Date: Tuesday, May 7, 2019

ANNOUNCEMENTS:

- HW #4 due Friday @ midnight
- Quiz #4 on Thursday morning
  - → Will cover the pitfalls & challenges of A/B testing (2 slide decks + jupyter notebook)
  - → Multiple choice
- Final Exam
  - → Due Tuesday, May 14th @ 5 p.m.
  - → Take home
  - → Cumulative
  - → Multiple choice
  - → Will be up on Thursday
- Thursday : 3:15 p.m – 5:00 p.m.
  Make-up class
  1st floor rooms

Example:

TinyCO → we use device as our unit

- Design an A/B test on device
- User information gives us an idea of what user has what device, but it can be incorrect 5% of the time.
- TinyCo makes inference @ the user level instead of the device level.

* TinyCO measures
  - $\bar{X}_A$ → Sample mean for group A
  - $\bar{X}_B$ → Sample mean for group B
  - Effect size: $\hat{\delta} = \bar{X}_A - \bar{X}_B$

→ we have to account for the error ID rate through the values:

- π_{AB} = # of users thought to be in group A but actually in group B
- π_{BA} = Analogous...

→ The power calculations depend on π_{AB} & π_{BA} and are found in the paper "The Power of an A/B test under interference".

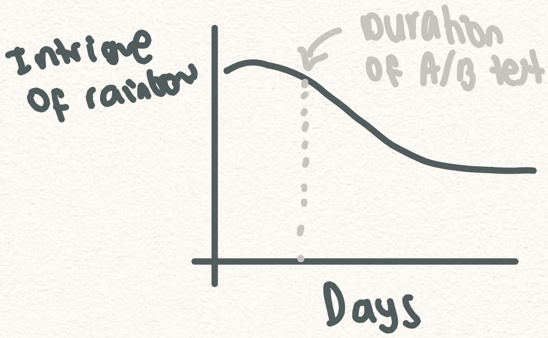
Note:
- Companies have a "customer value" metric which is often a function of feature metrics that can be measured.
- The main goal is always to optimize these "customer value" metrics.

Example: FastClick tests the effect of adding a rainbow to its front page.

- They run the test w/ session duration as the metric.

- At the beginning of the experiment users are really intrigued / excited by the rainbow.

* Mental image:



The intrigue of the rainbow is dictating the measured metric.

Question: How can we gauge the lift effect from new features?

Estimate expected lift based on lift from the past.

- You need to be aware of lift and make sure you run the experiment long enough to see true effect.

One vs Two-tailed tests for effect sizes

A vs. B

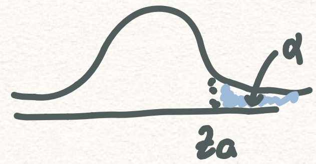
$H_0: \mu_A = \mu_B$ vs. $H_A: \mu_A \neq \mu_B$ } Two-sided

$H_A: \mu_A > \mu_B$
$H_A: \mu_A < \mu_B$ } One-sized (Have more power)

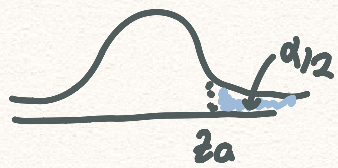
Power: (z- test)

$z_\alpha: \ P(z > z_\alpha) = \alpha$

* One-sided: $1-\beta = P\left(z > z_\alpha - \frac{\delta}{\sigma/\sqrt{n}}\right)$



* Two-sided: $1-\beta = P\left(z > z_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{n}}\right)$



$z_{\alpha/2} > z_\alpha$: This leads to power being greater for one-sized tests for all things $(\alpha, n, \delta, \sigma)$ held constant.