

Discussion 1

8/6/2024

Dan Chen

Set up R and RStudio

Install R

- Go to <https://cran.rstudio.com/>
- Under “Download and Install R” select the link that corresponds to the type of system on your computer (Mac, Windows, Linux)
- Navigate to link for the latest version of R (4.4.1) and download the installer
- Accept the default settings on the installer

Install RStudio

- Go to <https://posit.co/download/rstudio-desktop/>
- Download the version of RStudio appropriate for your system
- Accept the default settings on the installer

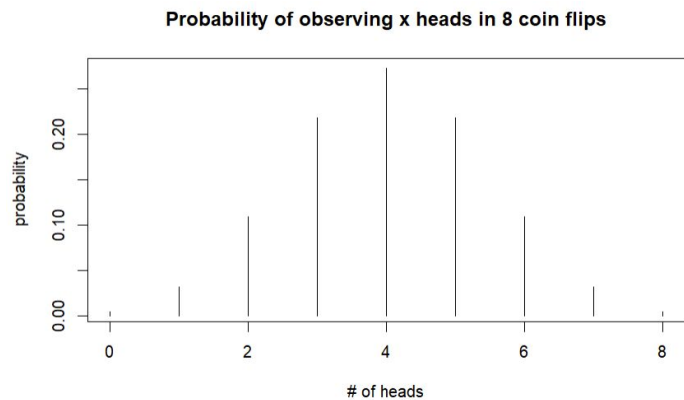
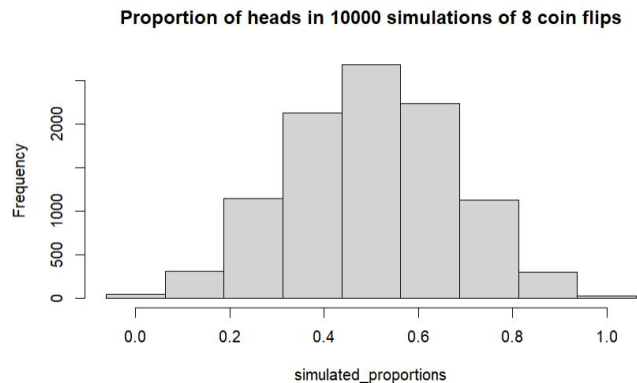
Introduction to coding in R

See *“Discussion 1.Rmd”*

Distributions and Population parameters

Example: Coin flips

- Population parameter: 0.5 chance of heads
- Sample: HHTHTHHH
- Sample statistic: .75
- Simulated distribution and Theoretical distribution

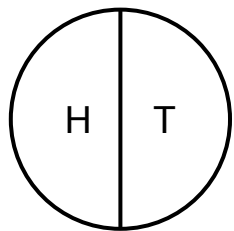


See "Discussion 1.Rmd"
for code

Simulation steps

True population parameter

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



Simulation 1: HHTTHTTT

Sample Statistic 1: 0.375

Simulation 2: THTTHHHTH

Sample Statistic 2: 0.5

⋮

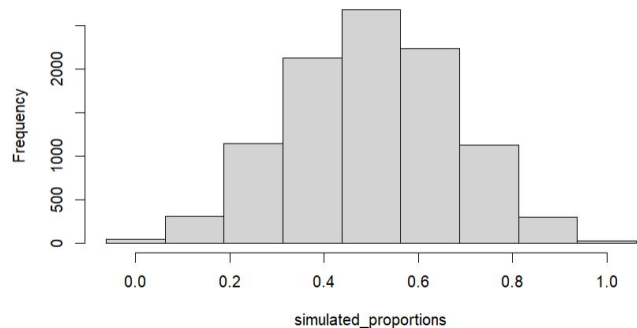
⋮

⋮

Simulation 10000: HTHHHTTH

Sample Statistic 10000: 0.625

Proportion of heads in 10000 simulations of 8 coin flips

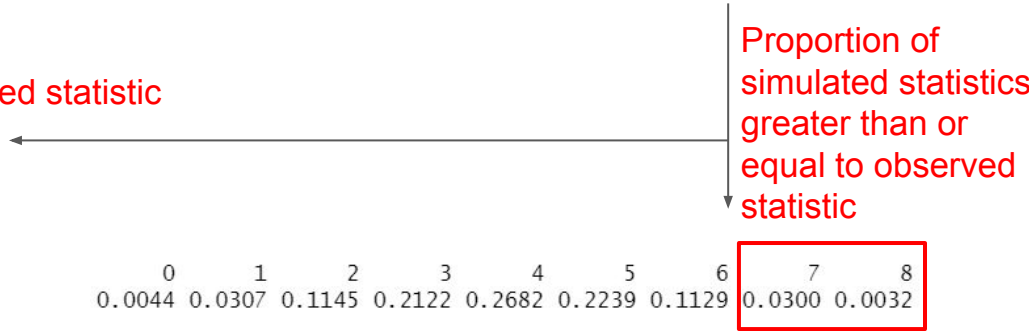
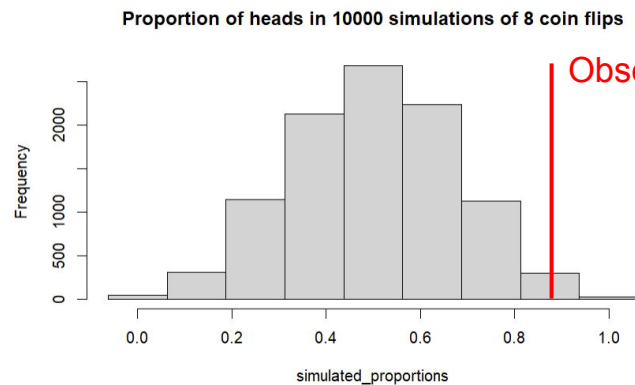
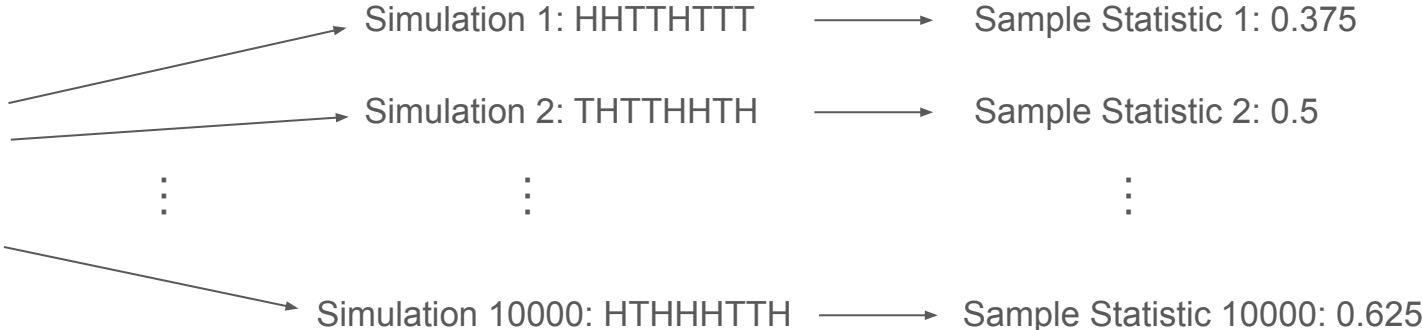
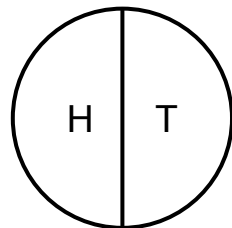


0 1 2 3 4 5 6 7 8
0.0044 0.0307 0.1145 0.2122 0.2682 0.2239 0.1129 0.0300 0.0032

Review of hypothesis testing

Observed: HHTHHHHH
Observed statistic: 0.875

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



Hypothesis testing: Theory

CENTRAL LIMIT THEOREM

If the sample size (n) is large enough, the distribution of sample proportions will be bell-shaped (or normal), centered at the long-run proportion (π), with a standard deviation of $\sqrt{\pi(1 - \pi)/n}$.

VALIDITY CONDITIONS

The normal approximation can be thought of as a prediction of what would occur if simulation was done. Many times this prediction is valid, but not always. We will consider the prediction valid when the validity condition (at least 10 successes and at least 10 failures) is met.

Hypothesis testing: Theory

Example: We want to know whether less than 40 percent of adults get more than 7 hours of sleep each night, so we sample 100 people from the population of adult humans and record whether they get more than 7 hours of sleep each night.

More than 7 hours of sleep: 30 people

Less than 7 hours of sleep: 70 people

H_0 : 40 percent of adults get more than 7 hours of sleep

H_a : Less than 40 percent of adults get more than 7 hours of sleep

Hypothesis testing: Theory

H0: 40 percent of adults get more than 7 hours of sleep

- Which means under the null hypothesis we expect 40 people in the sample to get more the 7 hours of sleep and 60 people to get less than 7 hours of sleep

Validity Condition: At least 10 people in each group ($40 > 10$ and $60 > 10$)

Standard Error = $\text{sqrt}(0.4 * 0.6 / 100) = 0.04898979$

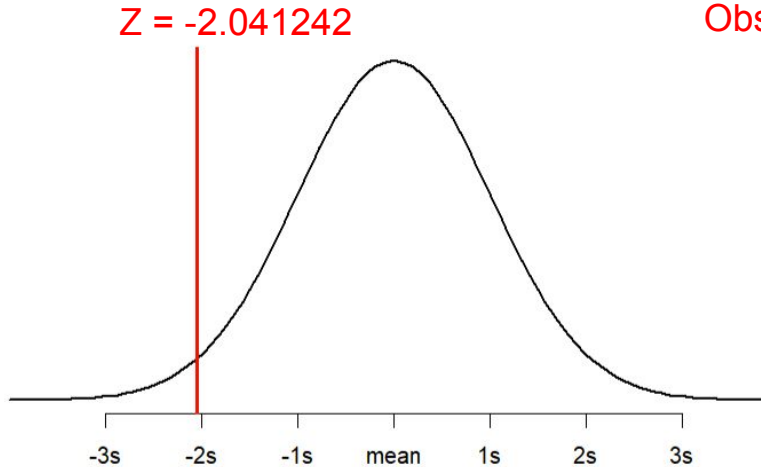
Expected Value = 0.4 (From null hypothesis)

Observed Value = $30/100 = 0.3$

Z value = $(\text{observed} - \text{expected})/\text{SE} = -2.041242$

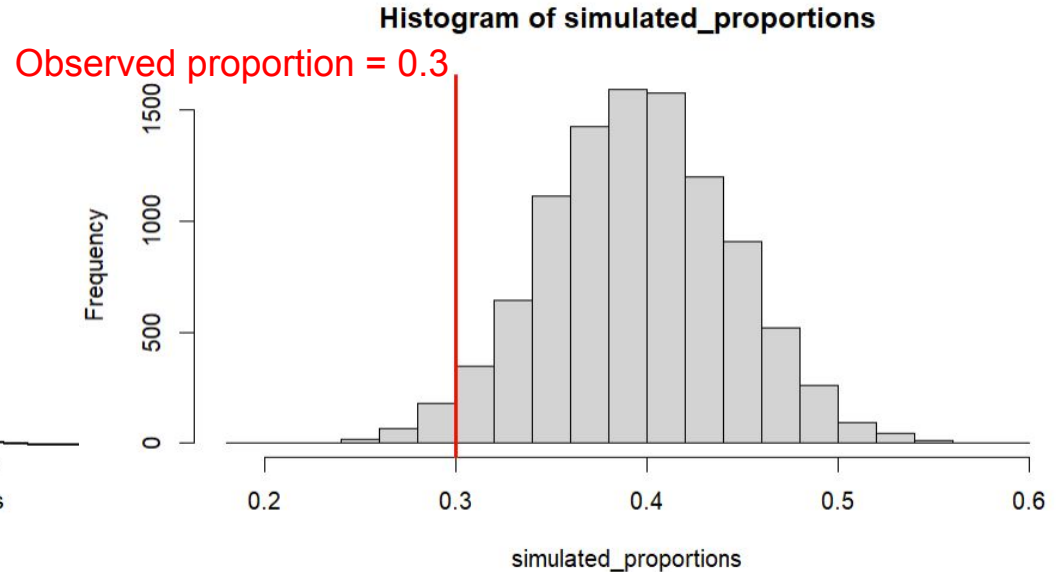
Hypothesis testing: Theory vs Simulation

Theoretical distribution:



Probability of observing $Z = -2.041242$ or less is 0.02061339

Simulated distribution:



Proportion of simulated statistics equal to 0.3 or less is 0.0268