Final Project - Free Trial Screener

At the time of this experiment, Udacity courses currently have two options on the home page: "start free trial", and "access course materials". If the student clicks "start free trial", they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid version of the course. After 14 days, they will automatically be charged unless they cancel first. If the student clicks "access course materials", they will be able to view the videos and take the quizzes for free, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.

In the experiment, Udacity tested a change where if the student clicked "start free trial", they were asked how much time they had available to devote to the course. If the student indicated 5 or more hours per week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours per week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free. At this point, the student would have the option to continue enrolling in the free trial, or access the course materials for free instead.

The hypothesis was that this might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have enough time—without significantly reducing the number of students to continue past the free trial and eventually complete the course. If this hypothesis held true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

The unit of diversion is a cookie, although if the student enrolls in the free trial, they are tracked by user-id from that point forward. The same user-id cannot enroll in the free trial twice. For users that do not enroll, their user-id is not tracked in the experiment, even if they were signed in when they visited the course overview page.

Experiment Design

Metric Choice

Invariant Metrics

Number of cookies

The additional free trial screener feature comes after the a user visits the course overview page and therefore the number of cookies should not be affected by this change. I expect this measure to be the same in both control and experiment groups.

Number of clicks

This measure is an invariant metric because The number of clicks is a tricker of the new trial screener feature and it comes before the new feature. I expect this measure to be the same in both control and experiment groups.

Click-through-probability

Again, it is an invariant metric because both the cookies and clicks are recorded before the free trial screener is triggered. I expect this measure to be the same in both control and experiment groups.

Evaluation Metrics

Gross conversion

The gross conversion is an evaluation metric because it capture the change in percentage of users who started the free trial. One of our objective is to see it this screener change direct some of the users who aren't ready for Nano degree to free course material option. We should expect a significant but slight drop in gross conversion.

Retention

Additional free trial screener is designed to keep students committed to spend more than 5 hours a week to the program. Retention is a great way to see how this new feature can improve student experience by informing students of the amount of effort needed. I would expect an increase in this metric.

Net Conversion

Because Udacity is setting clear expectations for the students before they enroll but also keep students who are committed in the program, Net Conversion is a great evaluation metric to measure its effect on committed students. I expect the net conversion to be the same or slight increased.

Metrics not selected

Number of user-ids

Number of user-ids provides similar information as Net Conversion does but since the number of users visit varies greatly between days, Net Conversion is a better choice as an evaluation metric.

Measuring Standard Deviation

Gross conversion:

P = probability of enrolling, given click = .20625
N = Number of clicks = Unique cookies to view page per day * Click-through-probability on "Start free trial" = 40,000 * .08 = 3,200
Standard deviation = sqrt((p*(1-p))/N) = **.0202**

I would expect the analytic estimate to be comparable to the empirical variability since the unit of analysis and unit of diversion in this case is the same.

Retention

P = probability of payment, given enroll = .53
N = Enrollments per day * (5000 / Unique cookies to view page per day) = 660 * (5000 / 40000) = 82.5
Standard deviation = **.0549**
I would expect the analytic estimate and empirical variability to be different. Because the units aren't the same.

Net conversion

P = probability of payment, given click = .1093
N = Number of clicks = Unique cookies to view page per day * Click-through-probability on "Start free trial" = 40,000 * .08 = 3,200
Standard deviation of Net conversion: **.0156**
I would expect the analytic estimate to be comparable to the empirical variability since the unit of analysis and unit of diversion in this case is the same.

Sizing

Number of Samples vs. Power

No, I will not use the Bonferroni correction during my analysis phase because the chosen metrics are very likely to be correlated.

Here I used an online sample size calculator to compute the size needed. With alpha = 5% and beta = 20%
http://www.evanmiller.org/ab-testing/sample-size.html

Metric	Baseline Conversion rate	Ratio of pageview	Sample size	Number of pageview
Gross Conversion	20.6%	.08	25,812	322,650 * 2 = 645,300
Retention rate	53%	660/40000 = .0165	39,115	4,741,213
Net Conversion	10.93%	.08	27,411	342,637.5 * 2 = 685,275

Duration vs. Exposure

If I include retention rate in our study, it would take more than 119 days to complete the experiment. Here I decided to only use gross conversion and net conversion as evaluation metrics so the time it takes to complete the experiment can be significantly reduced.

I think this is a relatively low risk experiment because the change should help students better identify if this program is a fit for them and does not impact the student experience. I choose .8 as the fraction of traffic exposed to have a shorter experiment.

Number of pageview requried: 685,275
Fraction of traffic exposed: 80%
Length of experiment: 22 days

Experiment Analysis

Sanity Checks

Z-score of 95% confidence interval: 1.96

Number of cookies:
Lower bound = .4988
Upper bound = .5012
Observed = .5006
Pass

Number of clicks on "Start free trial":
Lower bound = .49588
Upper bound = .5041
Observed = .5005
Pass

Click-through-probability on "Start free trial":
Lower bound = .0812
Upper bound = .0830
Observed = .0822
Pass

Result Analysis

Effect Size Tests

Gross Conversion:
Practical significance level = .01

Ncont = 17,293
Nexp = 17,260
Xcont = 3,785
Xexp = 3,423
Ppool = (3785 + 3423) / (17293 + 17260) = .2086
SEpool = sqrt(.2086*(1-.2086)*(1/17293 + 1/17260)) = .0044
m = 1.96 * .0044 = .0086
Pcont = .2189
Pexp = .1983
d = -.0206
Lower bound = -0.0291
Upper bound = -0.0120

It is statistically significant as the lower and upper bound does not include zero and it is also practically significant as it does not include practical significance level either.

Net Conversion:
Practical significance level = .0075
Ncont = 17,293
Nexp = 17,260
Xcont = 2,033
Xexp = 1,945
Ppool = (2033 + 1945) / (17293 + 17260) = .1151
SEpool = sqrt(.1151 * (1 - .1151) * ((1/17293) + (1/17260)) = .0034
m = .0034 * 1.96 = .0067
Pcont = 2033 / 17293 = .1176
Pexp = 1945 / 17260 = .1127
d = .0687 - .0716 = -.0049
Lower bound = -.0116
Upper bound = .0018

It is not statistically significant as the lower and upper bound include zero. Therefore it is also not practically significant.

Sign Tests

Here I will be using an online calculator to do a sign test.
http://graphpad.com/quickcalcs/binomial1.cfm
Gross conversion:
Number of days = 23
Number of successes = 19
P-value = .0026
The result is statistically significant at a 95% confidence level.

Net conversion:
Number of days = 23
Number of successes = 13
P-value = .6776
The result is not statistically significant at a 95% confidence level.

Summary

I used the Bonferroni correction. The two evaluation metrics that I have selected, gross conversion and net conversion are correlated and definitely not independent of each other. Bonferroni correction would have been too conservative. This is no discrepancy between the effect size hypothesis tests and the sign tests. The experiment has shown a significant decrease in gross conversion which is probability of enrolling, given click and no difference in net conversion, which is probability of payment, given click.

Recommendation
The result is expected and I would recommend to launch the change.
A significant decrease in gross conversion rate and a same net conversion rate mean the number of frustrated students who leave the program before first payment has been decreased but the number of students who continue the program has remained the same. This can be translated into an improvement on overall student experience and the experiment has met my expectation.

Follow-Up Experiment

Give a high-level description of the follow up experiment you would run, what your hypothesis would be, what metrics you would want to measure, what your unit of diversion would be, and your reasoning for these choices.

I would run a follow-up experiment to aiming to reduce the cancellation rate. After the student clicks the cancellation page and choose to cancel the subscription, he will be offered free online coaching / diagnostic service for 30 minutes if the student choose to stay.

Null Hypothesis: Offering free coaching service to students will not affect the cancellation rate.
Alternative Hypothesis: Offering free coaching service to students will decrease the cancellation rate during free-trial period
Evaluation metrics:
Probability of cancellation in 7 days (To measure if cancellation rate will be affected)
Probability of first payment (Expect to see a increase if the experiment is as expected)

Invariant metrics:

Number of user-ids (Track enrolled students and number of user-id should be unaffected)

Unit of diversion:

User-ids (The experiment happens on students who enroll the program.)