Ana Javed

P7. Design an A/B Test

Experiment Design

Metric Choice

The invariant metrics chosen were number of cookies, number of clicks, and click through probability. This is because invariant metrics demonstrate what is happening before the experiment takes place, and they will remain constant. Number of cookies is an essential measurement to monitor unique views to the course overview page, and was chosen because this is the unit of diversion for the experiment. Number of clicks will display number of unique cookies that clicked on the "start free trial", and help quantify how much traffic from the site is continuing forward by clicking the button. Click through probability provides a ratio of the unique cookies that visited the site and clicked on the button, which better captures the portion of unique cookies exploring the free trial option than simply assessing the raw values of cookies and clicks. Number of user-ids was not selected because it is a metric that would better serve as an evaluation metric, and give a lot of information once the experiment was completed. Furthermore, number of user-ids was not selected as an evaluation metric either. Even though it is similar to the gross conversion metric, gross conversion is a better choice since it provides more context to the daily number of users instead of providing a raw count.

Evaluation metrics, on the other hand, are expected to change and measure the results of the experiment. The evaluation metrics chosen were gross conversion and net conversion. Gross conversion provides a good measurement of the number of users who signed up for a free trial, and continued through the customer funnel model as discussed in the course. Net conversion goes a further step to document the number of users who continued enrollment after the free-trial period. Retention was initially chosen as an evaluation metric, but it was determined that the metric would make the experiment run longer than intended – about 119 days (if diverting all traffic). To shorten the experiment and still acquire the needed results, retention was removed.

In order to launch the experiment, I will expect results that are statistically significant in reducing the amount of students leaving the free trial (gross conversion). To ensure the number of students continuing past the free trial and eventually completing the course is not reduced, net conversion can be statistically significant, but does not need to be in order to launch. However

the CI and practical significance boundaries for net conversion should be closely examined to confirm no negative effects are introduced by the experiment.

Measuring Standard Deviation

- Standard Deviation of Gross Conversion: 0.0202
  - The analytical estimate would be comparable to the empirical variability since the unit of analysis (cookies clicking the start trial button) matches with the unit of diversion (cookies on overview page). This will decrease the variability between the two.
- Standard Deviation of Net Conversion: 0.0156
  - The unit of analysis for this metric (cookies clicking the start trial button) also matches with the unit of diversion (cookies on overview page), so we would expect the analytical estimate to match the empirical variability.

Sizing

Number of Samples vs. Power

I will not be using Bonferroni's correction during the analysis phase. The total number of pageviews needed to power the experiment is 685,325. The pageviews required for the Gross Conversion metric was 322,938 and for Net Conversion was 342,663. The metric with the highest number of views was selected and multiplied by two, to ensure both control and experimental groups had the same number of views.

Duration vs. Exposure

I would divert 100% of web traffic to this experiment, and would need approximately 18 days to obtain the complete number of pageviews. This was calculated by 685,325/(40000). The risk level for this experiment is very low since there is a small chance that a user is negatively impacted by the duration, and no sensitive or personal user data is being collected. Considering the low risk, I also chose to divert all traffic to reduce the expenses and efforts required to conduct the experiment.

Experiment Analysis
Sanity Checks

Invariant Metric	95% C.I.	Observed	Passes check?
Number of Cookies	0.4988-0.5012	0.5006	Yes
Number of Clicks	0.4958-0.5041	0.5005	Yes
Click Through Probability	0.0812-0.0830	0.08218	Yes

Result Analysis

Effect Size Tests

Evaluation Metric	95% C.I.	Statistically Significant?	Practically Significant?
Gross Conversion	-0.0291 to -0.0120	Yes	Yes
Net Conversion	-.01163 to .00183	No	No

Sign Tests

Evaluation Metric	p-value	Statistically Significant?
Gross Conversion	0.0026	Yes
Net Conversion	0.6776	No

Summary

      The Bonferroni correction was not used, since I am already relying to see statistical significance in both evaluation metrics before launching the change. Applying the correction would make it difficult to detect true positive results. However, if I were only using one evaluation metric to determine the launch, then it would be appropriate to implement the Bonferroni correction to reduce the chance of false positive results from occurring. In addition, there were no discrepancies between the effect size and sign test for this experiment.

Recommendation

      Taking in consideration the results of the experiment, I would not launch this experiment. Gross conversion results were statistically significant, as expected by this metric. The net

conversion metric however, did not produce statistically significant results and when exploring its confidence interval (CI) further, it included the negative value of the practical significance boundary. Another experiment with more power and samples could be conducted to minimize the CI further, but this is not likely to alter the net conversion metric results and it would be best to reevaluate the intervention.

In addition, after consulting with the business team and engineers and taking in consider their recommendations, I would explore other evaluation metrics that would better capture retention (number of users enrolled past the 14 day mark) and wouldn't result in a lengthy experiment.

Follow-Up Experiment

To reduce the number of frustrated students who cancel early, I would try to engage trial users with one of the coaches in the first few days of their enrollment. Ideally, this will not only expose users to the benefits of being enrolled in the nanodegree (as opposed to simply taking classes for free on Udacity or elsewhere) but will also display the degree which Udacity is committed to assisting their students. Trial users are offered a 30-minute online chat with a coach, where they are free to discuss a range of concerns including their expectations of the program, professional/education goals, and if Udacity is the best place for them. I hypothesize that this 30-minute coaching conversation will encourage users to stay enrolled in the program after the 14 day trial, despite the obstacles they may face as students.

My unit of diversion would be user-id because it is unchanging across electronic devices, and easy to monitor user activity throughout the trial. The invariant metric I would monitor would be number of user-ids. Some evaluations metrics I would consider are Coaching Engagement (number of trial users who opt for coaching), and Net User Conversion (number of users who remain enrolled past 14 days and make first payment ÷ number of users enrolled in free trial).

References

- http://www.graphpad.com/quickcalcs/binomial1/
- http://mathworld.wolfram.com/BonferroniCorrection.html