Experiment Design

Metric Choice

Invariant

- Number of cookies
- Number of clicks
- Click-through-probability

Evaluation

- Gross Conversion
- Net Conversion

Explanation:

- Number of cookies
  - This is the number of cookies viewing the course overview page. This happens before the user would see any change that was made due to the test, so it should be invariant and it would not be useful for evaluation.
- Number of user-ids
  - This isn't good as an invariant metric because the number of users enrolling in the free trial may change due to the extra step introduced by the test. However, it isn't very useful as an evaluation metric since we are more interested in users completing the free trial and remaining enrolled, rather than just a count of those starting the free trial.
- Number of clicks
  - This is another metric that would be measured before the change being tested, making it an invariant metric and not an evaluation metric.
- Click-through-probability
  - This is also another metric that would be measured before the change being tested, making it an invariant metric and not an evaluation metric.
- Gross Conversion
  - This is an essential evaluation metric. One of the goals of the test is to decrease the gross conversion. Ideally this would be done by decreasing the number of users that enroll in the free trial but would eventually quit due to lack of time.
- Retention
  - This appears to be a good evaluation metric- we are hoping that the change being tested will increases this ratio. However, sizing calculations indicate that it would require over 4.5 million pageviews to test with sufficient power and that would make the experiment take far too long.
- Net conversion
  - This is an essential evaluation metric- the goal of the change is to decrease the gross conversion without significantly decreasing the net conversion.

The goal is to decrease the number of users leaving the free trial due to a lack of time without significantly decreasing the number of users completing the free trial (which would affect revenue). In order to launch the change I would like to see a practically significant decrease in gross conversion without a significant decrease in net conversion.

Measuring Standard Deviation

- Standard Deviation of Gross Conversion = 0.0202
- Standard Deviation of Net conversion = 0.0156

It's possible that the analytic estimate underestimates the true variability in both metrics. We are dealing with ratios rather than individual counts in a distribution, so the actual distribution may be more complex than a normal distribution. Additionally, in both metrics the unit of diversion (cookies) is different from the unit of analysis (user id). One user could click on the "start free trial" button on different devices or in different browsers before enrolling. This would count them as multiple cookies but (once they enroll) as only a single user id, which would invalidate the independence assumption.

Sizing

Number of Samples vs. Power

The Bonferroni correction will be used since I am looking for results from both metrics. I will need 829,275 total pageviews for the experiment in order to have enough power to detect a 0.75% change in net conversion. Detecting a 1.0% change in gross conversion requires fewer pageviews (781,775), but I will need to go with the larger of the two numbers.

Duration vs. Exposure

I will divert 100% of traffic for 21 days.

I don't think this will be a very risky experiment. The hypothesis is that the change will not significantly reduce the number of users that continue past the free trial. It may even encourage some users to stick around longer (make more payments) since coaches may have more time. I think it is better to finish the experiment sooner, so diverting all traffic is fine.

Experiment Analysis

Sanity Checks

Metric	Lower Bound 95% CI	Upper Bound 95% CI	Actual	Passed?
Number of cookies	.4988	.5012	.5006	Yes
Number of clicks	.4959	.5041	.5005	Yes
Click Through Probability	-.0013	.0013	.0001	Yes

Result Analysis

Effect Size Tests

Metric	Lower Bound 95% CI	Upper Bound 95% CI	Statistically Significant?	Practically Significant?
Gross Conversion	-0.03035	-0.01076	Yes	Yes
Net Conversion	-0.01257	0.00282	No	No

Sign Tests

Metric	P value	Statistically Significant
Gross Conversion	.0026	Yes
Net Conversion	.6776	No

Summary

I did use the Bonferroni correction since I was running two tests and looking for significance of either/both of them (for comparison to the other significance results). In this case a statistically significant result would have a p value of less than 0.025. There does not appear to be any discrepancy- the gross conversion shows a statistically significant decrease while the net conversion difference is not statistically significant.

Recommendation

I recommend launching this change. We have successfully reduced the gross conversion in a statistically and practically significant way. We have also not found a statistically significant change in net conversion.

Introducing an expectation-setting message does appear to have been a useful change for Udacity. However, it may be worth running follow up experiments to improve the message, such as showing the message at different rates. For example, it may be better to ask the user how many hours a month or how many days a week they can dedicate some time. It may be better to ask how many hours the user plans on dedicating during the free trial period specifically, and to encourage them to delay enrolling in the free trial if needed.

Follow-Up Experiment

Experimental change:

I think that some portion of students may sign up for the free trial with good intentions but get busy at some point during the two weeks and procrastinate. They don't use the website enough to become committed and cancel before they get charged. I think sending one or more motivational reminder emails to users that have enrolled in the free trial may motivate them to continue. For a specific experiment, there could be an email sent one week into the free trial with stories from successful graduates about how Udacity helped them in their careers.

Hypothesis:

Sending a motivational reminder email to users enrolled in the free trial will help some of them avoid procrastination, become engaged in using the site, and eventually become paying users.

Metrics:

Retention is a great metric to evaluate. However, we should increase the limit for a practical change in order to lower the amount of pageviews we need- 2 or 3 % would work much better than 1%.

Unit of Diversion:

The unit of diversion will have to be user ID since the change will only affect users enrolled in the free trial.