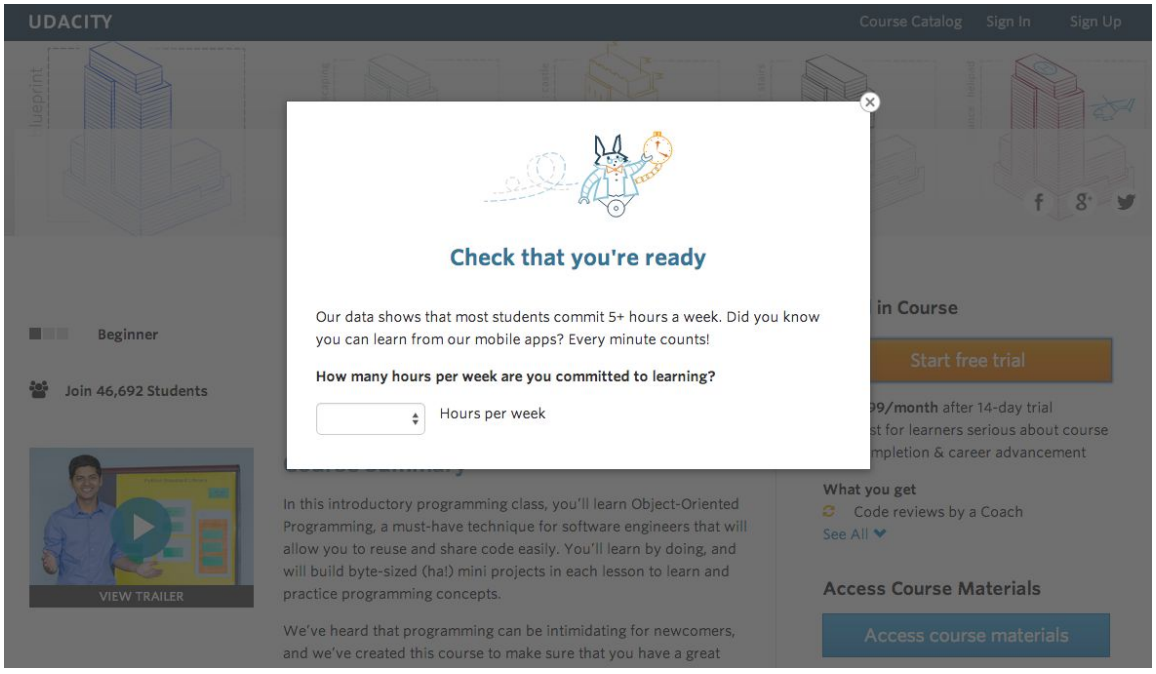
Design An A/B Test (Free Trial Screener)

Experiment Overview:

At the time of this experiment, Udacity courses currently have two options on the course overview 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.

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. 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 screenshot of the experiment is shown below:



Experiment Design

Metric Choice

Invariant Metrics: Invariant metrics are those metrics which should not change across the experimental and control groups. These metrics can play a good role in sanity checking. As part of this experiment, the following can be chosen as the invariant metrics.

- **Number of cookies (That is, number of unique cookies to view the course overview page.):** No of cookies is a good population sizing metric. As unit of diversion is cookie for this experiment, they should be randomly distributed across the control and experiment groups and both the groups should have approximately equal no of them to pass the initial sanity test.

- **Number of clicks (That is, number of unique cookies to click the "Start free trial" button, which happens before the free trial screener is trigger.) :** As the changes are made to the flow post the click on “start free trial” there should be no change in this parameter due to experiment and this parameter should be comparable across both the groups.

- **Click-through-probability (That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page.):** Click through probability is defined as the ratio of no of clicks to the no of pageviews, since both the parameters are invariant, this metric should be invariant as well and should be comparable across both the groups.

Evaluation Mertics: Evaluation metrics are those metrics which vary across the experimental and control groups, changes shown in these metrics are due to the changes in the system which are made for the experiment.

- **Gross conversion (That is, 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)** : Due to the change in the flow of “start free trial”, students who can spend less than five hours per week might not go through the regular flow and might opt for the option “access course materials” in experiemnt group, which would decrease the numerator of this metric, thus decreasing the overall metric when compared to the control group.

- **Net conversion (That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button.):** In the experiment group, students who can spend more then five hours might complete the checkout process and stay enrolled after the 14 day boundary period, so the numerator of this metric might remain constant/decrease in the experiment group which would decrease the overall metric in experiment group, when compared to the control group.

Measuring Standard Deviation

Since both the choosen evaluation metrics involve probabilities, the standard deviation of these metrics can be calculated using the formula of standard deviation for binomial distribution. This spreadsheet contains rough estimates of the baseline values for these metrics. The baseline values mentioned the sheet correspond to 40,000 pageviews, but we need to estimate the standard deviation for the choosen evaluation metrics which correspond to 5,000 pageviews.

S No.	Evaluation Metric	For 40k Pageviews	For 5k Pageviews
1.	Gross conversion	0.007152	0.0202
2.	Net Conversion	0.005515	0.0156

Since the unit of diversion and unit of analysis of the choosen evaluation metrics is same, both the emperical and analytical estimates will be comparable.

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.)

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?

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.**

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.)

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.)

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.

Recommendation

Make a recommendation and briefly describe your reasoning.

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.