Hypothesis Testing

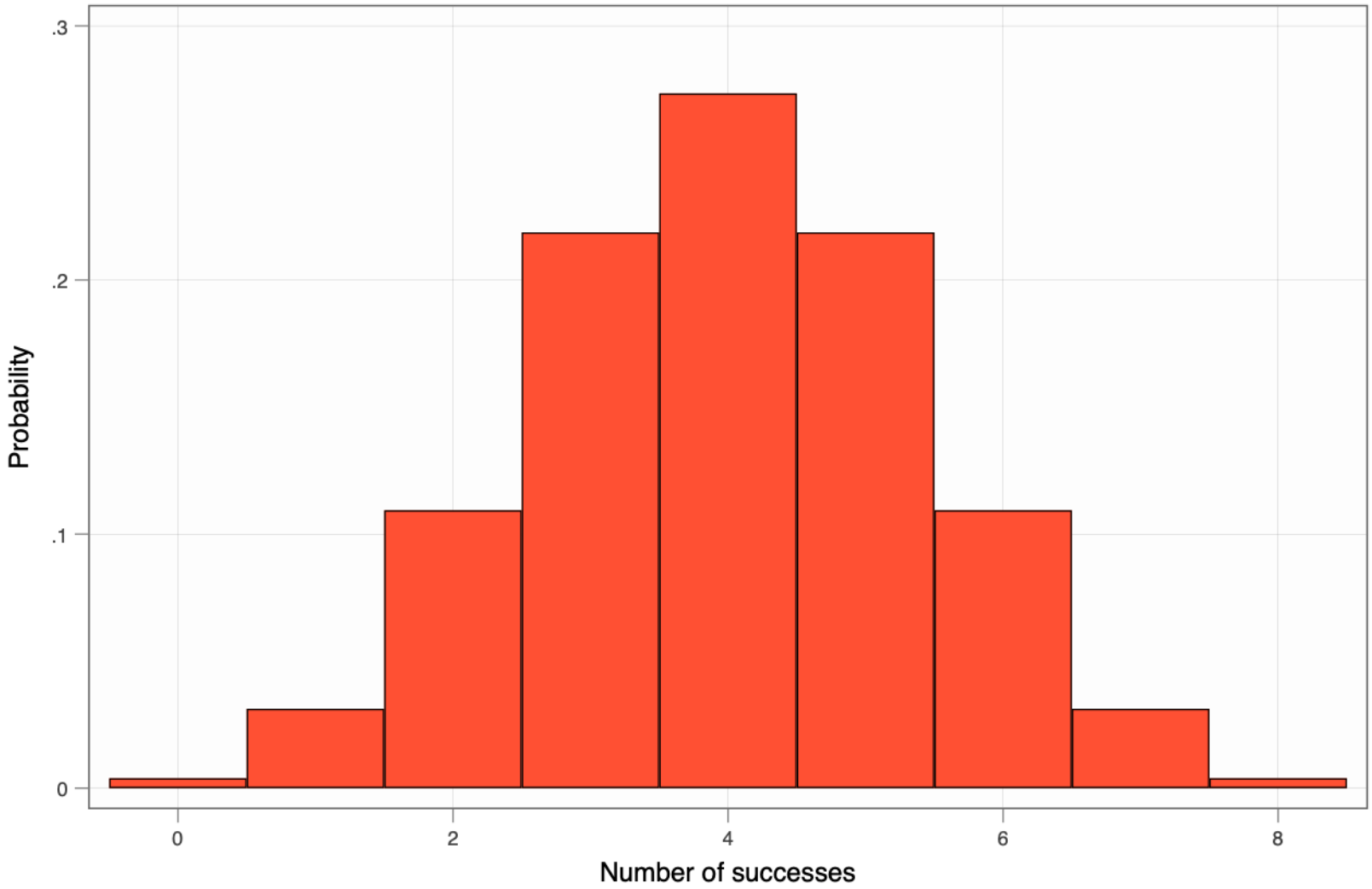
Expert Testimony

How to test hypotheses?

The lady tasting tea

- Miss Muriel Bristol claims she can taste whether tea or milk was first poured in a cup.
- Mr William Roach (her future husband) suggests an experiment: taste eight cups, each randomly (and to her) unknowingly assigned to "tea first" or "milk first."
- How would you test if Miss Bristol is correct?

Probability of k successful trials *by chance alone*



Hypothesis testing

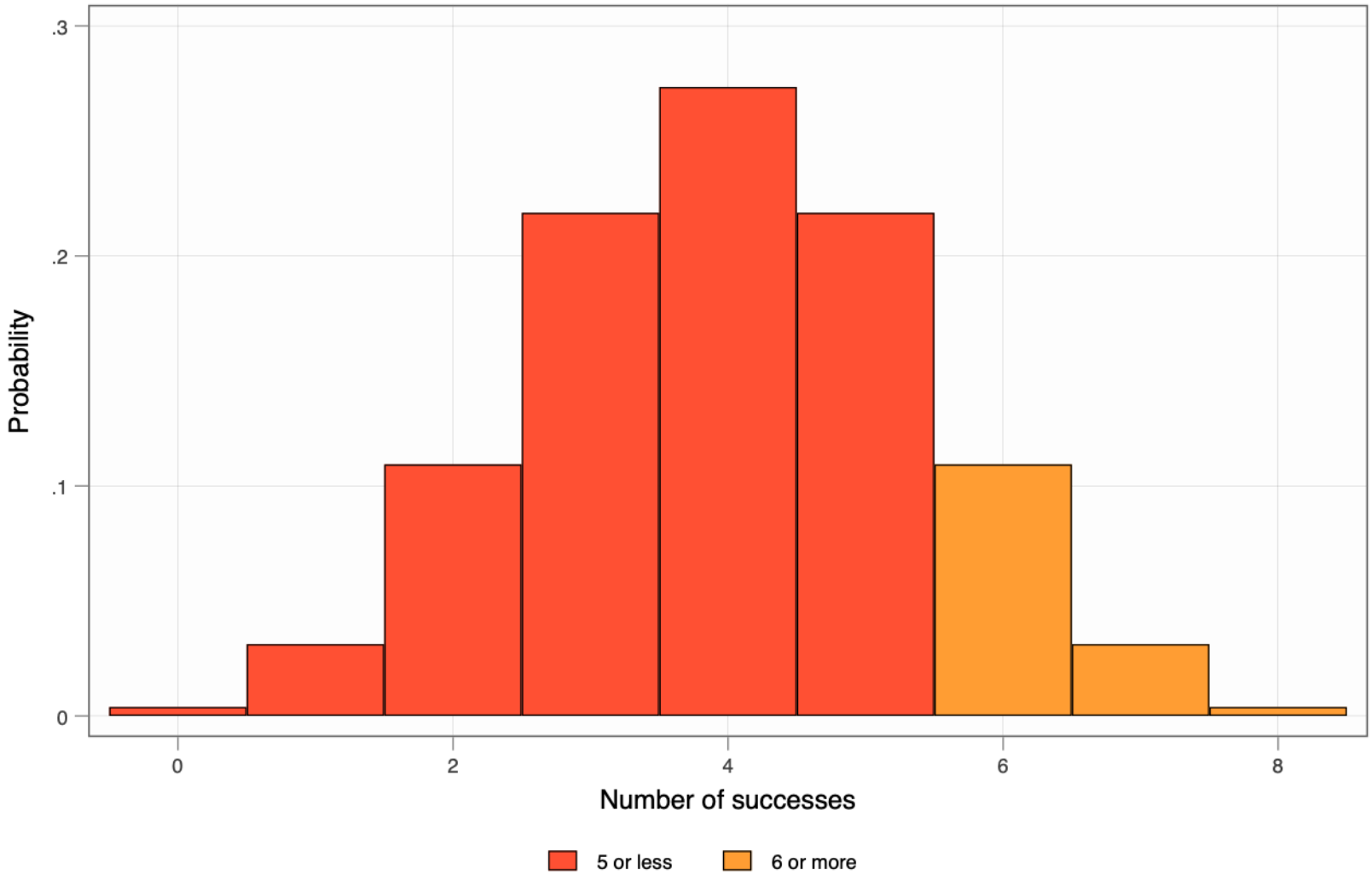
- There is a **truth** out there we want to find ("data generating process").
- The observed data varies **by chance**, as well. This somewhat masks the truth.

Null hypothesis (H0) Miss Bristol cannot tell apart "tea first" or "milk first." Formally, the probability of success is $q = 0.5$.

Alternative hypothesis (Ha) Miss Bristol is right, $q > 0.5$.

- If truth is H0, but data varies by chance, what is the probability that Miss Bristol succeeds 6 or more times?

Probability of 6 or more successful trials *by chance alone*



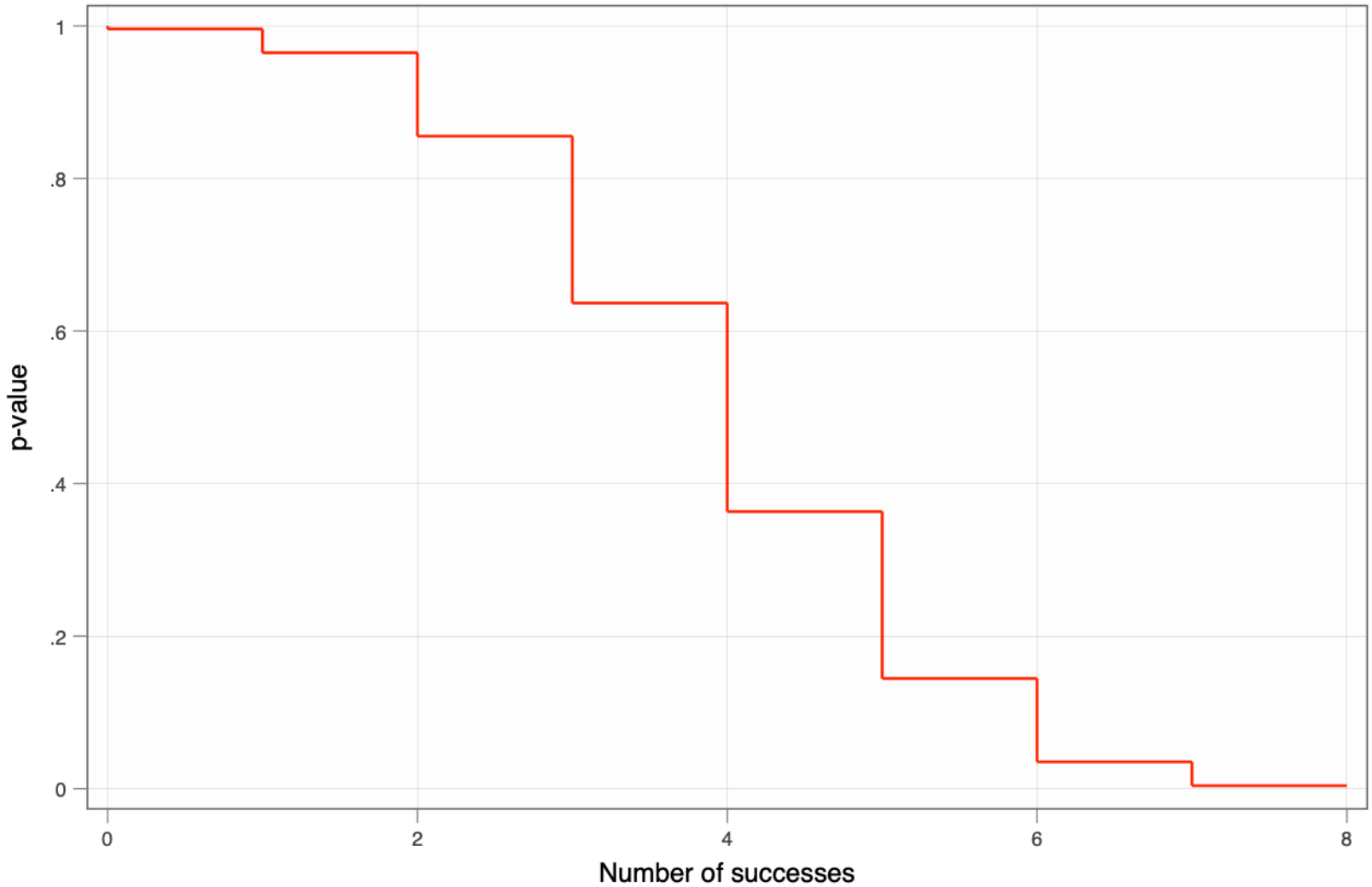
Probability of 6 or more successful trials *by chance alone*

- In this case, $p = 0.145$. There is a 14.5% probability that 6 or more successes happen by chance.
- In reality, Miss Bristol was right 8 out 8 cases. This means $p = 0.004$, or a 0.4% probability that this is purely chance.

The p-value

- p-value: Probability that we observe the data, or more extreme, *by chance alone*.
- The lower the p, the more confident we are that the null hypothesis is not true.
- We *reject the null hypothesis* if $p < 0.05$ (convention).

The p-value after k successes



What the p-value is not

- It's not the probability of H0 being true. (Though positively related.)
- $1 - p$ is not the probability of Ha being true.
- We never accept Ha. We just reject H0.

Testing in big data

Why do hypothesis testing?

When using internal data, we will often have data on the *whole population*, not a sample.

Why engage in hypothesis testing? ## An example - A casino is sued for a biased roulette wheel. - Out of 100 spins, 8 came out as 0 or 00 (casino wins). - Expected number: $2/(2+36) \times 100 = 5.3$. - Is the casino cheating?

Test it

H0 roulette wheel is unbiased ($\alpha = 0.0526$)
HA roulette wheel is biased ($\alpha \neq 0.0526$)

Compute p-value for rejecting H0.

Result

Cannot reject H0

$$p = 0.436 > 0.05$$

Lessons for hypothesis testing

- There is a **truth** out there! (discriminating practice, roulette bias)
- Even if we observe all data (not sample), it will be a noisy representation of the truth.
- Conduct hypothesis testing *as if* data is a sample.