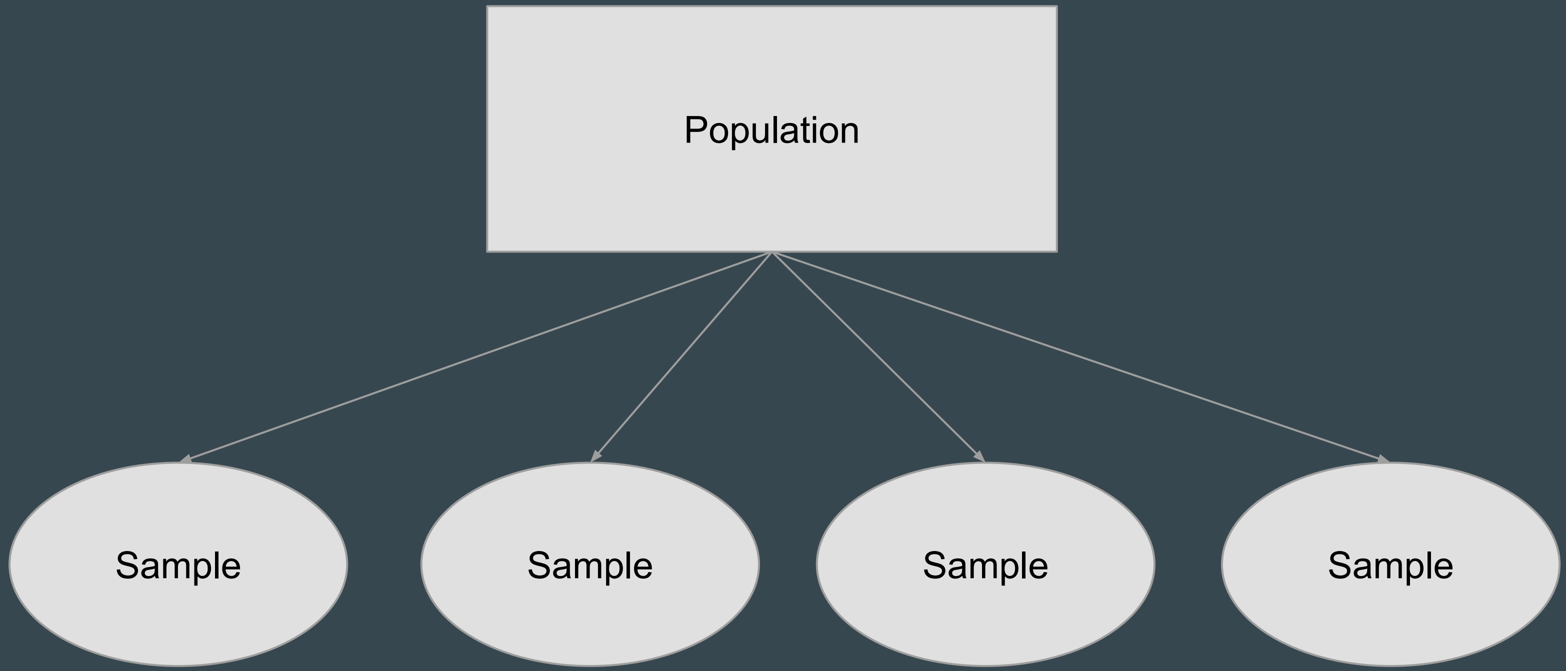
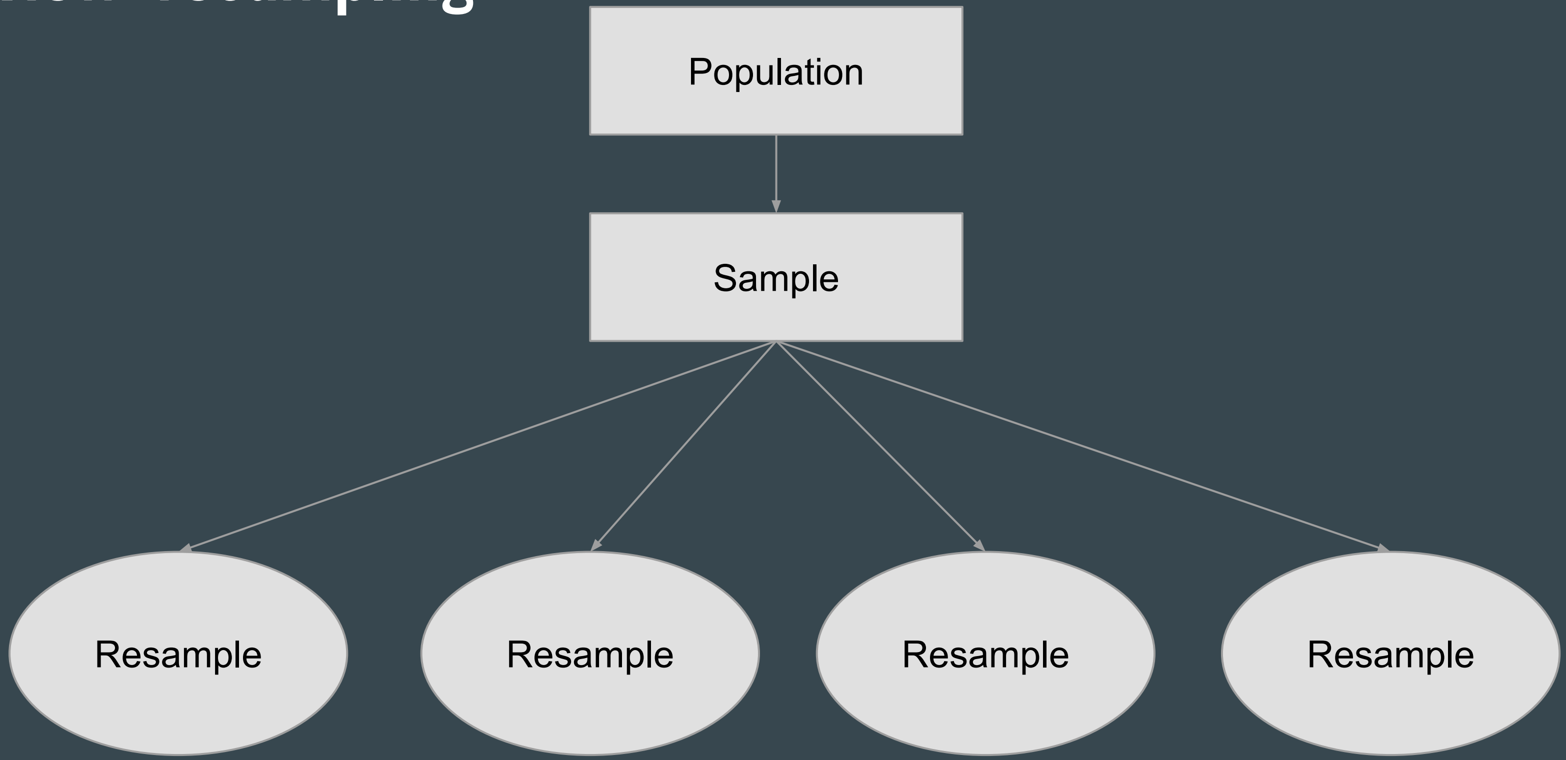
P-values the hard way

27 February 2019
PLSC 309

Review: resampling



Review: resampling



Review: bootstrap

1. Take at least 20 samples with replacement from existing sample
2. Calculate the mean of each of those resamples
3. Subtract the sample mean from the population mean μ to get δ
4. Take the α, 1 - α percentiles of your data for a 1 - α% confidence
    a. 80% confidence: 10th and 90th percentile
    b. 90% confidence: 5th and 95th percentile
    c. 95% confidence: 2.5th and 97.5th percentile
    d. 99% confidence: 0.5th and 99.5th percentile

Review: bootstrap

You are running a campaign. Your candidate currently has an approval rating of 45%. After you announce a new policy, you take a random sample of 100 likely voters.

- Now we can test our hypothesis by creating a distribution of difference in means (mean(X) - μ), δ^*

-0.2	1.2	1.8	2.1
2.5	3.2	3.7	4.1
4.9	5.4	6.2	6.8
7.6	8.1	8.6	9.2
9.9	10.1	10.7	11.2

Parametric Bootstrap

Logic of the bootstrap

- Resampling from our original sample mimics the variation from the population
- In other words, we're deliberately inserting some randomness into our data to account for random chance
  - But this random chance *is on the same scale* as our variable of interest (X)
  - Aka drawn from a similar distribution
- How else could we induce variation that's on the same scale as X?

Parametric Bootstrap

- The regular bootstrap provides us with *confidence intervals* around a given statistic
  - Mean
  - Median
  - Etc.
- The parametric bootstrap gives us confidence intervals around *the parameter(s)* of a distribution

Logic of the parametric bootstrap

- We have a random sample of a variable X that we'd like to model
- We assume that it follows a certain distribution
  - Counts -> Poisson
  - Wait times -> Exponential
  - Etc.
- For our resamples, we take data at random from the distribution that we just assumed

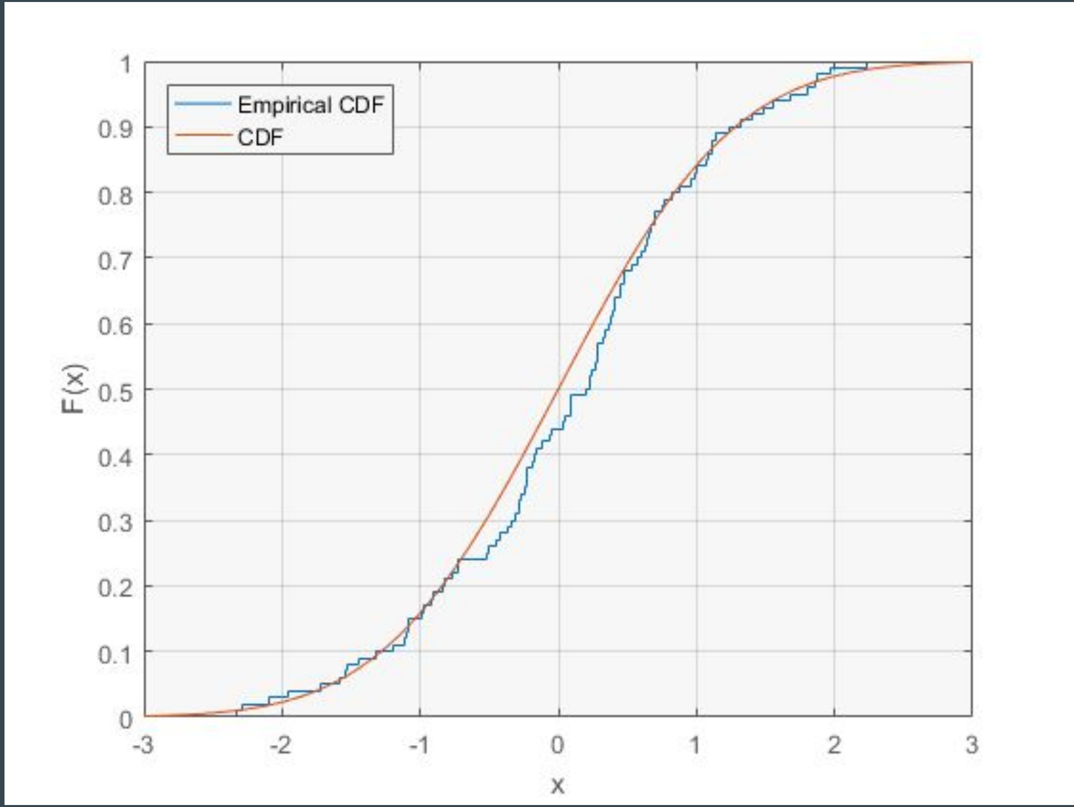
Difference: sampling with replacement vs. random draws

- The regular bootstrap estimates the distribution F from the sample
  - Sample the original sample *with replacement*
- The parametric bootstrap assumes that the distribution of F follows a known distribution
  - Create random draws from a distribution with the same parameter as the original sample

How do we get random numbers from distributions?

- We want to get random numbers that follow a probability distribution
- For example, random draws from a normal distribution with mean 0:
  - We'd expect most of our random draws to be centered around 0
  - We'd expect numbers larger or smaller with decreasing probability as they get further away from 0
- Problem: computers are only able to generate random variables from a uniform distribution

How do we get random numbers from distributions?



From random probabilities to random draws from X

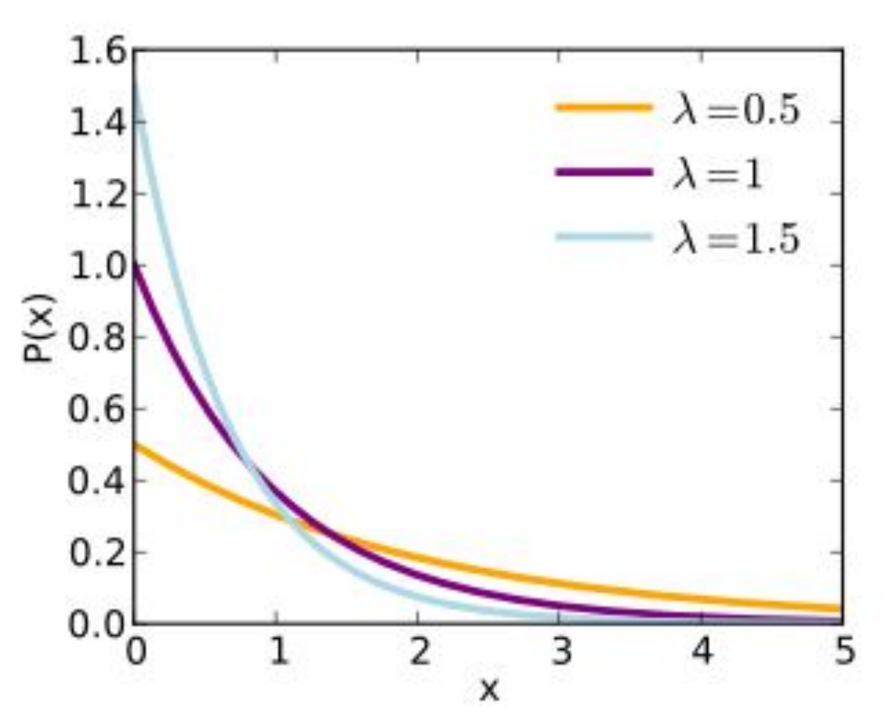
- A CDF is a function that gives us a percentile from 0 to 1

$$F(X) = p(X < x)$$

- A computer can generate percentiles, because they're uniform from 0 to 1
- So to get our X values out of our percentiles we invert the function

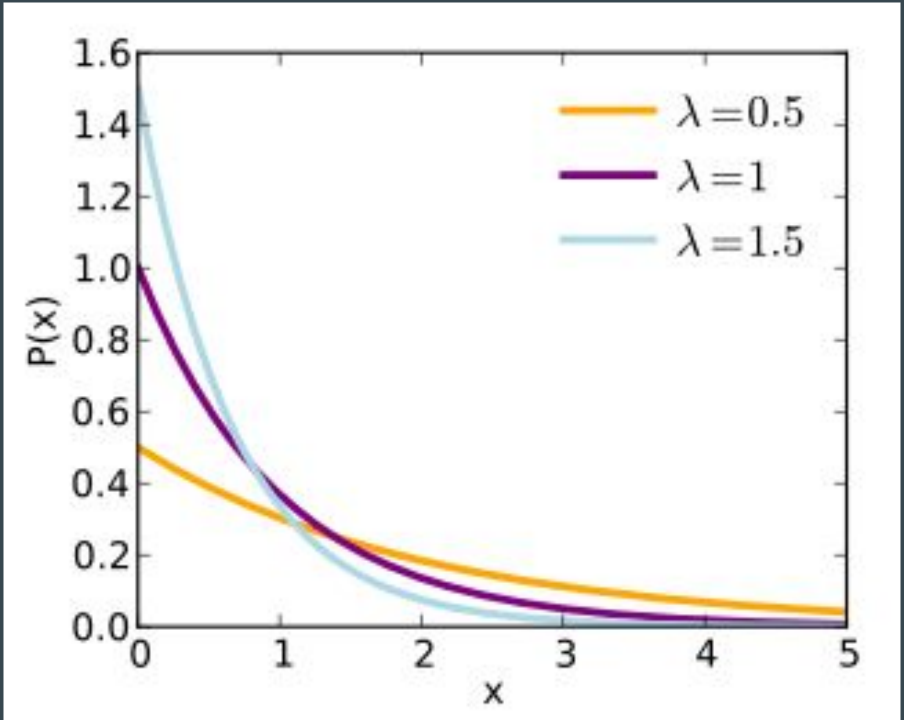
$$F^{-1}(p) = X$$

Random draws from an exponential



$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

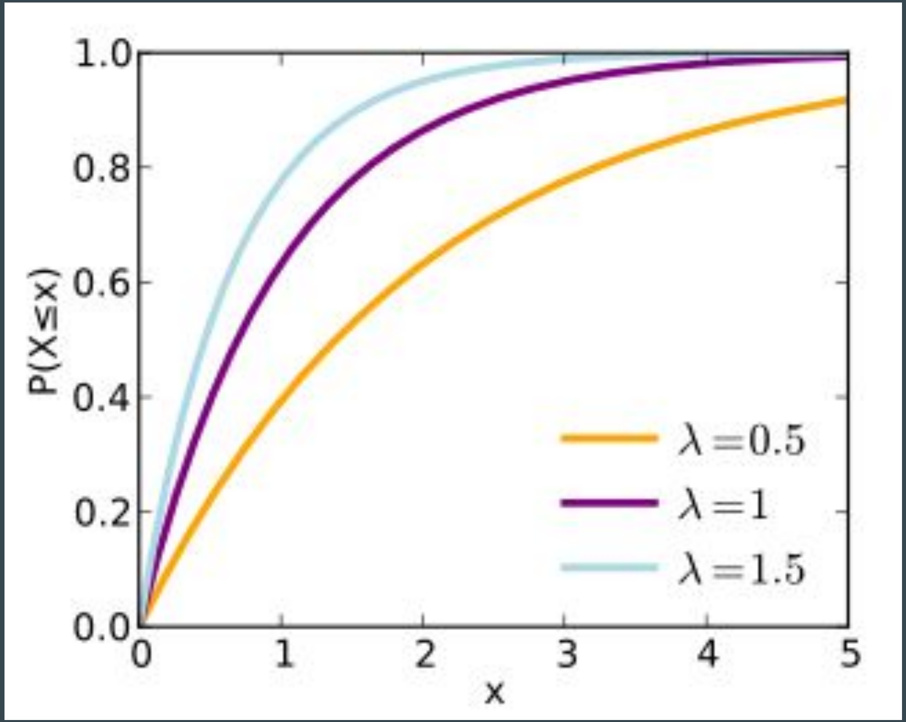
Random draws from an exponential



The goal of the parametric bootstrap is to get a confidence interval for the parameter λ, the "rate of change"

$$\lambda = 1 / \text{mean}(X)$$

CDF allows us to draw random numbers



- $F(\lambda,x) = 1 - e^{-\lambda x}$
- $F^{-1}(\lambda,x) = -1/\lambda * \ln(X)$
- We take a uniform variable U
    - =RAND() in excel
- And replace X for U in the inverse CDF
    - $F^{-1}(\lambda,x) = -1/\lambda * \ln(U)$

Parametric bootstrap cookbook

1. Assume a distribution that your random sample follows
2. Calculate the parameter for that distribution for your sample
   a. E.g. exponential is 1/mean(X); poisson is mean(X),;normal is mean(X) and stdev(X)
3. Using that parameter, create 100 random draws from the distribution the same size as your sample
4. For each resample, calculate your parameter of interest
5. Take the α, 1 - α percentiles of your those parameters for a 1 - α% confidence
   a. 80% confidence: 10th and 90th percentile
   b. 90% confidence: 5th and 95th percentile
   c. 95% confidence: 2.5th and 97.5th percentile
   d. 99% confidence: 0.5th and 99.5th percentile

Parametric bootstrap example

You are interested in measuring the length of time before an election in a parliamentary system. In a random sample of 100 elections, there is on average one every 4.8 years

Parametric bootstrap example

You are interested in measuring the length of time before an election in a parliamentary system. In a random sample of 100 elections, there is on average one every 4.8 years

- Assume your distribution
  a. Time before something happens (continuous) = exponential
- Estimate parameter of interest
  a. Exponential has one parameter λ
  b. $\lambda = 1 / \text{mean}(X)$
  c. For our example, $\lambda = .208$

Parametric bootstrap example

You are interested in measuring the length of time before an election in a parliamentary system. In a random sample of 100 elections, there is on average one every 4.8 years

- Generate 20 random draws from an exponential distribution with $\lambda = .208$
- Calculate the 1 / mean(X) for each random draw

Parametric bootstrap example

You are interested in measuring the length of time before an election in a parliamentary system. In a random sample of 100 elections, there is on average one every 4.8 years

0.164	0.176	0.178	0.182	0.183
0.186	0.194	0.194	0.197	0.202
0.203	0.206	0.209	0.213	0.221
0.229	0.239	0.240	0.249	0.254

Parametric bootstrap example

You are interested in measuring the length of time before an election in a parliamentary system. In a random sample of 100 elections, there is on average onee every 4.8 years

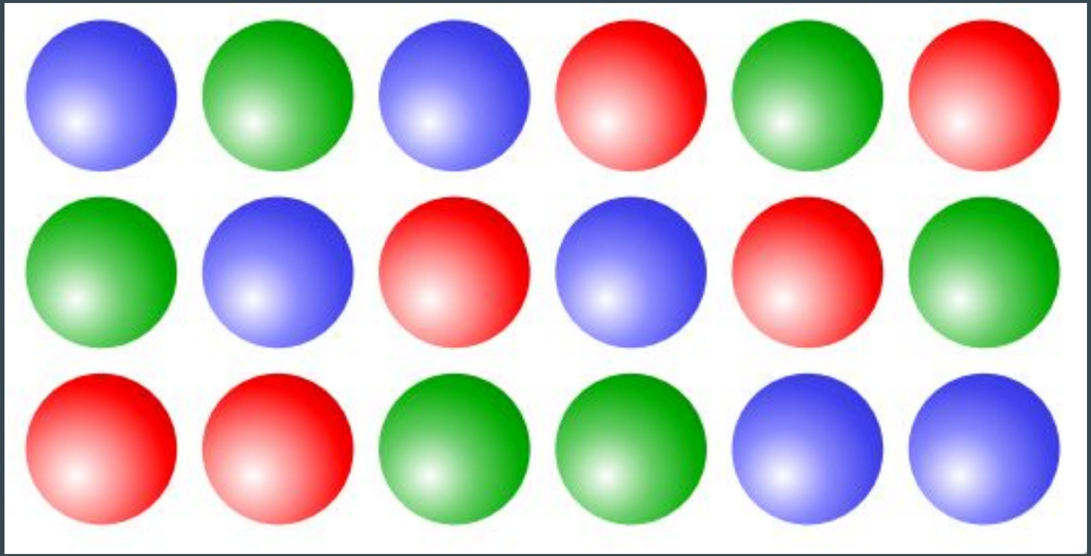
- 90% CI for $\lambda = [0.176, 0.249]$

0.164	0.176	0.178	0.182	0.183
0.186	0.194	0.194	0.197	0.202
0.203	0.206	0.209	0.213	0.221
0.229	0.239	0.240	0.249	0.254

Permutation Test

What is a permutation?

Say you have three marbles: red; blue; and green. Their permutations are all the different order these marbles could be placed in (3! = 3 * 2 * 1 = 6)

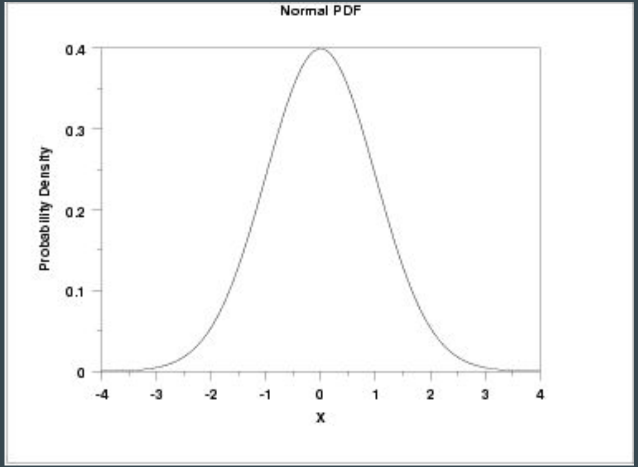


The Bootstrap for two samples?

- The bootstrap helps us model the variation in a sample
- This provides us with confidence intervals for comparing a mean to a population value
- But what about when you want to compare two samples?
- You can’t subtract a confidence interval from another confidence interval

How we solved for two samples in NHST

- The null distribution is the distribution of differences when nothing happens
- For two samples:
  - $H_0 = X_1 - X_2$
- For NHST, we *assumed the null distribution*
- This was based on the central limit theorem:



Are two samples really two distinct variables?

You are a political consultant. You are trying to compare two different ads. You show one to one group of 100 likely voters, and a different ad to another group. Then you measure whether they approve or support your candidate.

Ad 1
Approve
Not approve
Approve
Approve
Not approve

Ad 2
Not approve
Not approve
Approve
Not approve
Approve

Are two samples really two distinct variables?

You are a political consultant. You are trying to compare two different ads. You show one to one group of 100 likely voters, and a different ad to another group. Then you measure whether they approve or support your candidate.

Ad	Approve
1	Yes
1	No
1	Yes
1	Yes
1	No
2	No
2	No
2	Yes
2	No
2	Yes

The Empirical Null Distribution

- The CLT theorem gave us a theoretical null distribution for the difference between samples when that difference is 0
- The theoretical distribution for a two sample test is based on the idea that both samples are drawn from the same population
    - That means X_1 and X_2 follow the same distribution
- In other words, if the null distribution is true then we can combine X_1 and X_2
- If we can combine them, then which group each observation belongs to is irrelevant
- The difference between groups when the group is randomly assigned is the empirical null distribution

Combining X_1 and X_2 by permuting the groups

1. We calculate the observed difference in means between Group 1 and Group 2
2. We then reshuffle our groups b times (usually 1000), each time calculating the difference in means
3. This gives us a distribution of difference in means for our 1000 reshuffled experiments
    a. This is the *empirical null distribution*
4. We see what the probability of our observed value (calculated in step 1) is in relationship to the empirical null
    a. How small the p-value can be depends on how many reshuffles you make (1000 reshuffles can find a .001 p-value, 100 can find a .01 p-value, and so on)

Combining X_1 and X_2 by permuting the groups

- Our observed difference is 1

Ad	Approve
1	Yes
1	No
1	Yes
1	Yes
1	No
2	No
2	No
2	Yes
2	No
2	Yes

Combining X_1 and X_2 by permuting the groups

- Reshuffle our groups
- Now our observed difference is -2
- This is the first point in our empirical null distribution

Ad	Approve
1	Yes
2	No
1	Yes
2	Yes
1	No
2	No
1	No
2	Yes
1	No
2	Yes

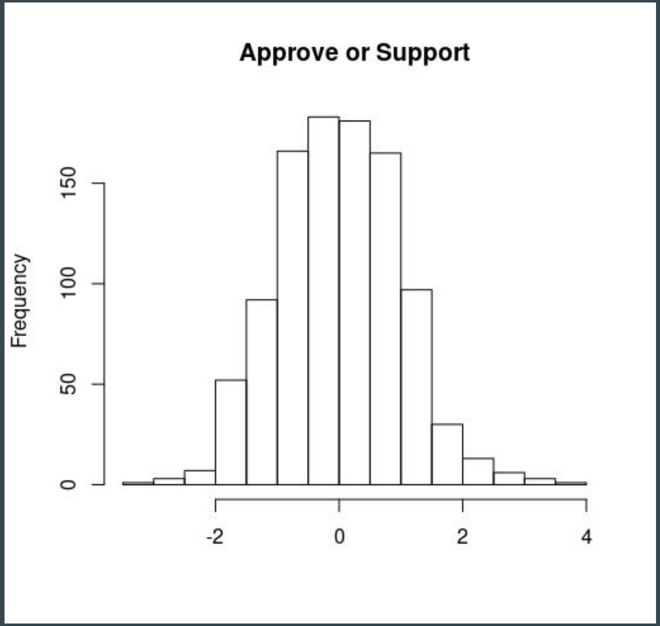
Combining X_1 and X_2 by permuting the groups

- Reshuffle our groups again
- Now our observed difference is -1
- This is the second point in our empirical null distribution

Ad	Approve
1	Yes
1	No
2	Yes
2	Yes
1	No
1	No
2	No
2	Yes
1	No
1	Yes

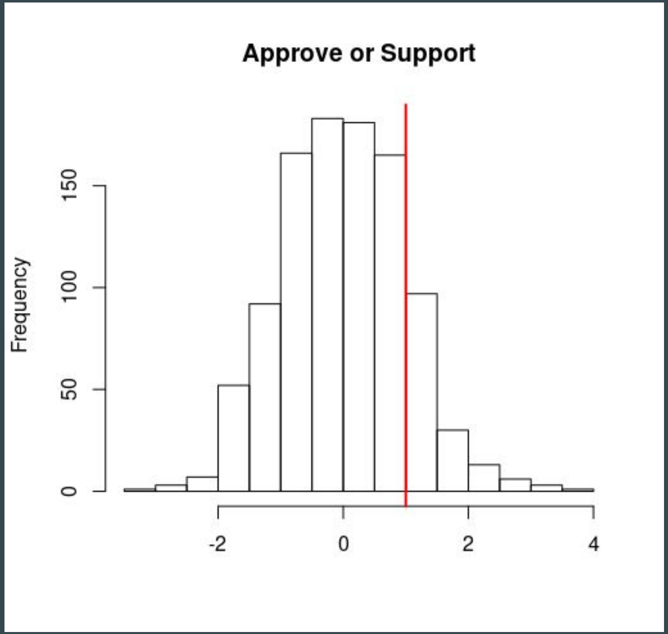
Combining X_1 and X_2 by permuting the groups

- We reshuffle the groups 1000 times, calculating the mean for each reshuffle, giving us our empirical null



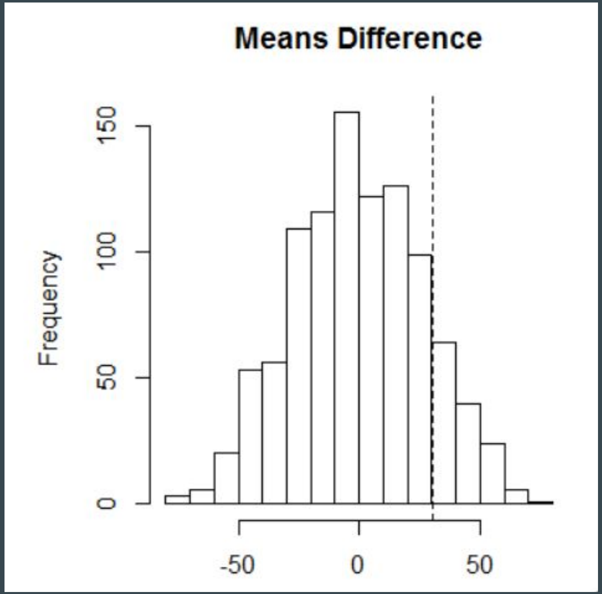
Combining X_1 and X_2 by permuting the groups

- We then calculate our the probability of our observed value by taking its percentile in the empirical null



Permutation test example

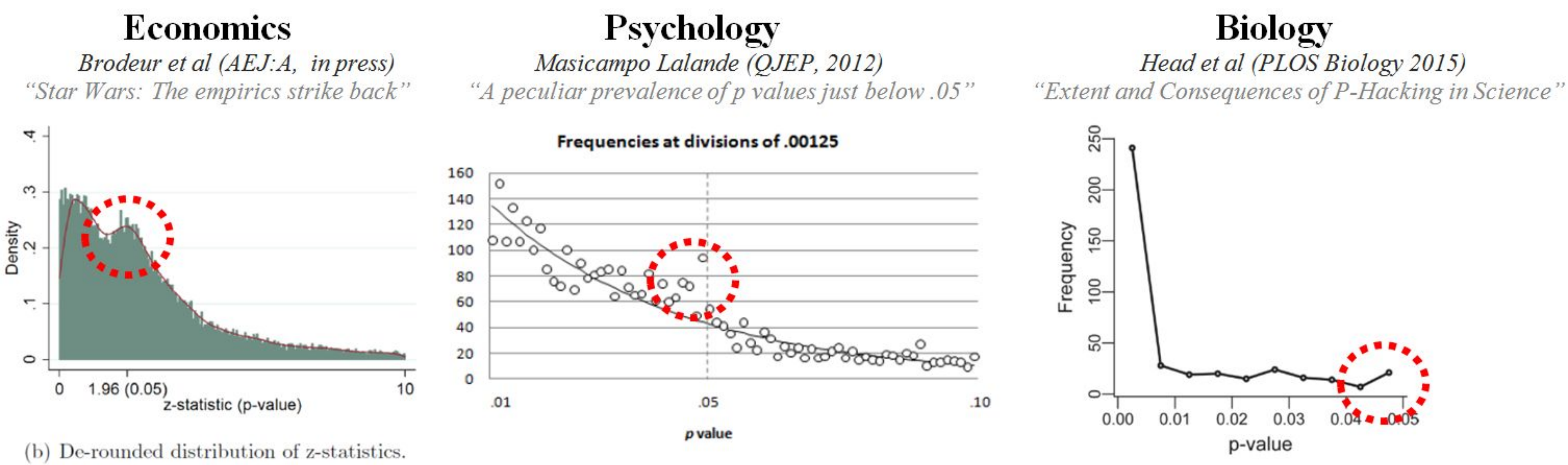
Scientists want to compare the survival times of two groups of mice, one of which was administered an experimental surgery, the other of which didn't. The observed difference in survival times was 30 days.



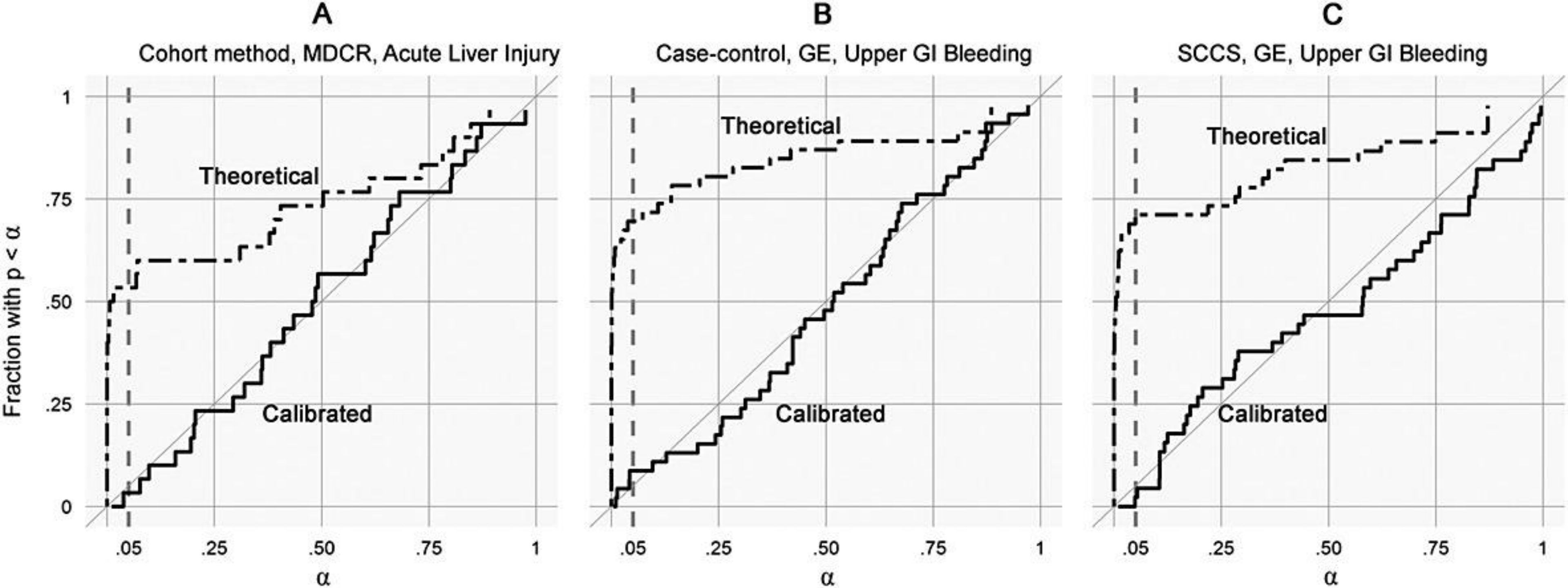
Benefits of the permutation test

- Still has the same conditional probability logic of NHST
- Is a stricter test because it does not require assumptions
- Automatically corrects for small samples
  - Small samples have higher variation, and reshuffling them will increase the variance of the empirical null, making a low p-value less likely

Benefits of the permutation test



Benefits of the permutation test



Review: parametric bootstrap

- In certain cases we can directly assume that our outcome of interest follows a certain distribution
- Instead of taking resamples, we take random draws from a distribution based on the parameter calculated in our sample
- This provides us with confidence intervals around the parameter of the distribution

Review: permutation test

- Instead of assuming a theoretical null distribution we create an empirical null
- The empirical distribution is based on the assumption that both samples are drawn from the same distribution
- We reshuffle the different groups and calculate a difference in means for each one
- We compare our original difference to this empirical null