Avery Hoffman, Samantha Bobman, Sheil Sarda, Sinan Onukar, and Yi Liu

OIDD 399 Pre-Simulation Strategy

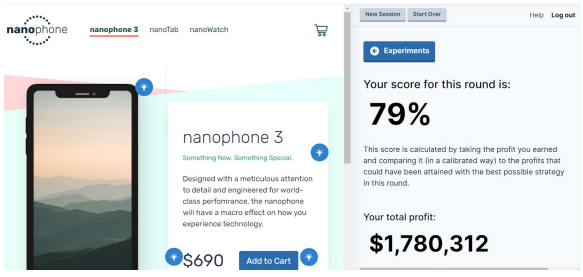
Getting Baseline values for different parameters

- 20 separate experiments (4 groups x 5 features): distinguish b/w US vs International, Desktop v Mobile
    - Why? There is no interaction between features or groups
- Start from an even split between variants
- Kick out the variants that are significantly worse than one of the others ($p<0.05$)
- For those not significant variants, assign traffic based on the conversion rate
    - E.g., Conversion rate A = 3%, Conversion rate B = 4 %, p-value = 0.30
    - Then we assign 30% traffic to A and 70% to B
- Do this iteration for 6 weeks

Tuning parameters

- P-value / Significance test
    - Experimentally, we found that Week 7 onward, the weights we assigned to different variants started stabilizing, allowing us to maintain the same weights for most tests
    - Once we hit this inflection point, we increased the P-value for the significance test from 0.05 to 0.1, and adjusted the weights of different variants in proportion to their significance
- Difference between conversion rates of each option
  If the conversion rate of 2 variants which are equally significant is not very different (e.g. within a 0.1% threshold), we hold-off on adjusting the weights assigned to them in the A/B test

Experimental Results



Concerns

- Our biggest concern with this strategy is that we may run out of time.
    - However, we will mitigate this concern by expanding our P-value for the significance test to 0.1