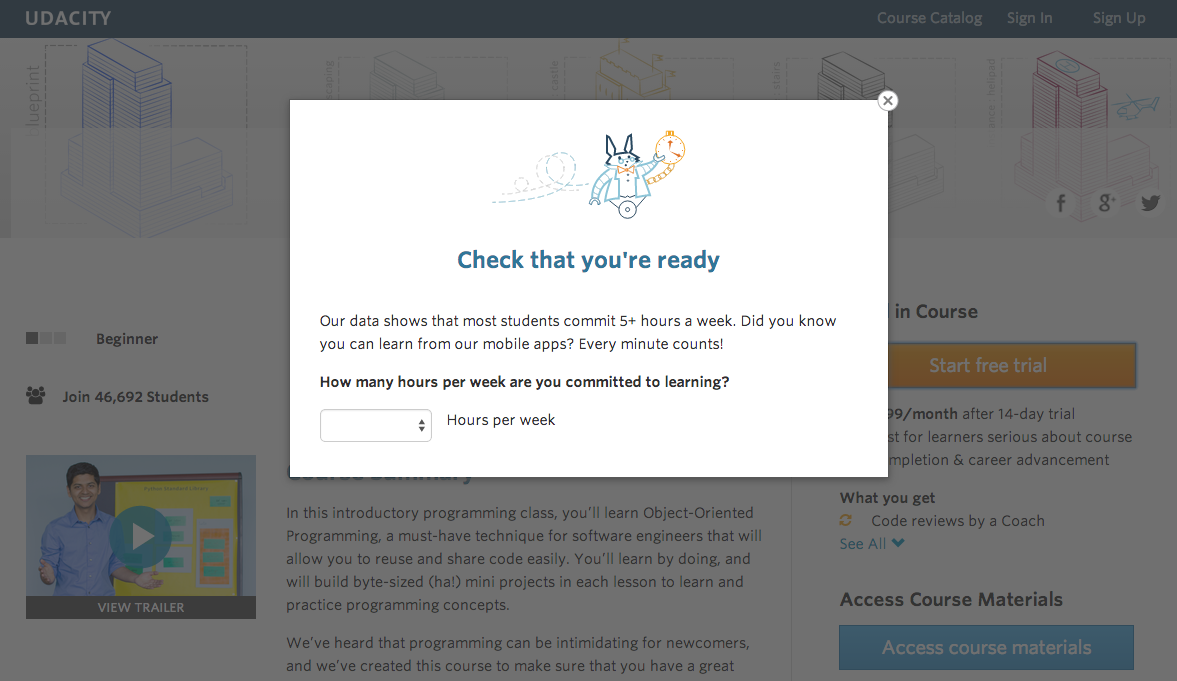
Experiment Design

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. This screenshot shows the experiment:



The primary aim of Udacity is to improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

**Null Hypothesis** : The null hypothesis is that this approach might not make a significant change and might not be effective in reducing the early Udacity course cancellation.

**Alternative Hypothesis** : The alternative hypothesis is that this might reduce 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.

**Metric Choice**

Any place "unique cookies" are mentioned, the uniqueness is determined by day. (That is, the same cookie visiting on different days would be counted twice.) User­ids are automatically unique since the site does not allow the same user­id to enroll twice.

● Number of cookies: That is, number of unique cookies to view the course overview page. (dmin=3000)

● Number of user­ids: That is, number of users who enroll in the free trial. (dmin=50)

● Number of clicks: T hat is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is trigger). (dmin=240)

● Click­through­probability: 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. (dmin=0.01) ● Gross conversion: T hat is, number of user­ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button. (dmin= 0.01)

● Retention: That is, number of user­ids to remain enrolled past the 14­day boundary (and thus make at least one payment) divided by number of user­ids to complete checkout. (dmin=0.01) ● Net conversion: That is, number of user­ids to remain enrolled past the 14­day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button. (dmin= 0.0075)

#### Invariant metrics (expected to be unchanged in the control and experimental groups)

#### number of cookies (cannot be affected by the experiment: users made a decision to visit the page before they were asked the question);

#### number of clicks (cannot be affected by the experiment: users clicked the button before they were asked the question);

#### click-through probability (cannot be affected by the experiment: it equals to the number of clicks divided by the number of cookies).

#### Evaluation metrics (expected to be different in the control and experimental groups):

#### gross conversion (can be affected by the experiment / can decrease: users could make a decision to enroll in the free trial in the experimental group less than in the control group because they did not plan to learn 5+ hours per week);

#### retention (can be affected by the experiment / can increase: enrolled users could be disappointed in the learning process less and make more payments in the experimental group than in the control group because they paid attention to studying 5+ hours per week);

#### net conversion (can be affected by the experiment / can decrease: users could enroll in the free trial less in the experimental group than in the control group, thus could decrease the number of people who paid).

#### Unused Metrics

#### Number of user-ids: The number of users who enroll in the free trial. User-ids are tracked only after enrolling in the free trial and equal distribution between the control and experimental branches would not be expected. User-id count could be used to evaluate how many enrollments stayed beyond the 14 day free trial boundary, but since it isn't normalized, I have elected not to use it.

#### The goals of the experiment in the practical meaning:

#### the number of payments should not be decreased;

#### the number of students who were disappointed and had not paid because they could not study enough time should be reduced.

#### The goals of the experiment in terms of our metrics:

#### the gross conversion should significantly decrease;

#### the retention should significantly increase;

#### the net conversion should not decrease.

**Measuring Standard Deviation**

This list contains rough estimates of the baseline values for these metrics (again, these numbers have been changed from Udacity's true numbers).

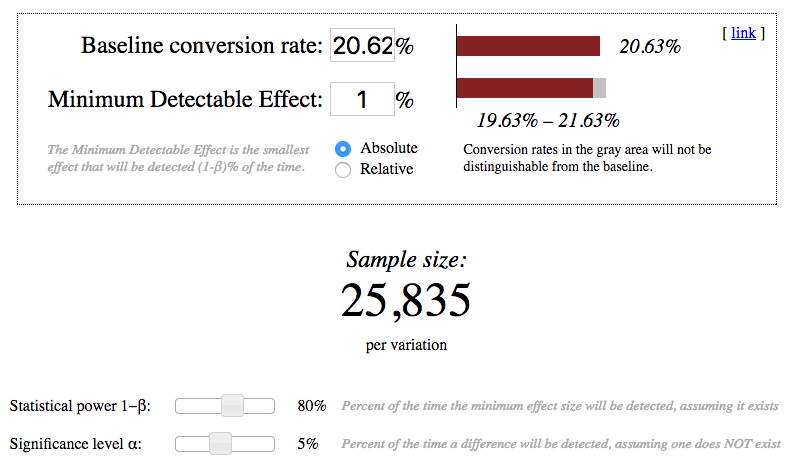
* Unique cookies to view page per day: 40000
* Unique cookies to click "Start free trial" per day: 3200
* Enrollments per day: 660
* Click-through-probability on "Start free trial": 0.08
* Probability of enrolling, given click: 0.20625
* Probability of payment, given enroll: 0.53
* Probability of payment, given click 0.1093125
* Number of cookies = 5000
* Number of clicks on "Start free trial" = 5000 × 0.08 = 400
* Number of enrollments = 5000 × 0.08 × 0.20625 = 82.5
* "SD Gross conversion = ", math.sqrt(0.20625 \* (1 - 0.20625) / 400)
* "SD Retention = ", math.sqrt(0.53 \* (1 - 0.53) / 82.5)
* "SD Net conversion = ", math.sqrt(0.1093125 \* (1 - 0.1093125) / 400)
* SD Gross conversion = 0.020230604137
* SD Retention = 0.0549490121785
* SD Net conversion = 0.0156015445825

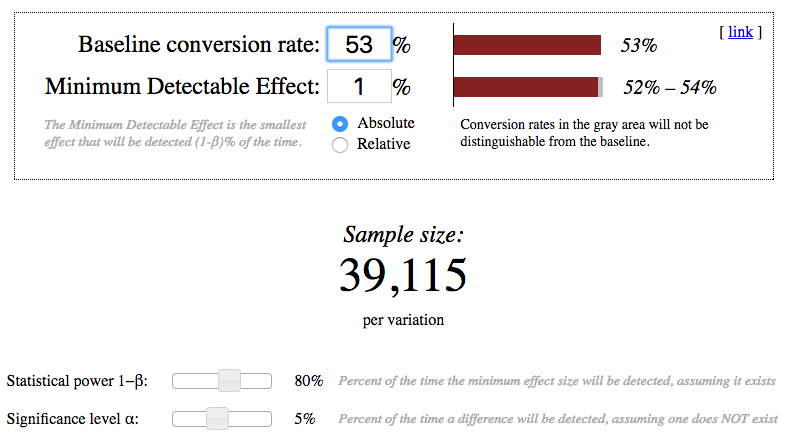
**Sizing**

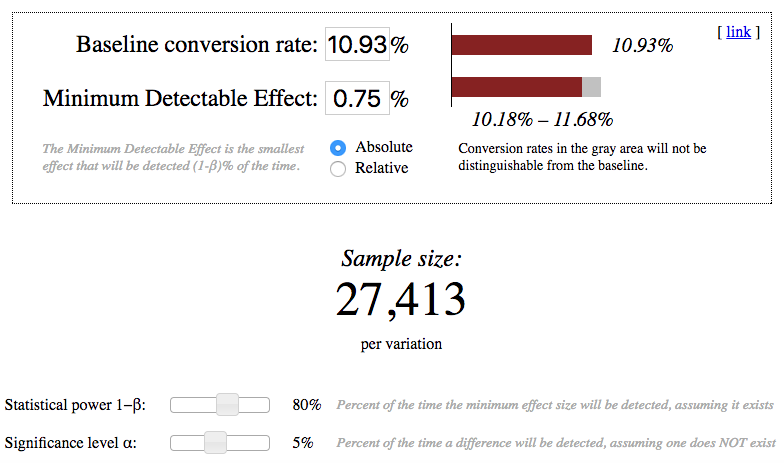
**Number of Samples vs. Power**

❓ Using the analytic estimates of variance, how many pageviews total (across both groups) would you need to collect to adequately power the experiment? Use an alpha of 0.05 and a beta of 0.2. Make sure you have enough power for each metric.

❗ I have used the [online calculator](http://www.evanmiller.org/ab-testing/sample-size.html) for calculating the sample sizes and have chosen the largest.







Gross conversion: 2 × 25835 × 40000 ÷ 3200 = 645875.0

Retention: 2 × 39115 × 40000 ÷ 660 = 4741212.12121

Net conversion: 2 × 27413 × 40000 ÷ 3200 = 685325.0

Pageviews required is maximum of pageviews required for Gross Conversion, Retention, Net Conversion. Therefore, the required pageviews is 47,41,212.

**Duration vs. Exposure**

If we divert 100% of traffic, given 40,000 page views per day, the experiment would take ~ 119 days. If we eliminate retention, we are left with Gross Conversion and Net Conversion. This reduces the number of required pageviews to 685,325, and an ~ 18 day experiment with 100% diversion and ~ 35 days given 50% diversion.

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 daybyday

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 daybyday

data, and report the

pvalue

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 highlevel

description of the follow up experiment you would run, what your hypothesis

would be, what metrics you would want