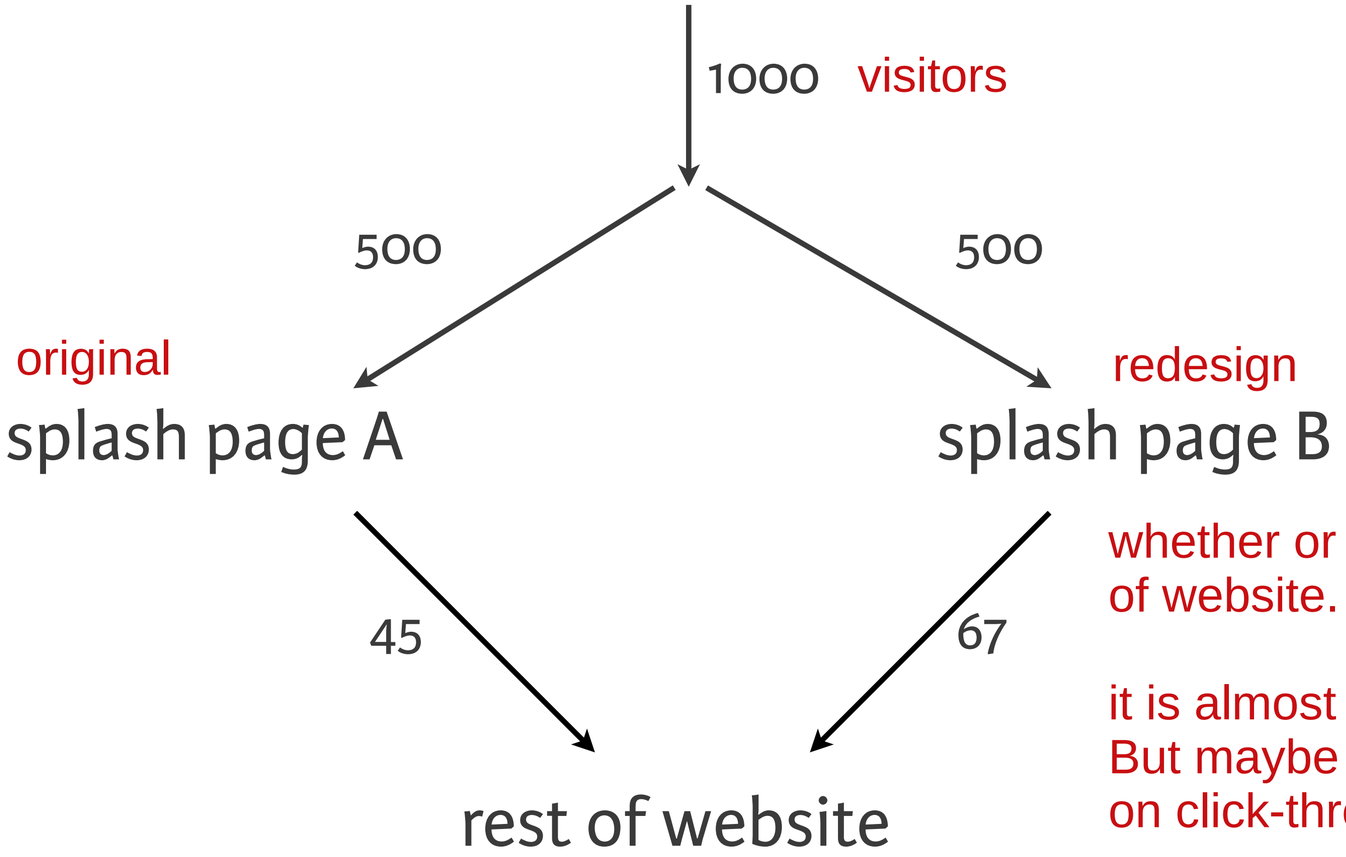
STATISTICAL THINKING IN PYTHON II

A/B testing

EX: how many more users click through to the website for the redesign vs original design

Is your redesign effective?



whether or not each of them clicks through to the rest of website.

it is almost a 50% increase in the click through rate.
But maybe there really is no diff btw the effect of 2 designs on click-through rate and diff u saw is due the random chance.

Null hypothesis

to check: what is the probability that u would observe at least the observed diff in no of clicks through if that were the case?

a permutation test is a good choice here b/c u can simulate the results as if the redesign had no effect on the click through rate.

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

Permutation test of clicks through

```
In [1]: import numpy as np

In [2]: # clickthrough_A, clickthrough_B: arr. of 1s and 0s

In [3]: 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
   ...:

In [4]: diff_frac_obs = diff_frac(clickthrough_A,
   ...:                           clickthrough_B)
```

fraction of visitors who click through

observed value of test statistic

Permutation test of clicks through

generate our permutation replicates of test statistic

```
In [1]: perm_replicates = np.empty(10000)

In [2]: for i in range(10000):
    ...:     perm_replicates[i] = permutation_replicate(
    ...:                 clickthrough_A, clickthrough_B, diff_frac)
    ...:

In [3]: p_value = np.sum(perm_replicates >= diff_frac_obs) / 10000
```

no of repl where test statistic was at least as great as what we observed

```
In [4]: p_value
Out[4]: 0.016
```

relatively small --> might reasonably think that the redesign is real improvement

A/B test

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

  give different, hopefuly, better results

Null hypothesis of an A/B test

A/B testing is just a special case of the hypo testing framework we have already been working and informative one

- The test statistic is impervious to the change

  a low p-value implies that the change in strategy lead to a change in performance.

  note: be warned that significance does not mean practical significance.
  (a diff in click through rate may be statistically significant, but if is only a couple people more per day, your marketing team may not consider the change worth the cost

STATISTICAL THINKING IN PYTHON II

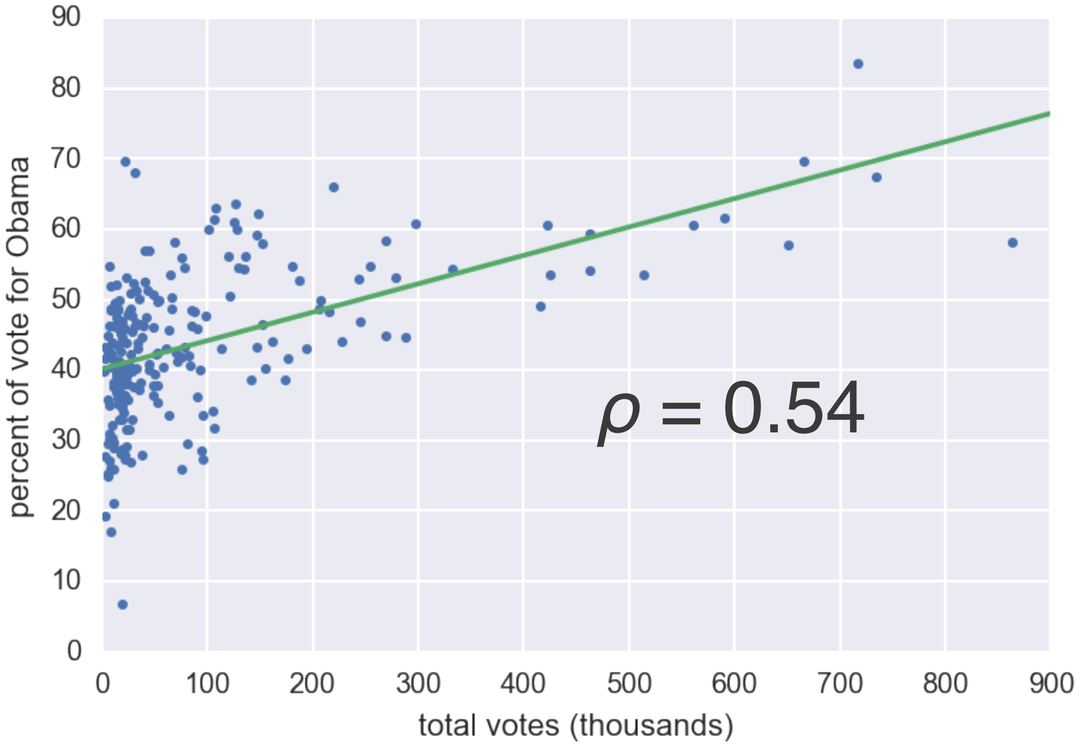
Let’s practice!

STATISTICAL THINKING IN PYTHON II

Test of correlation

2008 US swing state election results

Pearson corr coeff [-1, 1]



0.54 indicates that data are not perfectly correlated but are correlated nonetheless.

how can we know for sure if this corr is real, of if it could have happened just by chance?

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

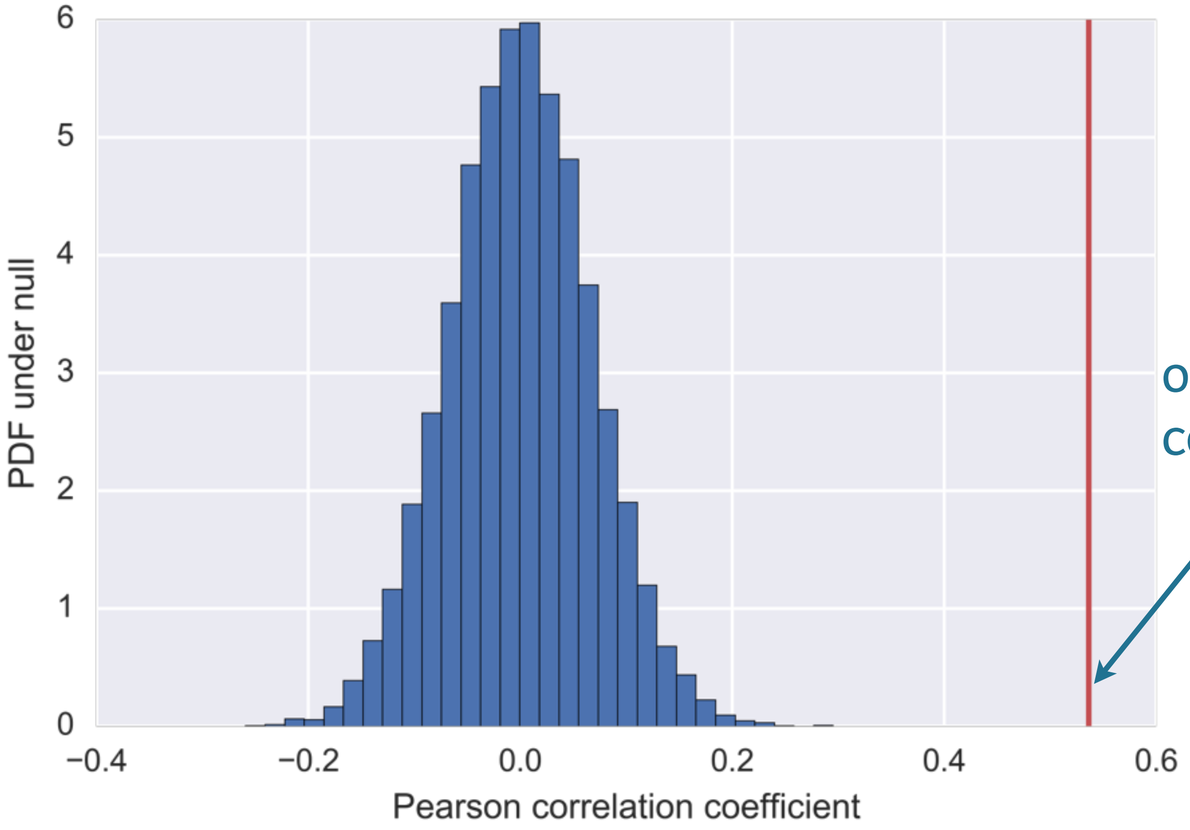
Hypothesis test of correlation

Ho: there is no correlation btw 2 variables

- 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.

did with all 10000 of replicates under Ho

More populous counties voted for Obama



Not one had a Pearson corr coeff as high as the observed value (0.54).
Then, try with 1e5 and 1e6 replicates.
In all cases, not one rep had a Pearson corr coeff as high as 0.54

this does not mean that pvalue =0

observed Pearson correlation coefficient

It means that it is so low that we would have to generate an enormous number of repl to have even one that has a test statistic sufficiently extreme.

p-value is very very small

essentially no doubt that counties with higher vote count tended to vote for Obama

STATISTICAL THINKING IN PYTHON II

Let’s practice!

The observed correlation between female illiteracy and fertility may just be by chance; the fertility of a given country may actually be totally independent of its illiteracy. You will test this hypothesis. **To do so, permute the illiteracy values but leave the fertility values fixed.** This simulates the hypothesis that they are totally independent of each other. For each permutation, compute the Pearson correlation coefficient and assess how many of your permutation replicates have a Pearson correlation coefficient greater than the observed one.