Testing ad content with survey experiments

Patrick Miller
pmiller@civisanalytics.com
@patr1ck_mil
Testing ad content with a Survey Experiment

1. We find a representative sample using our state-of-the-art sampling and weighting correction methods.

2. All respondents were asked demographic and other relevant Pre-Screen questions.

3. Respondents were then randomly split into treatment and control groups. Those in a treatment group were shown a single creative.

4. Respondents were then asked Post-Message questions to measure message effectiveness on key metrics.
Examples

Tests we learned concrete things from
## Overall Treatment Effects

### Brand Consideration

<table>
<thead>
<tr>
<th>Average Treatment Effects</th>
<th>Best Message Probability</th>
<th>Backlash Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3%</td>
<td>8%</td>
<td>92%</td>
</tr>
</tbody>
</table>
Treatment Effects by Gender

GENDER

BRAND CONSIDERATION

FEMALE

MALE

30.0% 40.0% 50.0%

-5%

-2%

DOVE
Tide

LIKE BIKINIS, THEY SAVE MEN A LOT OF GUESSWORK.

DON’T PAY FOR WATER, PAY FOR CLEAN

15% CLEANING INGREDIENTS*

90% CLEANING INGREDIENTS

*leading detergent—detergent, base variant vs. Tide PODS® packs.
Like all household detergents, wash away from children.
## Overall Treatment Effects

### Brand Favorability

<table>
<thead>
<tr>
<th>Average Treatment Effects</th>
<th>Best Message Probability</th>
<th>Backlash Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10%</td>
<td>98%</td>
<td>0%</td>
</tr>
<tr>
<td>+2%</td>
<td>2%</td>
<td>24%</td>
</tr>
<tr>
<td>-10%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Nike
Overall Treatment Effects

Brand Consideration

<table>
<thead>
<tr>
<th>Average Treatment Effects</th>
<th>Best Message Probability</th>
<th>Backlash Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4%</td>
<td>76%</td>
<td>24%</td>
</tr>
<tr>
<td>-5%</td>
<td>2%</td>
<td>93%</td>
</tr>
</tbody>
</table>
Treatment effects by Ideology

Conservatives had the most backlash (-7%, -5%) and had the lowest consideration (38%) while Liberals showed no backlash at all and had the highest consideration (76%)
Meta analysis
Most ads are ineffective, but testing improves efficiency

Some ads are definitely ineffective
  ● 11% of ads have backlash

A lot of ads are probably ineffective
  ● 26% of ads have a treatment effect < 1pp
  ● 43% of ads have treatment effects not conclusively different from 0pp

Testing multiple ads improves efficiency overall
  ● The best ad was 13% better on average than the worst ad in the experiment
Summary
Implementation
What we learned the hard way
Overview

Steps
1. Data collection
2. Survey weighting
3. Modeling
4. Reporting

Goals
1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable
1. Data Collection
It's small data, and measurement matters

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable
2. Survey Weighting
Your sample is biased, correct it with weighting

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable
3. Modeling
Keep it simple with a parametric model

“It's just logistic regression”

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable

$$\text{glm}(y \sim tx \ast age + tx \ast female, \text{family} = \text{'binomial'})$$
... and make it a service

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable

https://devrant.com/rants/1854993/package-tsunami
4. Reporting
Overall Treatment Effects

Brand Consideration

<table>
<thead>
<tr>
<th>Average Treatment Effects</th>
<th>Best Message Probability</th>
<th>Backlash Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4%</td>
<td>76%</td>
<td>24%</td>
</tr>
<tr>
<td>-5%</td>
<td>2%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Goals

1. Accurate
2. **Interpretable**
3. Trustworthy
4. Reusable
Baselines

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable
Weighted Marginal Treatment Effects

Goals

1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable

Average Treatment Effect (ATE) is the incremental gain over the control group.

An ATE can be positive or negative. Backlash is a negative reaction to a piece of creative.
Uncertainty

Goals
1. Accurate
2. Interpretable
3. Trustworthy
4. Reusable
Testing ad content with survey experiments

Answer questions about ad effectiveness unambiguously, but testing allows your company to learn which ones are effective.

Avoid bad ads that cause twitter/internet firestorms.

For implementation prioritize trustworthiness and interpretability; make the model reusable by deploying as a service.
Testing ad content with survey experiments

Patrick Miller
pmiller@civisanalytics.com
@patr1ck_mil