

POIR 613: Computational Social Science

Pablo Barberá

University of Southern California

`pablobarbera.com`

Course website:

pablobarbera.com/POIR613/

Today

1. Zoom discussion sessions (for now):
 - ▶ September 22nd
 - ▶ October 20th
 - ▶ November 10th
2. Experimental research in the digital age
3. Solutions for last week's challenge
4. Webscraping

Experimental research in the digital age

Experimental research in the digital age

Field experiments combine the control of laboratory experiments ([high internal validity](#)) with the generalizability of a real setting ([external/convergent validity](#)).

Challenge: cost, particularly if scale is sufficient to study high-variance social phenomena.

[Digital technologies](#) offer practical and cost-effective venues for conducting field experiments (*aka* **A/B tests**).

Given sufficient [access](#) and existence of [software](#) that allows randomization, researchers can study both short- and long-term effects of manipulations

How Obama raised \$60 million using experiments

The image shows a screenshot of the Obama '08 'GET INVOLVED' website. The page features the Obama '08 logo at the top, followed by the text 'GET INVOLVED' in large, white, sans-serif capital letters. Below this text is a photograph of Barack Obama smiling, surrounded by a large number of blue campaign buttons, each featuring the Obama '08 logo. At the bottom of the page, there is a white registration form with the following elements:

- A link labeled 'JOIN THE MOVEMENT' on the left.
- An input field labeled 'Email Address'.
- An input field labeled 'Zip Code'.
- A prominent red button labeled 'SIGN UP'.

Two white arrows with labels point to specific elements on the right side of the page:

- An arrow labeled 'Media' points to the Obama '08 campaign buttons.
- An arrow labeled 'Button' points to the red 'SIGN UP' button.

At the very bottom of the page, there is a dark blue footer containing the text 'PAID FOR BY OBAMA FOR AMERICA' on the left, the Obama '08 logo in the center, and the text 'CONTINUE to WEBSITE' on the right.

How Obama raised \$60 million using experiments

IMAGES



VIDEOS



JOIN US NOW

LEARN MORE

SIGN UP NOW

SIGN UP

6 Media variation \times 4 button combinations = 24 combinations

Which one do you think will get a higher conversion rate?

How Obama raised \$60 million using experiments

Combinations (24)		Page Sections (2)		Download: XML CSV TSV Print		
Relevance Rating	Variation	Est. conv. rate	Chance to Beat Orig.	Observed Improvement	Conv./Visitors	
Button <div>5 / 5</div>	Original	7.51% ± 0.2%	—	—	5851 / 77858	
	Learn More	8.91% ± 0.2%	100%	18.6%	6927 / 77729	
	Join Us Now	7.62% ± 0.2%	73.5%	1.37%	5915 / 77644	
	Sign Up Now	7.34% ± 0.2%	13.7%	-2.38%	5660 / 77151	
Media <div>5 / 5</div>	Original	8.54% ± 0.2%	—	—	4425 / 51794	
	Family Image	9.66% ± 0.2%	100%	13.1%	4996 / 51696	
	Change Image	8.87% ± 0.2%	92.2%	3.85%	4595 / 51790	
	Barack's Video	7.76% ± 0.2%	0.04%	-9.14%	3992 / 51427	
	Sam's Video	6.29% ± 0.2%	0.00%	-26.4%	3261 / 51864	
	Springfield Video	5.95% ± 0.2%	0.00%	-30.3%	3084 / 51811	

Outcome variable: sign-up rates

Dashboard shows sign-up rates for each separate variation

How Obama raised \$60 million using experiments

Combinations (24)		Page Sections (2)		Download: XML CSV TSV Print		
Disable	All Combinations (24) ▼		Key: Winner Inconclusive Loser ?			
<input type="checkbox"/> Combination	Status ?	Est. conv. rate ?	Chance to Beat Orig. ?	Observed Improvement ?	Conv./Visitors ?	
Original	Enabled	8.26% ± 0.5%	—	—	1088 / 13167	
★ Top high-confidence winners. Run a follow-up experiment »						
<input type="checkbox"/> Combination 11	Enabled	11.6% ± 0.6%	100%	40.6%	1504 / 12947	
<input type="checkbox"/> Combination 7	Enabled	10.3% ± 0.6%	100%	24.0%	1340 / 13073	
<input type="checkbox"/> Combination 3	Enabled	9.80% ± 0.6%	99.7%	18.7%	1277 / 13025	
<input type="checkbox"/> Combination 10	Enabled	9.23% ± 0.6%	95.9%	11.7%	1203 / 13031	
<input type="checkbox"/> Combination 8	Enabled	9.03% ± 0.6%	91.6%	9.28%	1178 / 13046	
<input type="checkbox"/> Combination 9	Enabled	8.77% ± 0.6%	81.8%	6.10%	1111 / 12672	
<input type="checkbox"/> Combination 6	Enabled	8.64% ± 0.5%	75.3%	4.58%	1108 / 12822	

Dashboard shows sign-up rates for each separate variation

The winner



The image shows a screenshot of the Obama 2008 campaign website's splash page. At the top is the Obama '08 logo. Below it, the word "CHANGE" is written in large blue letters, with "WE CAN BELIEVE IN" in smaller blue letters underneath. A black and white photograph of Barack and Michelle Obama with their children is centered on the page. At the bottom, there is a registration form with the text "JOIN THE MOVEMENT" on the left. The form includes two input fields: "Email Address" and "Zip Code". To the right of these fields is a red button labeled "LEARN MORE". At the very bottom of the page, there is a blue bar containing the text "PAID FOR BY OBAMA FOR AMERICA" on the left and "CONTINUE ~ WEBSITE" on the right.

Original sign-up rate: 8.26%

New sign-up rate: 11.6%

Change: +40.6 lift in sign-up rate

10MM people signed-up through splash page during campaign

Without experiment, number would have been 7.2MM

That's 2.8MM fewer email addresses

Average donation per email address is \$21

2.8MM x \$21 = \$60MM !!!

Experimental research in the digital age

Experimental technologies for online interventions:

1. Email and text messages

- ▶ More likely to get subjects' attention
- ▶ e.g. [Blair et al \(2017\)](#): randomized text messages in India to encourage people to report corruption

2. Modified web interface

- ▶ Manipulation: platform features, exposure to information, display of specific web elements, etc.
- ▶ e.g. [Bakshy et al \(2012\)](#): social cues on FB ads

3. Bots

- ▶ Program or script that makes automated requests
- ▶ e.g. [Munger \(2016\)](#): reducing harassment on Twitter

4. Add-ons

- ▶ Additional software that nudges or tracks subjects
- ▶ e.g. [Guess \(2016\)](#): web tracking software to observe individuals' news consumption in response to monetary encouragement to seek information

Experimental research in the digital age

What can go wrong? (And potential solutions)

1. Logging errors: covariate balance in pre-treatment variables, A/A tests
2. Novelty effects: longer experiments
3. Multiple testing: Bonferroni correction
4. High significance due to large sample sizes: Cohen's D
5. SUTVA (interference between units): better research design
6. The 'free beer' problem: social science theory!

Side note: power calculations

- ▶ **Power** is the probability of detecting a specified *effect size* with specified sample characteristics (*size* and *variability*)
- ▶ Four interrelated components:
 1. Sample size
 2. Effect size you want to detect
 3. Desired significance level (false positive rate you're fine with)
 4. Power
- ▶ Before you run an experiment, you can compute necessary sample size assuming other 3 components:

```
> power.prop.test (p1=0.30, p2=0.35,  
sig.level=0.05, power=0.80)
```