Date: Tuesday, April 30, 2019

[Case study question]

Creative type: {logo, brand name, none}

Ad-format: {static, cinematic, autoplay}

QUESTION: How do creative type and ad format affect Pinterest's platform?

STRATEGY: Full factorial design
↳ To analyze the causal effects of creative type and ad format on user experience on Pinterest.

Reasons:
- We want casual relationships (not strictly an optimization problem).
- We don't want to run separate experiments because these don't account for interactions of creative type and ad format.

1) **Design:** Create 9 versions of the site using different combination of each pair.

   Example of Condition: {logo, static},
   {logo, cinematic}, etc.

2) Allocate say 100/9 % of the users to each version of the site.

3) Calculate sample size needed to reach a power of say 0.8 for an effect size of δ effect.

   ↳ gives a plot of power across sample size for differing values of δ.

4) Run experiment to calculate a metric/KPI for Pinterest for each version.

↳ Make sure sample size is reached before stopping.

5) Analysis: Run regression of metric on condition dummy variables and pairwise interaction effects.

↳ Linear if mean of metric.
↳ Logistic if proportion.

Suppose we have cts metrics and run linear regression.

Then we can test for main effects of each factor (creative type & ad format) and interaction effects using partial F-tests.

↳ This analysis allows us to answer the question.

Metrics: Click-thru rate of pin
session duration/Time on website

Guardrail metrics — daily active users, $\frac{1}{\text{load-time}}$
(things that should not decrease)

↳ could view this as metrics that do not increase/decrease.

Example: Suppose we're testing for effects on CTR between version 1 or 2 of the site.

P_1 = True CTR of version 1

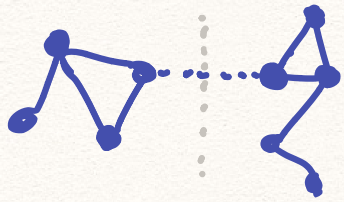
P_2 = True CTR of version 2

$H_0: \delta = 0$ vs $H_1: \delta \neq 0$

δ = treatment effect
$= P_1 - P_2$

z-test
↳ we apply the above test to all pairs of conditions and apply B-H to control false-discovery rate.

* Nice visualizations for graphs → Gephi



Aim: To identify the partition of these nodes

Strategy:

1) $Observed_j$: # of edges in C_j

2) $Expected_j$ = Expected # of edges in C_j if the edges of the graph were permuted at random.

3) $$\text{Modularity} = \frac{1}{\#\text{edges}} \sum_{j=1}^{2} (Obs_j - Exp_j)$$

↑ modularity = strong clustering

↓ modularity : no community structure.

* High values of modularity

  ⇒ we see a lot more edges in each community than we expect.

  ↳ Strong clustering.

The aim of identifying the best partition becomes identifying the partition w/ the highest modularity.