Experiment Design

Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here. (These should be the same metrics you chose in the "Choosing Invariant Metrics" and "Choosing Evaluation Metrics" quizzes.)

For each metric, explain both why you did or did not use it as an invariant metric and why you did or did not use it as an evaluation metric. Also, state what results you will look for in your evaluation metrics in order to launch the experiment.

- Number of cookies: (Invariant) The number of cookies should be the same in the control and experiment sets. The unit of diversion is cookies so therefore it should be about the same in both.
- Number of user-ids: This metric could be different in the control and experiment set, since you could have individuals who did enroll in the free trial and those who didn't because of the notice of the time commitments. This metric could be used as an evaluation metric since it is the number of enrollments but since it is not normalized, gross conversion is a better evaluation metric to evaluate enrollments.
- Number of clicks: (Invariant) This metric could be different in the sets. Individuals would have been assigned to a control or experimental group by the time they click the "Start Free trial" button but this is the population that we are interesting in so I am choosing it as an invariant.
- Click-through-probability: This metric would be random both sets and it has to do with the number of unique cookies. By the time a user clicks on the "Start Free Trial" button, he/she would have already been assigned to a group when viewing the course overview page. I don't expect this metric to be significantly different between the sets. Since I am choosing cookies and clicks already as invariant, I will leave this one out.
- Gross conversion: (Evaluation) This metric would help analyze the results if the change (getting the user's time commitment) affected the decision of going through the free trial by comparing the results of the two groups. If this metric decreases in the experiment group, then the change is successful and it should be launched.
- Retention: I had chosen this metric as an evaluation metric because would help analyze if the change actually helped retain students and get them to go through the course. However, since it made the experiment too long in duration, it was removed.
- Net conversion: (Evaluation) This metric should be different in the two sets and it will help to compare the results with gross conversion, meaning, if gross conversion decreases this metric should not decrease significantly since the change shouldn't affect the number of students that move on with the course after the trial.

Measuring Standard Deviation

List the standard deviation of each of your evaluation metrics. (These should be the answers from the "Calculating standard deviation" quiz.)

For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the the empirical variability, or whether you expect them to be different (in which case it might be worth doing an empirical estimate if there is time). Briefly give your reasoning in each case.

Metric	Description	Value	**Standard Deviation**	
Gross Conversion	Probability of enrolling, given click:	0.20625	***0.0202***	Not expected to be different to the empirical variability since the unit of analysis is the same as the unit of diversion.
Retention	Probability of payment, given enroll:	0.53	***0.0549***	Because the unit of analysis(enrollment) is different from the unit of diversion, empirical variability is expected to be different. *This metric was removed since it made the experiment take too long.*
Net Conversion	Probability of payment, given click	0.1093125	***0.0156***	Not expected to be different to the empirical variability since the unit of analysis is the same as the unit of diversion

When a unit of analysis is different from the unit of diversion is recommended to use empirical variability since when computing analytical variability, you are making the assumption that the data is normally distributed and independent. However, when the unit of analysis is different from the unit of diversion, then that independent assumption is no longer valid because you are diverting group of events and therefore they are correlated to each other, which will cause the variability to be much higher.

Sizing

Number of Samples vs. Power

Indicate whether you will use the Bonferroni correction during your analysis phase, and give the number of pageviews you will need to power you experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

I will not use Bonferroni. The original number of pageviews needed was 4,741,212. However, after removing retention as an evaluation metric, the new number of pageviews is ***683,625***.

$$Pageviews = (pageviews\ (enrolled/payment) \div Number\ of\ Cookies\ start\ free\ trial) * 2$$

Duration vs. Exposure

Indicate what fraction of traffic you would divert to this experiment and, given this, how many days you would need to run the experiment. (These should be the answers from the "Choosing Duration and Exposure" quiz.)

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?

I decided to divert all the traffic for 18 days since the change is not risky enough to limit how many users should go through the experiment, it is within a reasonable timeframe and there is nothing that indicates that we are dealing with sensitive data such as health history or political attitudes. The risk I see is if the prompt asking for the time commitment fails and doesn't allow the student to go any further, but this shouldn't happen if it had been tested before launching.

Experiment Analysis

Sanity Checks

For each of your invariant metrics, give the 95% confidence interval for the value you expect to observe, the actual observed value, and whether the metric passes your sanity check. (These should be the answers from the "Sanity Checks" quiz.)

For any sanity check that did not pass, explain your best guess as to what went wrong based on the day-by-day data. **Do not proceed to the rest of the analysis unless all sanity checks pass.**

	Lower Bound	Upper Bound	Observed	Pass
Number of cookies	0.4988203921	0.5011796079	0.5006396669	Yes
Number Clicks	0.4958844957	0.5041155043	0.5004673474	Yes

Result Analysis

Effect Size Tests

For each of your evaluation metrics, give a 95% confidence interval around the difference between the experiment and control groups. Indicate whether each metric is statistically and practically significant. (These should be the answers from the "Effect Size Tests" quiz.)

	Gross Conversion	Net Conversion
Lower Bound	-0.02912335834	-0.01160462436
Upper Bound	-0.01198639083	0.001857179011
Statistically Significant	Yes	No
Practically Significant	Yes	No

Sign Tests

For each of your evaluation metrics, do a sign test using the day-by-day data, and report the p-value of the sign test and whether the result is statistically significant. (These should be the answers from the "Sign Tests" quiz.)

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

Summary

State whether you used the Bonferroni correction, and explain why or why not. If there are any discrepancies between the effect size hypothesis tests and the sign tests, describe the discrepancy and why you think it arose.

Since the main reason of using the Bonferroni correction is to reduce the chance of a type I error (rejection of the null hypotheses when the null hypothesis is true) at the expense of a type II error(acceptance of the null hypothesis when the null hypothesis is false), in this case I think it is more logical not to use it since I will run into a higher risk of accepting a null hypothesis and implementing the change when I should reject it and keep looking for other ways to set students expectations prior to starting the free trial.

Bonferroni correction is also too conservative for this case since our metrics are dependent. In order to track the net conversion (user id's remaining enrolled after the 14 days trial) students have to go signup for the 14-day trial which is a variable used in the gross conversion. If gross conversion changes then net conversion is also more likely to change.

Recommendation

Make a recommendation and briefly describe your reasoning.

$H_0 =$ The change of prompting the student to enter the number of hours before signing up for a trial and advising to go through the course in audit mode if the minimum time commitment of 5 hours doesn't match with what the student has entered, will help reduce the number of students who drop out during the trial because of insufficient time. This being accomplished without reducing the number of students who go through the free trial and eventually completing the course.
My sanity checks passed and the observed values are within the 95% confidence interval.
The gross conversion 95% Confidence Interval is between -0.0291 and -0.012 which is statistically significant and it doesn't include the minimal boundary of $d_{min} = 0.01$. By this alone

we can recommend launching with 95% confidence that the change works. If we consider the sign test result of 0.0026, it agrees with the hypothesis since this value is lower than $\alpha = 0.05$ and therefore the result is unlikely to have come out by chance.
The net conversion 95% confidence interval is between -0.0116 and 0.0018 which is not statistically significant because it includes 0,it does include the negative value of the minimal boundary $d_{min} = 0.0075$, and the sign test result of 0.6776 also agrees that this change might have occurred by chance. Since the change in net conversion was negative then this is something that should be further explored as it might affect the business.
Going back to the hypothesis, we want gross conversion to be lower which is exactly what our analysis shows; however, we don't want net conversion to go down since the experiment should not affect the number of students that move forward after the trial and this analysis shows that net conversion does go down. Therefore, I do not recommend making the change since net conversion did change.

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.

One follow up experiment would be to do the same experiment but at a different time of the year. Because the experiment was done in October and November which are close to holiday season, the time that the students could spend on the nanodegree could be affected and therefore causing students to drop out during the trial at a higher rate than if it was at a different time of the year. We might find that the change actually does make a big difference and therefore the change should be implemented. Everything in the experiment stays the same.

Another follow up experiment would be to take the same approach but differentiating the nanodegree. For example, for the Data Analyst and Machine Learning nanodegrees 5hrs/week is probably not enough. Assuming Udacity has this data to prove that certain nanodegrees do require more time per week, the same experiment could be run per nanodegree.

Maybe the time commitment is not the main reason why students are dropping out from the free trial, but perhaps is the workload of nanodegree and the difficulty level. One experiment would be a quick test to highlight the main skillset needed to proceed with the nanodegree. The student must pass this test in order to continue to the free trial. If student doesn't pass then he/she can't proceed to enroll in the nanodegree and a suggestion for what course or nanodegree to take will be given according to the results.

Unit of diversion:
Cookie since we will start tracking from the very beginning when a student arrives to the page

Invariant metrics:

- Number of Cookies since we will set our population from this point
- Number of Click-test (Number of clicks to the test link) to track how many students are proceeding to this page
- Number of clicks-trial (Number of clicks to the start free trial version) to track how many students move forward to the course

Evaluation metric:

- Gross Conversion (number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button). This metric would help us analyze if adding the test affected the number of the students who went through the trial.
- Net Conversion (number of user-ids to remain enrolled past the 14-day boundary divided by the number of unique cookies to click the "Start free trial" button). This metric would help us analyze and compare if the change affected the number of student who successfully go through the trial and decide to continue.

Hypothesis:

The change of prompting the student to test their skills before continuing to the free trial, will help reduce the number of students who drop out the course after the free trial because of difficulty.