Udacity Data Analyst Nanodegree - Design an A/B Test

Project submitted by: Harish Garg(harish.garg@gmail.com)

Experiment Design

Metric Choice

Invariant metrics are defined as metrics that should not change across experimental or control groups whereas Evaluation metrics are expected to change over the experiment with differences observed between the experimental and control groups.

So, having above definitions as guide, I choose below metrics...

Invariant metrics

- Number of cookies: The number of unique cookies to view the course overview page. It's not expected to change across experimental groups.
- Number of clicks: The number of unique users/students (unique cookies) that click on the "start free trial" button. It's not expected to change across experimental groups.
- Click-through-probability: The number of unique cookies to click on the "start free trial" button divided by the number of unique cookies to view the course overview page. It's not expected to change across experimental groups.

Evaluation Metrics

- Gross conversion: The number of user-ids that enroll in the free trial divided by the number of unique cookies to click on the "start free trial" button. It’s a evaluation metric as number of enrollments could be affected by the experiment. A statistical significant reduction in enrollment is expected.
- Retention: The number of user-ids that stayed enrolled past the 14 day free trial (made a payment) divided by the number of unique cookies that clicked on the "start free trial" button. It is expected to be change in the experiment hence will be used as an evaluation metric. Retention is expected not to decrease.
- Net conversion: The number of user-ids to remain enrolled past the 14 day free trial (made a payment) divided by the number of unique cookies that clicked on the "start free trial" button. It is expected to change as part of experiment. Net conversion is expected not to decrease.

Metrics not used

- Number of User-Ids: We want to know the probability of enrolling and the metric useful here will be unique cookies that click on the "start free trial" button. Hence, this particular metric would not be much useful to us. Hence, we will not be using this metric.

Measuring Standard Deviation

Evaluation Metric	Standard Deviation
Gross Conversion	.0202
Retention	.0549
Net Conversion	.0156

- As the units of diversion for the experiment and analysis is cookies, it is very likely that the analytical std dev will match the empirical std dev.
- And for the metric retention, since units for the diversion and analysis are different i.e. user-Id vs cookies, it is likely that the analytical std dev and the empirical std dev will not match.

Sizing

Number of Samples vs. Power

- The Bonferroni Correction is not used.
- Number of page views needed to power the experiment is 4,741,212.

Duration vs. Exposure

The participants will be exposed to a minimal risk as,

- No new sensitive data is expected to be collected as part of the experiment. Experiment is not going to ask students to submit any kind of Political attitudes, personal disease history, sexual preferences related data.
- This experiment is being done for a Software skills enhancement course which doesn't put students in danger in any way. Hence, Duration of the experiment also doesn't pose any risk to the participants.

Hence, I would divert 100% of the traffic. Given the no. of page views needed, it would require 100 days. But this would be too long a experiment and won't be acceptable by any company. So, we will drop Retention as an evaluation metric. With evaluation metrics as Gross Conversation and Net Conversion, the no. of days required will be 18 days.

Experiment Analysis

Sanity Checks

Invariant Metric	Lower Bound	Upper Bound	Observed	Pass/Fail
Number of Cookies	.4988	.5012	.5006	Pass
Number of Clicks	.4959	.5041	.5005	Pass
Click-through Probability	.0812	.0830	.0822	Pass

Result Analysis

Effect Size Tests

Evaluation Metric	Lower Bound	Upper Bound	Statistically Significant	Practically Significant
Gross Conversion	-0.0291	-0.012	Yes	Yes
Net Conversion	-0.0116	.0019	No	No

Sign Tests

Evaluation Metric	p-value	Statistically Significant
Gross Conversion	0.0026	Yes

Net Conversion	0.6776	No

Summary

- We are considering two evaluations to study the experiment affects - Gross Conversion and Net Conversion. And we want both of them to match our expectations. Hence, we will not be using Bonferroni correction in this case.
- Gross Conversion shows up statistically significant in both Sign and Effect size tests. However, Net conversion is not statistically significant in both tests.

Recommendation

Gross Conversion numbers does indicates that the the experiment reduced the # of students who enrolled in the trial but not enough prepared, which is a success. However, the change in Net Conversion is not statistically significant in both the tests. But the confidence interval of the net conversion does include the negative of the practical significance boundary. That is, it's possible that this number went down by an amount that would matter to the business.

Hence, I would recommend against the launch.

Follow-Up Experiment

Description
For the students enrolled in the trial period, Udacity can maintain a “Streak” record where Udacity can actively prompt the student about maintaining the streak, i.e. student spends at least certain minimum hours in learning otherwise Udacity prompts (via website or email) that the student should maintain the streak.

Hypothesis
This brings in certain gamification to the course. More students will be encouraged to keep the necessary pace and will be more likely to maintain it as they go from trial to paid.

Metrics
Invariant-metric will be # of user-ids. User-ids would not change between groups over the experiment. Evaluation metric will be # of payments (post trial) divided by the number of unique user-id's as an evaluation metric., which is expected to change.

Unit of diversion
Unit of diversion will be enrolled user-ids.