

# 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.

### Answer:

Invariant metrics:

- Number of cookies - This should be the same in both control and experiment group because experiment is after click.
- Number of clicks - This should be the same in both control and experiment group because experiment is after click.
- Click-through-probability - Same reason as above.

Evaluation metrics:

- Gross conversion - Number of enrollment is expected to decrease in the experiment group because the experiment is adding one more step before the enrollment. We need to see how much this change with the experiment.
- Retention - This is expected to increase in the experiment group because this experiment should reduce the number of students who leave after 14 days.
- Net conversion - Because this experiment reduces the number of enrollments, it is important to know whether the net conversion is affected (reduced) by this experiment.

## 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.

**Answer:**

- Gross conversion - 0.0202

Unit of analysis is click, and unit of diversion is cookie. Because they are different, it is expected that analytical estimation can be different from empirical variability, and it is worth doing an empirical estimate.

- Retention - 0.0549

Unit of analysis is userid (enrollment) and unit of diversion is user-id too because they are tracked with user-id once they start a free trial. So, analytical estimate should be comparable to the empirical variability.

- Net conversion - 0.0156

Unit of analysis is click and unit of diversion is cookie. Because they are different, it is expected that analytical estimation can be different from empirical variability, and it is worth doing an empirical estimate.

## **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 your experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

**Answer:**

No, I won't use Bonferroni correction because each metric is correlated with others, and it would give too conservative results.

The number of page views that I need is 685275

### **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.)

**Answer:**

I would divert 100% of traffic to this experiment, and it would require 18 days to run the experiment.

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

I chose to diver 100% of traffic because this experiment would not be risky for Udacity considering that most people who are willing to enroll should expect to spend more than 5 hours per week, and this experiment won't prevent those users from starting a free trial.

## 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.)

#### Answer:

Number of cookies - CI: (0.4988, 0.5012), Observed: 0.5006, pass

Number of clicks - CI: (0.4959, 0.5041), Observed: 0.5005, pass

Click-through-probability - CI: (-0.0013, 0.0013), Observed: 0 pass

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.**

#### Answer:

All sanity check passed.

### 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.)

**Answer:**

Gross conversion - CI: (-0.0291, -0.012). Statistically and practically significant

Net conversion - CI: (-0.0116, 0.0019). Statistically and practically insignificant

**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.)

**Answer:**

Gross conversion - p-value: 0.0026, Statistically significant

Net conversion - p-value: 0.6776, Statistically insignificant

**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.

**Answer:**

In this analysis, I didn't use Bonferroni correction because these metrics are correlated each other, and Bonferroni correction would be too conservative. Both effect size test and sign tests are consistent - the difference in gross conversion is statistically significant in both tests, and the difference in net conversion is statistically insignificant.

**Recommendation**

Make a recommendation and briefly describe your reasoning.

**Answer:**

Based on the test, this free trial screener improves the gross conversion without reducing net conversion in both statistical and practical senses. In other words, it will

reduce the number of students who will leave the free trial without significantly reducing the number of students who eventually stay after 14 days. Therefore, I would like to recommend to include the free trial screener. One thing that we should do is to ramp up the changes during the launch instead of applying the changes to all.

## 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.

### Answer:

In this analysis, the experiment was limited to a short time period. However, student behaviors may change as explained during the class. For example, it may be different during summer vacation, or winter time. Therefore, one follow-up experiment I would like to do is to monitor a set of small groups for a longer time period without the free trial screener, and see if we can draw similar conclusions for a longer time period.

In this experiment, I will measure the same metrics using the same unit of diversions (cookie) except that the experiment will stay longer for a small set of experiment groups.