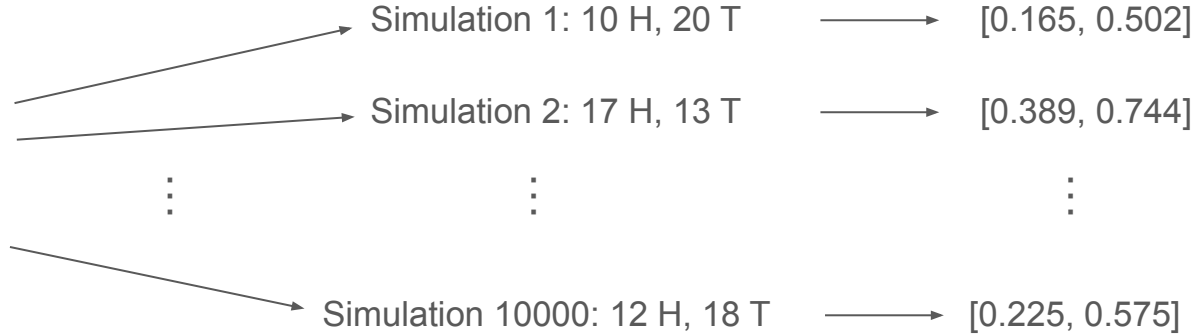
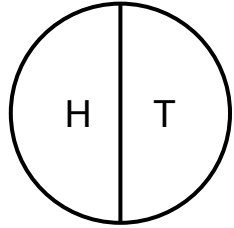
Confidence Interval equivalence to Hypothesis Tests

- A confidence interval is a random variable or statistic that is calculated from a random sample
- If we take multiple samples from a population and construct a 95% confidence interval from each sample. We would expect 95 percent of the intervals to contain the true population parameter.
- We can also conclude that the probability of observing a confidence interval that does not contain the true parameter is 5%
- A p value of a hypothesis test can be interpreted as the probability that a value equal to or more extreme than the observed value is observed under the null hypothesis

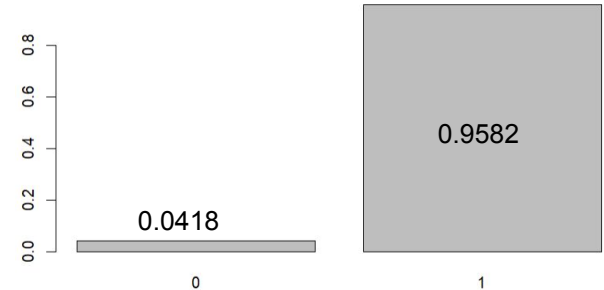
Simulation perspective

Observed: 21 Heads, 9 Tails
Confidence Interval: [0.536, 0.864]

Null hypothesis
 $P(\text{Heads}) = 0.5$



Confidence Interval contains 0.5



See Discussion 2.Rmd for
simulation code

The observed value falls within the
group that does not contain 0.5

Practice Problem (3.3.21)

Two Cal Poly freshmen gathered data on the prices for a random sample of 30 textbooks from the campus bookstore. They found the average price was \$65.02, and the standard deviation of prices was \$51.42. The data are not strongly skewed.

- a. Identify the observational unit for this study.
- b. Identify the variable of interest and whether it is categorical or quantitative.
- c. Explain how you know from the sample average and sample standard deviation that the distribution of textbook prices is not bell-shaped.
- d. In the context of this study, explain why it is valid to use the theory-based (t-distribution) approach to find a confidence interval?
- E. Calculate the 95% confidence interval for the parameter of interest

Practice Problem

Based on a four-year (2003–2007) study of over 30,000 people who were 45+ years, where all individuals were followed up regularly for diet and health changes, researchers Tsivgoulis et al. (Neurology, 2013) reported that the Mediterranean diet was linked to better memory and cognitive skills.

- a. Identify the explanatory and response variables.
- b. Identify a confounding variable whose effect on memory and cognitive skills might be confounded with that of the explanatory variable identified in part (a)