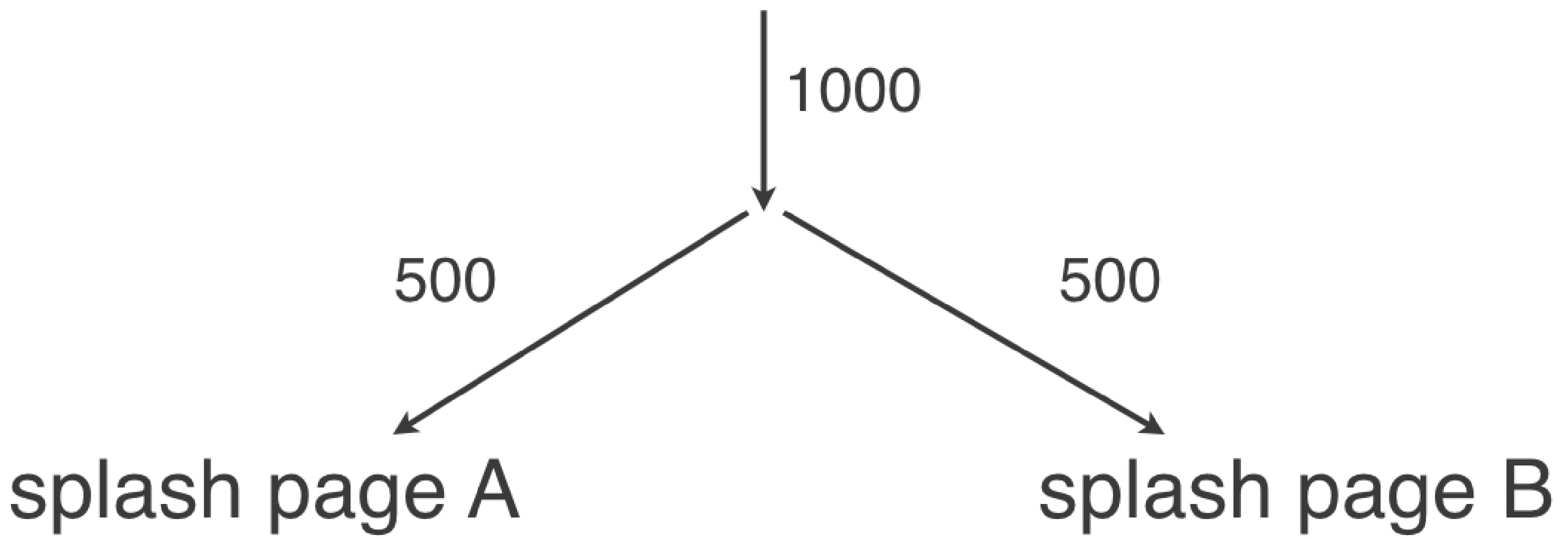
A/B testing

STATISTICAL THINKING IN PYTHON (PART 2)

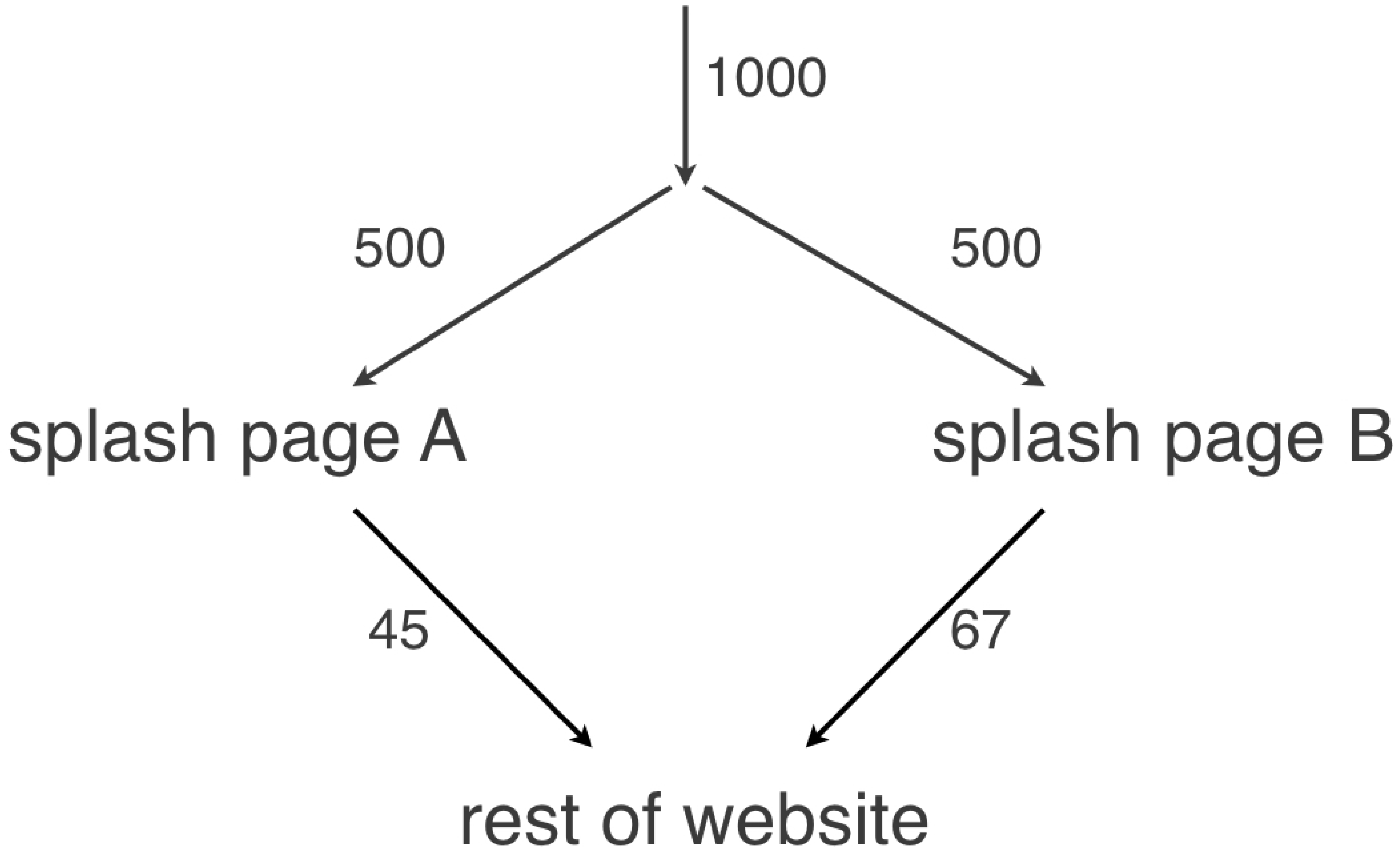
Justin Bois

Lecturer at the California Institute of Technology

Is your redesign effective?



Is your redesign effective?



Null hypothesis

- The click-through rate is not affected by the redesign

Permutation test of clicks through

```python
import numpy as np
# clickthrough_A, clickthrough_B: arr. of 1s and 0s
def diff_frac(data_A, data_B):
        frac_A = np.sum(data_A) / len(data_A)
        frac_B = np.sum(data_B) / len(data_B)
        return frac_B - frac_A
diff_frac_obs = diff_frac(clickthrough_A,
                          clickthrough_B)
```

Permutation test of clicks through

```
perm_replicates = np.empty(10000)
for i in range(10000):
     perm_replicates[i] = permutation_replicate(
                   clickthrough_A, clickthrough_B, diff_frac)
p_value = np.sum(perm_replicates >= diff_frac_obs) / 10000
p_value
```

```
0.016
```

A/B test

- Used by organizations to see if a strategy change gives a better result

Null hypothesis of an A/B test

- The test statistic is impervious to the change

Let's practice!

STATISTICAL THINKING IN PYTHON (PART 2)

Your p-value is 0.0001, which means that only one out of your 10,000 replicates had a result as extreme as the actual difference between the dead ball and live ball eras. This suggests strong statistical significance. Watch out, though, you could very well have gotten zero replicates that were as extreme as the observed value. This just means that the p-value is quite small, almost certainly smaller than 0.001.

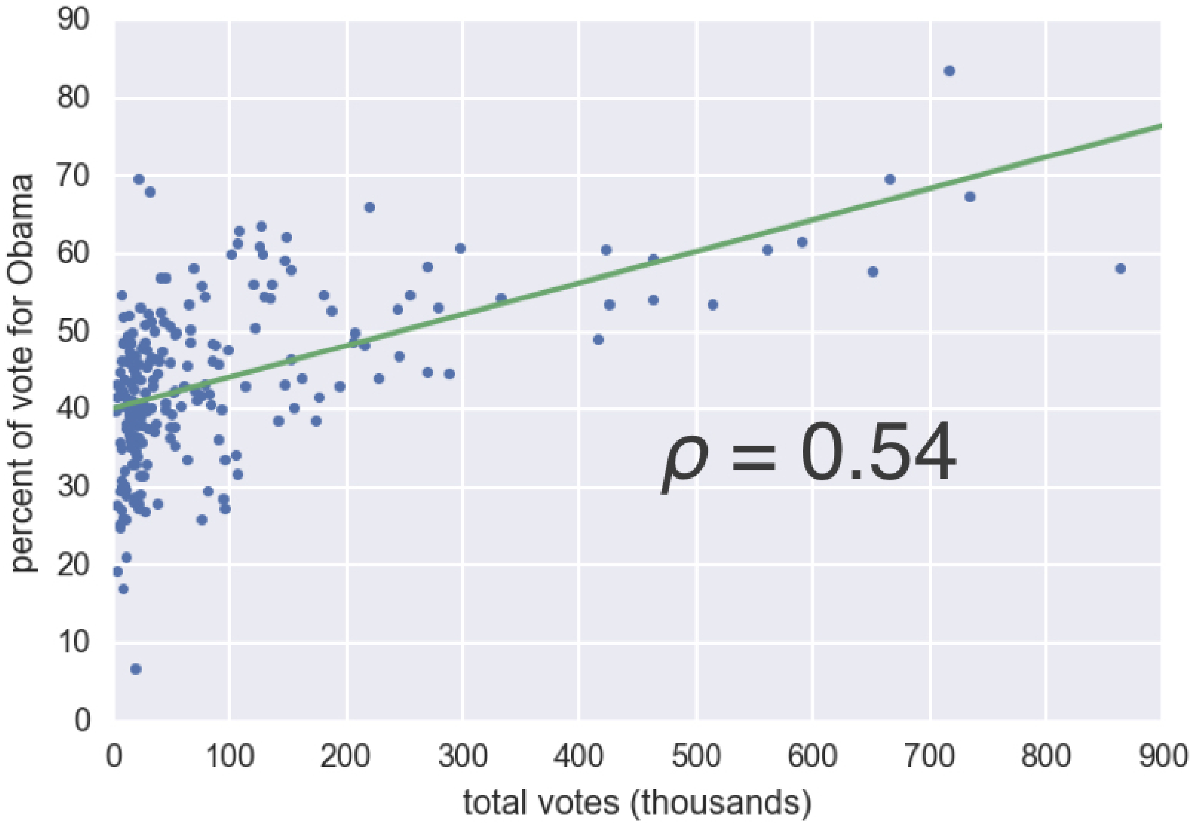
Test of correlation

STATISTICAL THINKING IN PYTHON (PART 2)

Justin Bois

Lecturer at the California Institute of Technology

2008 US swing state election results

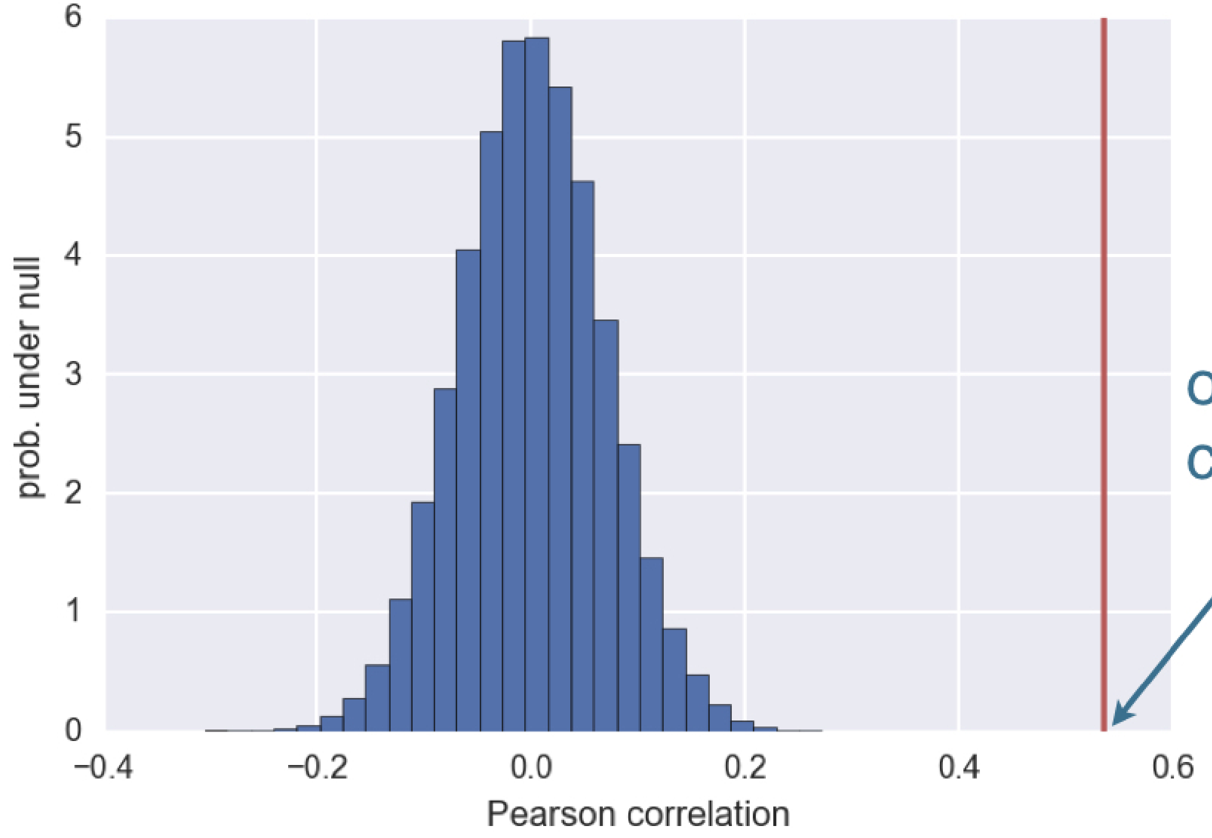


[1] Data retrieved from Data.gov (https://www.data.gov/)

Hypothesis test of correlation

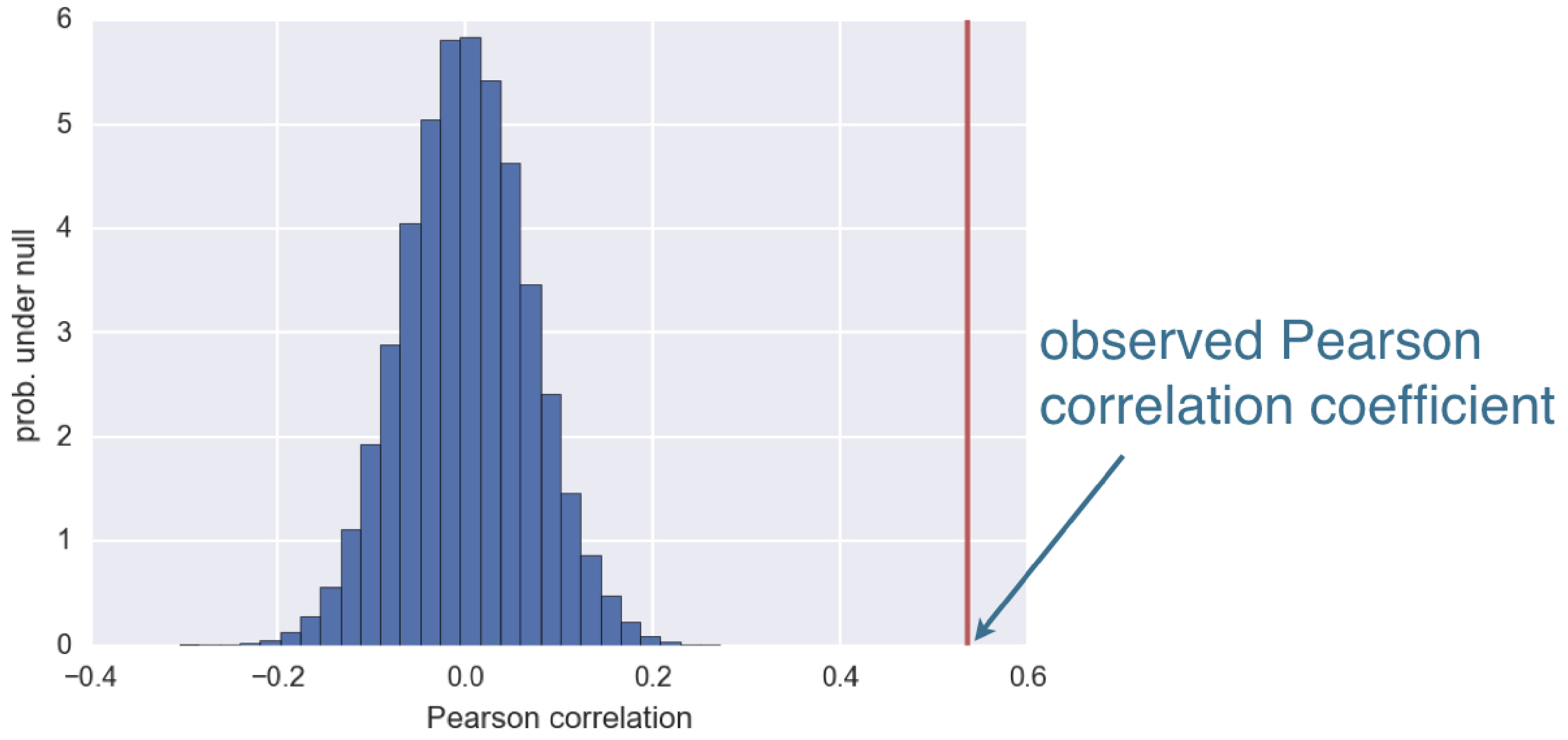
- Posit null hypothesis: the two variables are completely uncorrelated
- Simulate data assuming null hypothesis is true
- Use Pearson correlation, ρ, as test statistic
- Compute p-value as fraction of replicates that have ρ at least as large as observed.

More populous counties voted for Obama



observed Pearson correlation coefficient

More populous counties voted for Obama



observed Pearson correlation coefficient

p-value is very very small

You got a p-value of zero. In hacker statistics, this means that your p-value is very low, since you never got a single replicate in the 10,000 you took that had a Pearson correlation greater than the observed one. You could try increasing the number of replicates you take to continue to move the upper bound on your p-value lower and lower.

Let's practice!

STATISTICAL THINKING IN PYTHON (PART 2)

Nice work! The p-value is small, most likely less than 0.0001, since you never saw a bootstrap replicated with a difference of means at least as extreme as what was observed. In fact, when I did the calculation with 10 million replicates, I got a p-value of 2e-05