P7- Design an A/B Test

Metric Choice

Invariant Metrics:

Number of cookies- The number of cookies in each group should be about the same and unaffected by the content of the experiment.

Number of clicks- Because the experiment is only visible after clicking, the number of clicks should also be similar across groups

Click-through-probability- The ratio of the above two metrics should also be very close between groups.

Evaluation Metrics:

Gross conversion- The experiment could change the percent of users who enroll after clicking (and therefore seeing the prompt asking how many hours they are willing to commit). Hopefully this would result in fewer students signing up as they are screened out.

Net Conversion- This factor is the primary indirect metric for revenue. A decrease in this value is undesirable, as revenue would decrease.

I didn't use Retention as an evaluation metric because it would have required too many initial pageviews to get significant results.

I didn't use number of user ids as an evaluation metric because as it does not have a denominator, it is unlikely that it would follow a normal distribution and it would be hard to do significance tests or analytically calculate variability for it.

In order to launch the experiment, I would like to see a decrease in gross conversion and no change or a positive change in net conversion.

Measuring Standard Deviation

Gross conversion – 0.0202

Net conversion – 0.0156

I expect both of these metrics, being probabilities, would follow a binomial distribution. With a big enough sample size, I the distribution would likely be normal. And because the unit of analysis matches the unit of diversion (cookies), the assumption of independent samples is likely to hold true. Therefore, the analytic estimate of variability is likely to be pretty close to an empirical estimate.

Sizing

Number of Samples vs. Power

Not using Bonferroni Correction (see below for reasoning).

I will therefore need 679300 pageviews through both groups.

Duration vs. Exposure

Fraction diverted: 100%

Days needed: 17

Possibly, the dialogue could turn away large numbers of users who could have succeeded in the course. But it appears as if Udacity already recommends that students are able to allocate 5 hours/week, and simply making it more visible. Because of this and because the project guidelines suggest using more than “a few weeks” for the experiment is too much, I chose to divert all of the traffic.

The experiment is not particularly risky as it does not expose sensitive user data or dramatically change the user experience.

Experiment Analysis

Sanity Checks

Number of cookies: [0.4988-0.5012]; measured = 0.5006; passes sanity check

Number of clicks on “Start free trial”: [0.4959-0.5041]; measured = 0.5005; passes sanity check

Click-through-probability on “Start free trial”: difference [-0.0013-0.0013]; measured difference=0.001; passes sanity check

Result Analysis

Effect Size Tests

Gross conversion: [(-0.0291)-(-0.0120)]; statistically and practically significant.

Net conversion: [-0.0116-0.0019]; neither statistically nor practically significant

Sign Tests

Gross conversion: p-value = 0.0026; significant

Net conversion: p-value = 0.6776; not significant.

Summary

I did not use the Bonferroni correction because I am evaluating the experiment such that I require both metrics to have meet the conditions rather than simply one or the other.

Recommendation

Gross conversion did decrease as expected from the experiment. The results for net conversion were less easily interpretable. Within the 95% confidence interval, however, Net Conversion could decrease at a practically significant level. This means that Udacity's revenue could potentially decrease, which violates the goals of the experiment. Because of the results for net conversion, I would recommend running this experiment again, possibly for a longer time period, to further examine the potential decrease in Net Conversion.

Follow-Up Experiment

If given the chance, I would run an identical version of this experiment. I would still use Net and Gross Conversion as my evaluation metrics and number of clicks, number of cookies, and click-through-probability as my invariant metrics. I would also still use cookies as my unit of diversion. However, I would take out the sentences in the alert box that says "most students commit 5+ hours a week". I expect that many people just choose 5 hours to get past the dialogue to checkout- eliminating this "nudge" seems unlikely to actually change course commitment but likely to get a better assessment of how much a student is willing to commit. I hypothesize that gross conversion would decrease more than this experiment showed without decreasing net conversion.