General Notes

Darian S. Martos

AB Testing Notes

These are some preliminary notes for A/B testing, using the book as the main reference below. Some notes from a previous stats course taught by Dr. Art Owen, that are widely available, were also referenced for these notes. The chapter ordering for these notes is quite scattered, mostly due to me finding certain sections more fruitful than others.

Much of these notes are organized in a definition-based format. Examples are scarce unless I find them valuable.

Resources used (these are hyperlinks):

- Trustworthy Online Controlled Experiments [Kohavi, Tang, Xu]
- AB Testing Notes from Stanford Stats 263/363 [Owen]

Some baseline abbreviations:

- Controlled experiments - CEs
- Overall Evaluation Criterion - OEC (see Ch. 2.1)

Chapter 1. Preliminaries from the Owen Notes

TBD - TBD

Chapter 2. TOCE - Introduction and Motivation

Overall Evaluation Criterion (OEC) - Quantitative measure of an experiment's objective

1 Terms and Definitions

- Example - active days per user
- An ↑ in OEC implies an ↑ in site visits ⟶ good
- OEC should be short-term measurable, believed to causally drive long-term strategy

Parameter - Controllable experiment variable thought to influence OEC or other metrics, also known as factors or variables.

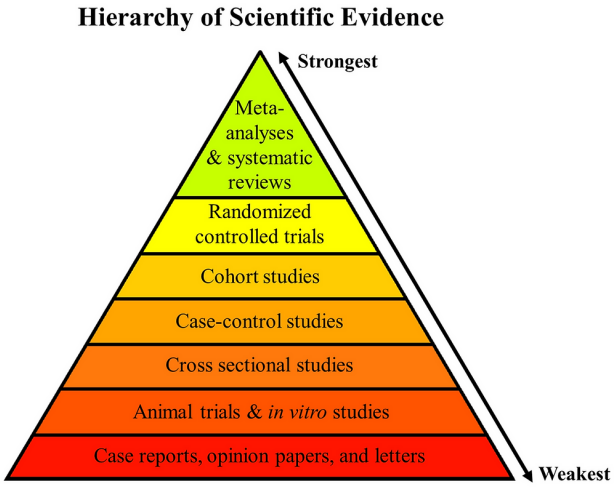
- The assigned values are known as *levels*.
- In simple A/B tests - univariate with two values
- Multivariate tests evaluate parameters together

Variant - User experience being tested, A and B are two variants - usually a control and treatment.

Randomization unit - Pseudo-random process that is applied to units to map to variants. The recommended set of units is usually users.

2 Why experiment?

- Don't want to assume causality off observation alone
- Want to follow the hierarchy of evidence in terms of causal strength:



- Online CEs can establish causality with high probability, are able to detect small changes that are harder to detect with other techniques, such as changes over time (sensitivity), and are able to detect unexpected changes

3 Necessary Ingredients for Experiments

- There are experimental units that can be assigned to different variants with no interference.
- There are enough experimental units (ideally thousands).
- Key metrics, ideally an OEC, that are agreed upon and can be practically evaluated. Data should be reliable and can be easily collected.
- Changes should be easy to make, this is context dependent. For example - recommender systems are easier to adjust versus flight software.

4 Tenets for Organizations to Run Online CEs

- The organization wants to make data-driven decisions and has formalized an OEC.
- The organization is willing to invest in the infrastructure and tests to run CEs and ensure that the results are trustworthy.
- The organization recognizes that it is poor at assessing the value of ideas.