Is Your Idea Worth Testing?
A/B Testing Theory for Models and Verizon Use Cases
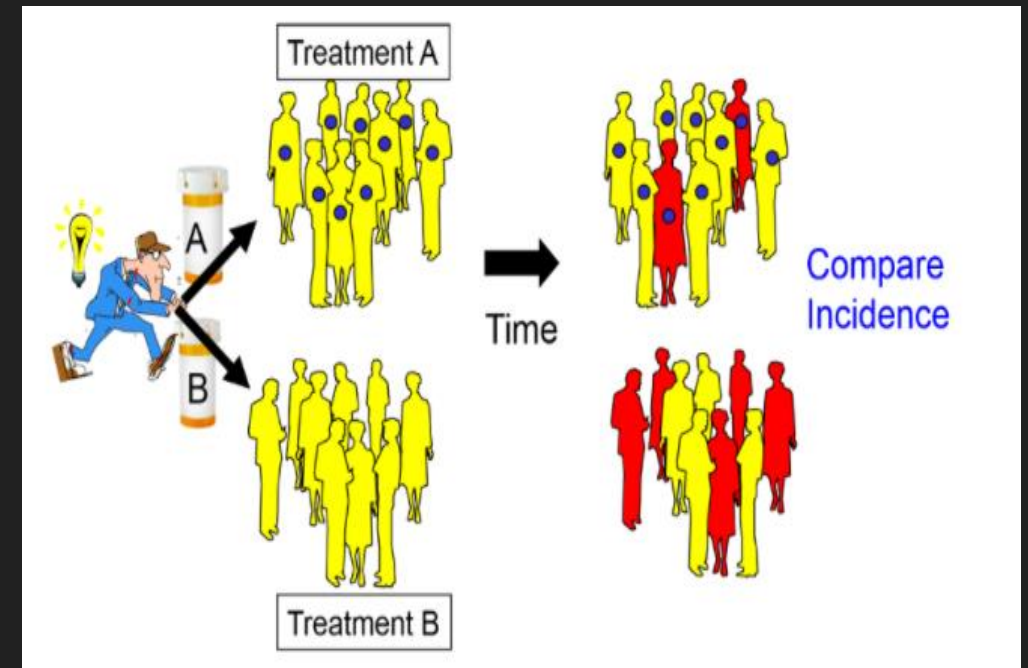
Sanjit Paliwal

8th November, 2021

Agenda

- History
- Benefits
- Methodology
- Terminology
- Verizon Use Cases
- Limitations
- References

History

- In the 1920s statistician Ronald Fisher ran agricultural experiments, asking questions such as, What happens if I put more fertilizer on this land? [1]

- The principles persisted and in the early 1950s scientists started running clinical trials in medicine.

- In the 1960s and 1970s the concept was adapted by marketers to evaluate direct response campaigns (e.g., would a postcard or a letter to target customers result in more sales?).

- A/B testing, in its current form, came into existence in the 1990s. Greg Linden at Amazon created a prototype to show personalized recommendations based on items in the shopping cart. [2]

Source: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module4-Cohort-RCT

1 : Amy Gallo. "A Refresher on A/B Testing" Harvard Business Review. https://hbr.org/2017/06/a-refresher-on-ab-testing
2. Linden, Greg. Early Amazon: Shopping cart recommendations. Geeking with Greg. [Online] April 25, 2006.
http://glinden.blogspot.com/2006/04/early-amazon-shoppingcart.html.

3

Benefits

- **Quickly evaluate ideas** : The web provides an unprecedented opportunity to evaluate ideas quickly using A/B tests also called randomized controlled experiments.

- **Causal Relationship** : Controlled experiments embody the best scientific design for establishing a causal relationship between changes and their influence on user-observable behavior.[3]

- **Know your Audience** : A/B testing provides a framework to learn more about your audience and make changes so that you're reaching them in the most effective way.[4]

- **Accelerate Innovation** : When a company builds a system for experimentation, the cost of testing and experimental failure becomes small, thus encouraging innovation through experimentation.

3. Kohavi, R., Henne, R.M., & Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. *KDD '07*.

4: Jon Simpson . "A/B Testing: The Benefits And How To Use It Efficiently". https://www.forbes.com/sites/forbesagencycouncil/2020/03/12/ab-testing-the-benefits-and-how-to-use-it-efficiently/?sh=198ce040786d
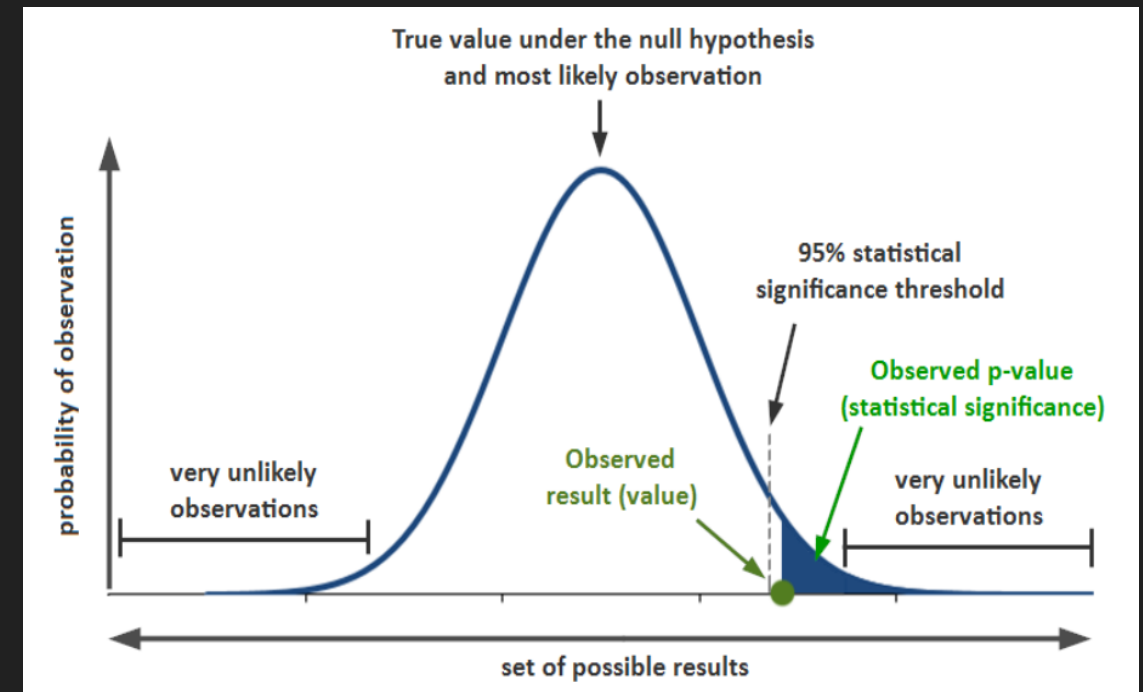
Methodology

- A/B testing, at its most basic, is a way to compare two versions of something to figure out which performs better.

- In the simplest manifestation of such experiments, live users are randomly assigned to one of two variants:
    i.   the Control, which is commonly the "existing" version, and
    ii. the Treatment, which is usually a new version being evaluated.

- Metrics of interest, ranging from runtime performance to implicit and explicit user behaviors and survey data, are collected.

- Statistical tests are then conducted on the collected data to evaluate whether there is a statistically significant difference between the two variants on metrics of interest.
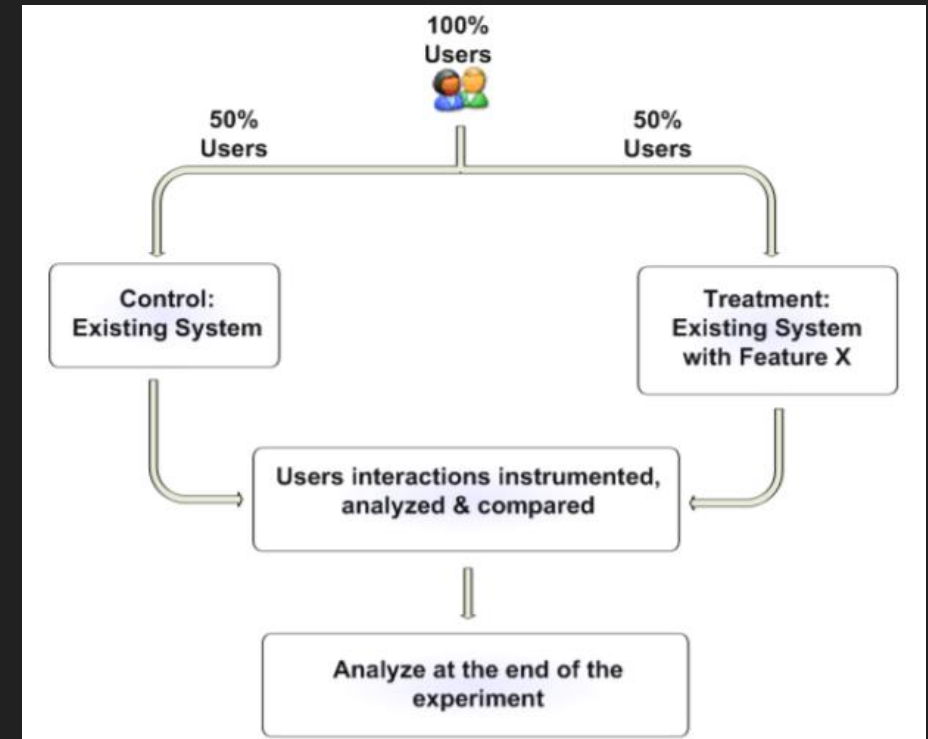
Source: Kohavi, R., Henne, R.M., & Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. *KDD '07.*

Terminology

- **Overall Evaluation Criterion (OEC)** : A quantitative measure of the experiment's objective.

- **F**actor: A controllable experimental variable that is thought to influence the OEC.

- **Experimentation Unit** : The entity on which observations are made. The units are assumed to be independent.

- **Null Hypothesis**: The hypothesis, that the OECs for the variants are not different and that any observed differences during the experiment are due to random fluctuations.

- **Confidence leve**l : Commonly set to 95%, this level implies that 5% of the time we will incorrectly conclude that there is a difference when there is none (Type I error).

- **Power** : The probability of correctly rejecting the null hypothesis, when it is false. Power measures our ability to detect a difference when it indeed exists.

Source: https://blog.analytics-toolkit.com/2017/statistical-significance-ab-testing-complete-guide/

Limitations

- **Quantitative Metrics, but No Explanations :** It is possible to know which variant is better, and by how much, but not "why".

- **Short term vs. Long Term Effects :** Controlled experiments measure the effect on the OEC during the experimentation period, typically a few weeks.

- **Primacy and Newness Effects :** When a new design or feature is introduced, some users will investigate it, click everywhere, and thus introduce a "newness" bias.

References

1. Amy Gallo. "A Refresher on A/B Testing" Harvard Business Review. https://hbr.org/2017/06/a-refresher-on-ab-testing.

2. Linden, Greg. Early Amazon: Shopping cart recommendations. Geeking with Greg. [Online] April 25, 2006. http://glinden.blogspot.com/2006/04/early-amazon-shoppingcart.html.

3. Kohavi, R., Henne, R.M., & Sommerfield, D. (2007). Practical guide to controlled experiments on the web: listen to your customers not to the hippo. *KDD '07*.

4. Jon Simpson . "A/B Testing: The Benefits And How To Use It Efficiently". https://www.forbes.com/sites/forbesagencycouncil/2020/03/12/ab-testing-the-benefits-and-how-to-use-it-efficiently/?sh=198ce040786d

5. A/B Testing by Google course on Udacity. https://www.udacity.com/course/ab-testing--ud257

Q & A