Inference for a single proportion

Bootstrap test for a proportion

Prof. Dr. Jan Kirenz

Statistical inference for categorical data

Note that there is only one variable being measured in a study which focuses on **one proportion**

For each observational unit,

the single variable is measured as either a **success** or **failure**:

e.g., “surgical complication”

vs. “no surgical complication”

Medical consultant example

Average complication rate for liver donor surgeries in the US is about 0.1 (**10%**)

Clients of the **medical consultant** have only had 3 complications in the 62 liver donor surgeries: 0.048 (**4.8%**)

Is the difference big enough?

Could the low complication rate of
p’= 0.048

have simply occurred by chance,

if her complication rate does not
differ from the US standard rate?

Variability of the statistic

We want to identify the **sampling distribution**

of the test statistic **p’**

if the **null hypothesis** was true.

We want to see the **variability**

we can expect from **sample proportions**

if the **null hypothesis** was true.

Can we reject the null hypothesis?

We plan to use this information

to decide whether there is enough evidence

to reject the **null hypothesis**.

How to simulate the variability of the statistic?

Under the null hypothesis,

___% of liver donors

have complications during or after surgery.

we could simulate 62 clients

to get a sample proportion for the complication rate

from the null distribution.

Simulating observations using a hypothesized null parameter value is often called a **parametric bootstrap simulation.**

How to simulate the variability of the statistic?

Each client can be simulated

using a bag of marbles

with 10% red marbles

and **90% white marbles**.

Sampling a marble from the bag (with 10% red marbles)

is one way of simulating whether a patient has a complication if the true complication rate is 10%.

If we select 62 marbles

and then compute the proportion of patients with complications in the simulation, p'_{sim1}

then the resulting sample proportion

is a sample from the null distribution.

How many patients had a complication rate **below 0.048** in our simulations?

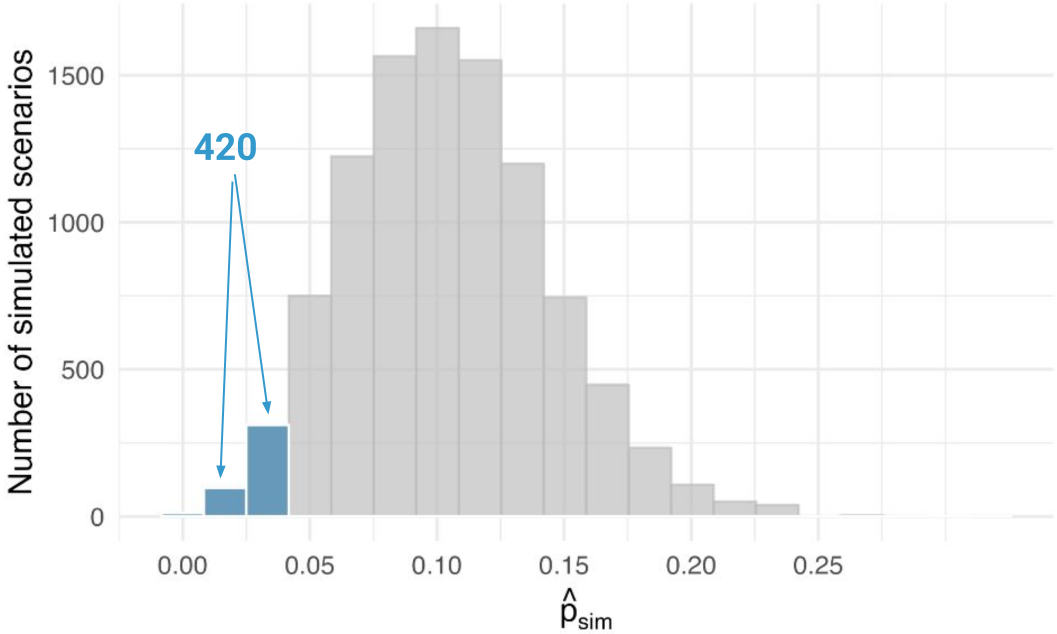


Figure 16.1: The null distribution for $\hat{p}$, created from 10,000 simulated studies. The left tail, representing the p-value for the hypothesis test, contains 4.2% of the simulations.

There were 420 simulated sample proportions with $p'_{sim} \leq 0.048$.

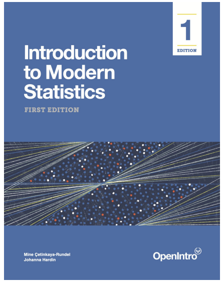
We use these to construct the null distribution's left-tail area and find the p-value:

$$\text{left tail area} = \frac{\text{Number of observed simulations with } \hat{p}_{sim} \leq \ 0.048}{10000}$$

The estimated **p-value** is 0.042

Explain what this means in plain language in the context of the problem.

Resources



The content of this presentation is mainly based on the excellent book “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin (2021).

The online version of the book can be accessed for free:

https://openintro-ims.netlify.app/index.html