Project 6 – A/B Testing

This template can be used to organize your answers to the final project. Items that should be copied from your answers to the quizzes should be given in blue.

Experiment Design

Metric Choice

Invariant metrics:	Evaluation metrics:
- Number of cookies	- Gross conversion
- Number of clicks	- Net conversion
- Click-through-probability on “Start free trial”	

Explanation on choice of metrics:

Invariant metrics:

Number of cookies

The metric is picked as one of the invariant metrics. The change being tested in the experiment is reflected after clicking “start free trial”. Thus, it will not affect visitors viewing the course overview page and the number of cookies should not be impacted under the current experiment.

Number of user-ids

The metric is not picked as one of the invariant metrics. Population with user-ids indicate that one has already enrolled in the course. As the change tested in the experiment is made after clicking the “start free trial” button, the metrics will be affected by the feature being tested and will vary.

Number of clicks on “Start free trial”

The metric is picked as one of the invariant metrics. Again, the change being tested in the experiment is reflected after clicking “start free trial”. Thus, the experiment should not affect one’s probability of clicking the button and the metrics will not be impacted before and after the new feature being implemented.

Click-through-probability on “Start free trial”

The metric is picked as one of the invariant metrics. The click through probability rate is a metric measuring the chance of viewers clicking the button. As the change tested in the experiment is made after clicking the “start free trial” button, the metrics will not be affected by the feature being tested and will vary.

Gross conversion

The metric is not picked as one of the invariant metrics. The change implemented in the experiment will affect the number of enrollment and thus the metrics. The metrics will thus vary and not suitable for the mentioned purpose.

Retention

The metric is not picked as one of the invariant metrics. Similar to gross conversion, the change implemented in the experiment will affect the number of enrollment and thus the metrics. The

metrics will thus vary and not suitable for the mentioned purpose.

Net conversion

The metric is not picked as one of the invariant metrics. Similar to the above two metrics, the change implemented in the experiment will affect the number of enrollment and thus the metrics. The metrics will thus vary and not suitable for the mentioned purpose.

Evaluation metrics:

Number of cookies

The metric is not picked as one of the evaluation metrics. The current experiment and hypothesis is looking at the number of students' enrollment an those ultimately remain in the program after the free trial. Number of cookies is not the concern in this experiment and thus not one of the metrics.

Number of user-ids

The metric is not picked as one of the evaluation metrics. Measuring the absolute number does cannot give us a full picture on the change of behavior of the users, i.e. the number of enrolled users may increase by 100 which seems a lot, the number of cookies may have in other way increased by 100K which dwarf the increment of enrolled users.

Number of clicks on "Start free trial"

The metric is not picked as one of the evaluation metrics. The experiment is concerned with the effect of the addition of new pop up feature. Users' experience will thus be the same before clicking the "start free trial" button and the issue being examined cannot be reflected by the metric here.

Click-through-probability on "Start free trial"

The metric is not picked as one of the evaluation metrics. The experiment is interested in the number of users who have enrolled in the free trial and complete checkout, and the number of users to remain enrolled afterwards. The click-through-probability only measures the number of clicks on the free trial button and is not the focus of the experiment.

Gross conversion

The metric is picked as one of the evaluation metrics. According to the experiment hypothesis, the additional feature should set clearer expectations to students upfront before they start the free trial. Thus the less determined students should be deterred from taking up the free trial. The gross conversion metric can measure such effect and is therefore chosen as the metric. The gross conversion metric is expected to decrease in the experiment.

Retention

The metric is not picked as one of the evaluation metrics. The metric is indeed a fitting metric to measure the hypothesis on whether the number of students continuing past the free trial and complete the course would decrease or not. However, it would take a very large number of page views in order to generate significant statistical power to support the metric (~0.5 year time) and the duration is considered too long. The metric is thus not considered as the evaluation metric.

Net conversion

The metric is picked as one of the evaluation metrics. The net conversion metric can measure whether users are still enrolling or have dropped out of the course after the free trial and this can be used to validate our proposed hypothesis that the addition of the button should be decrease the number of students ultimately completing the course. The metric is expected to remain constant.

Measuring Standard Deviation

Evaluation metrics:	Standard deviation
- Gross conversion	0.020230604
- Net conversion	0.015601545

The unit of diversion in this experiment is a cookie. Since the unit of analysis of the two evaluation metrics are user-ids. There will be great diversion between their analytic and empirical variability. The reason behind is that the way of grouping data is different between user-id and cookies, users may be switching to 5 browsers to access the course material and in turn one user-id record will correspond to 5 cookies. This circumstances takes place in every data point and explain the variability issue.

Sizing

Number of Samples vs. Power

Bonferroni correction?	No
Pageviews	685,275

Duration vs. Exposure

Fractions of traffic diverted	0.5
Days required	35

The fraction is chosen based on time and risk consideration. We would like to collect data within a reasonable frame in order to smoothen the weekdays effect (people may be more likely to visit Udacity site and make trial decision during weekend) and also to provide actionable insight at an earlier timeframe as we are dealing with a time critical business decision. One month with 50% of the traffic will be enough to serve the purpose. From a risk perspective, one should not divert all traffic to the new feature given the possibility of bugs or unpleasant user experience which may greatly discourage users' enrollments.

From my point of view, the experiment should not be very risky for the business. People choosing Udacity should understand the time commitment for a self-study course and the addition of a pop up reminder on time commitment will not discourage a great deal of users from enrolling the course. Technically, the implementation of the new feature should be simple and will not cause major issues with the website structure or traffic.

Experiment Analysis

Sanity Checks

	Lower bound	Upper bound	Observed	Passed?
Number of cookies	0.4988	0.5012	0.5006	Y
Number of clicks on "Start free trial"	0.4959	0.5041	0.5005	Y
Click-through-probability on "Start free trial"	-0.0013	0.0013	0	Y

Result Analysis

Effect Size Tests

	Lower bound	Upper bound	Statistical significance	Practical significance
Gross conversion	-0.0291	-0.0120	Y	Y
Net conversion	-0.0116	0.0019	N	N

Sign Tests

	p-value	Statistical significance
Gross conversion	0.0026	Y
Net conversion	0.6776	N

Summary

Bonferroni correction will not be used in the experiment. In this A/B testing, we would like to have gross conversion decreasing and net conversion remains unchanged (null-hypothesis not rejected), and this is a AND condition to be considered. The main purpose Bonferroni correction is to reduce false positive, and in turns increase false negative. Under an "AND" case, only if both metrics have fulfilled our criteria should lead to a launch of the new feature. Bonferroni correction decreases the false positive rate will have minimal effect to the launch, but increasing the false negative rate will have an impact and reduce the number of successful launches. Considering that the correction will turn the experiment too conservative, the technique should not be used here.

Recommendation

I will recommend to launch this experiment. From the result analysis, we can see that gross conversion decreases with statistical and practical significance, indicating that number of users completing the checkout and enrolled in the free trial has decreased, potentially because the pop up has discouraged undetermined learners. In turn, net conversion remains unchanged as indicated by the insignificance as shown in the different tests. This indicates that number of learners remaining in the program after free trials did not decreases significantly. The above two findings confirm our hypothesis and the experiment should thus be launched.

Follow-Up Experiment: How to Reduce Early Cancellations

Experiment Background

Users first enrolled in Udacity Courses may feel overwhelmed as the online self learning experience is totally new and different from our daily classroom learning method. One of the major causes of frustration is lack of guidance or involvement and thus leading to drop out from the program.

Changes made in experiment

Tailored personal communication with suggested relevant discussions based on student's progress and other useful onboarding tips will be sent to their email address. The strengthened sense of involvement will reduce student cancellation during the free trial.

Evaluation Metrics

I would use "net conversion" and "total time spent with course content" as my evaluation metrics. Measuring net conversion can determine whether the change can impact the number of users remaining enrolled in the program. Our expected outcome is that the net conversion metric will increase with the change implemented. The total time spent with course content metrics can measure the engagement level of the users, with personalized communication sent to help our users, they should be more engaged in their learnings and the metric will increase based on our hypothesis.

Unit of diversion

The unit of diversion which will be used in this experiment will be user-id.
As we are only concerned with about the students who have already enrolled to the program, user-id will be a sufficient unit of diversion for our purpose.