Final Project Instructions

Experiment Overview: 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. This screenshot shows what the experiment looks like.

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.

Metric Choice

Which of the following metrics would you choose to measure for this experiment and why? For each metric you choose, indicate whether you would use it as an invariant metric or an evaluation metric. The practical significance boundary for each metric, that is, the difference that would have to be observed before that was a meaningful change for the business, is given in parentheses. All practical significance boundaries are given as absolute changes.

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.) User-ids are automatically unique since the site does not allow the same user-id to enroll twice.

- **Number of cookies:** That is, number of unique cookies to view the course overview page. (d_{min}=3000)
- **Number of user-ids:** That is, number of users who enroll in the free trial. (d_{min}=50)
- **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). (d_{min}=240)
- **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. (d_{min}=0.01)
- **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. (d_{min}= 0.01)
- **Retention:** That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout. (d_{min}=0.01)
- **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. (d_{min}= 0.0075)

You should also decide now what results you will be looking for in order to launch the experiment. Would a change in any one of your evaluation metrics be sufficient? Would you want to see multiple metrics all move or not move at the same time in order to launch? This decision will inform your choices while designing the experiment.

Measuring Variability

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

For each metric you selected as an evaluation metric, estimate its standard deviation analytically. Do expect the analytic estimates to be accurate? That is, for which metrics, if any, would you want to collect an empirical estimate of the variability if you had time?

Sizing

Choosing Number of Samples given 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.

Choosing Duration vs. Exposure

What percentage of Udacity's traffic would you divert to this experiment (assuming there were no other experiments you wanted to run simultaneously)? Is the change risky enough that you wouldn't want to run on all traffic?

Given the percentage you chose, how long would the experiment take to run, using the analytic estimates of variance? If the answer is longer than a few weeks, then this is unreasonably long, and you should reconsider an earlier decision.

Analysis

The data for you to analyze is [here](). This data contains the raw information needed to compute the above metrics, broken down day by day. Note that there are two sheets within the spreadsheet - one for the experiment group, and one for the control group.

The meaning of each column is:
- **Pageviews:** Number of unique cookies to view the course overview page that day.
- **Clicks:** Number of unique cookies to click the course overview page that day.
- **Enrollments:** Number of user-ids to enroll in the free trial that day.
- **Payments:** Number of user-ids who who enrolled on that day to remain enrolled for 14 days and thus make a payment. (Note that the date for this column is the start date, that is, the date of enrollment, rather than the date of the payment. The payment happened 14 days later. Because of this, the enrollments and payments are tracked for 14 fewer days than the other columns.)

Sanity Checks

Start by checking whether your invariant metrics are equivalent between the two groups. If the invariant metric is a simple count that should be randomly split between the 2 groups, you can use a binomial test as demonstrated in Lesson 5. Otherwise, you will need to construct a confidence interval around your control value using a similar strategy as in Lesson 1, then check whether the experiment group value falls within that confidence level.

If your sanity checks fail, look at the day by day data and see if you can offer any insight into what is causing the problem.

Check for Practical and Statistical Significance

Next, for your evaluation metrics, calculate a confidence interval for the difference between the experiment and control groups, and check whether each metric is statistically and/or practically significance. A metric is statistically significant if the confidence interval does not include 0 (that is, you can be confident there was a change), and it is practically significant if

the confidence interval does not include the practical significance boundary
(that is, you can be confident there is a change that matters to the
business.)

If you have chosen multiple evaluation metrics, you will need to decide
whether to use the Bonferroni correction. When deciding, keep in mind the
results you are looking for in order to launch the experiment. Will the fact
that you have multiple metrics make those results more likely to occur by
chance than the alpha level of 0.05?

Run Sign Tests

For each evaluation metric, do a sign test using the day-by-day breakdown.
If the sign test does not agree with the confidence interval for the difference,
see if you can figure out why.

Make a Recommendation

Finally, make a recommendation. Would you launch this experiment, not
launch it, dig deeper, run a follow-up experiment, or is it a judgment call? If
you would dig deeper, explain what area you would investigate. If you would
run follow-up experiments, briefly describe that experiment. If it is a
judgment call, explain what factors would be relevant to the decision.

Follow-Up Experiment: How to Reduce Early Cancellations

If you wanted to reduce the number of frustrated students who cancel early
in the course, what experiment would you try? Give a brief description of the
change you would make, what your hypothesis would be about the effect of
the change, what metrics you would want to measure, and what unit of
diversion you would use. Include an explanation of each of your choices.