# A/B Test Project

# Experiment Design

## Metric Choice

**List which metrics you will use as invariant metrics and evaluation metrics here.**

1. Invariant metrics should not change in the experiment and control groups. Number of cookies and number of clicks are qualified.
2. Evaluation metrics are expected to change in the experiment and control groups. Gross conversion, retention, and net conversion are qualified.

**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.**

1. Number of cookies: The experiment starts after the users clicks “start free trail” button. Good invariant metric because the number of cookies should not change. Not a good evaluation metric since the number of cookies is independent from the experiment.
2. Number of user ids: Not a good invariant metric because they depend on the experiment design. Not a good evaluation metric because it cannot justify the different group sizes of the control and experiment groups.
3. Number of clicks: The experiment starts after the users clicks “start free trail” button, therefore, the number of cookies and number of clicks should not change.
4. Click through probability: Not a good evaluation metric because the experiment happens after the user clicks the “start free trail” button. This metric is independent from the experiment. Good invariant metric because the click through probability are expected to be the same, and the message in the experiment appears after users click the “start free trail” button.
5. Gross conversion: This metric can be used to measure whether the message has a effects on the course drop-off rate. The gross conversion in the experiment group is expected to be lower.
6. Retention: This is a good evaluation metric. The retention in the experiment group is expected to be higher.
7. Net conversion: This is also a good evaluation metric. The net conversions in the experiment and control groups are expected to be different due to the message.

The experiment group should have lower gross conversion, higher retention, and higher net conversion. The message should help student to clearly identify their study outcomes and stay in the program if they commit to.

## Measuring Standard Deviation

**List the standard deviation of each of your evaluation metrics.**

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.

The rate of the three evaluation metrics are , , and

Given the sample size is 5000, the expected unique cookies that click the free trail button, and the unique ids that completed checkout are:

The standard deviation of each evaluation metric can be calculated as follows.

Standard deviation of gross conversion:

For gross conversion and net conversion, the unit of diversion matches the unit of analysis (both are unique cookies that clicked the free trial button). Therefore, their analytical standard deviation will likely match the empirical standard deviation in the experiment.

The unit of diversion for retention doesn’t not match the unit of analysis, therefore, its analytical standard deviation will not likely match the empirical standard deviation in the experiment.

## 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.**

The metrics are highly correlated, and the Bonferroni correction will likely to be too conservative. Therefore, I won’t use this correction.

In the baseline, and

Using , , and the sample sizes from Evan Miller, the required sample sizes are as follows.

### 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.**

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

I will direct 100% of the unique cookies that clicked the free trail button to the experiment. The message is designed to help students have a clear understanding of the preferred time spent per week in order to complete the program. It will not cause any physical harm or get any hurt. There is no sensitive data either since the experiment tracks the user ids after they enroll in the free trail, there is no evolvement in the student’s political attitudes, etc. The experiment will not cause significant harm to the users and it is considered as minimal risk.

If we use retention as the evaluation metric, it will take approximately days to complete the experiment and it’s not realistic for an A/B test. Therefore, for the rest of this report, we will evaluate gross conversion and net conversion only.

For gross conversion, it will take approximately days to complete the experiment.

For net conversion, it will take approximately days to complete the experiment.

# 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.**

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.**

We expect that the control and experiment groups have the same numbers of cookies and clicks, so the unit of diversion is 0.5.

For cookies:

For number of clicks:

Both cookies and number of clicks passed the sanity check.

## 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.**

Gross conversion:

. The observed difference is both practically significant and statistically significant.

Net conversion:

. The observed difference is neither practically significant nor statistically significant.

### 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.**

Gross conversion:

There are 4 out of 23 success experiment. Based on graphpad, the p-value of a two-tailed test is 0.0026. Therefore, the test result is statistically significant.

Gross conversion:

There are 10 out of 23 success experiment. Based on graphpad, the p-value of a two-tailed test is 0.6776. Therefore, the test result is not statistically significant.

### 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.

I did not use Bonferroni correction since the changes of both the gross conversion and the net conversion should meet the expectations. With two metrics, the chance for getting an error almost doubles. The hypothesis tests and the sign tests have the same results.

## Recommendation

**Make a recommendation and briefly describe your reasoning.**

The recommendation is based on net conversions in the experiment and control groups. The differences of net conversion is neither practically and statistically significantly. Although the observation of the difference is negative, but its confidence interval includes zero. Therefore, it’s likely that the experiment will not cause a difference in net conversion. It is not recommended to launch the experiment.

# 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.**

In the follow-up experiment, after the users join the free trail, the system will send an email to users in the experiment group including the time they spent in the first week, and remind they that users who spend at least five hours per week are most likely to complete the nanodegree.

Null hypothesis: by sending an email with the time users spent studying, it will not affect the retention of the users.

Unit of diversion: user ids. The users are tracked by user ids after they join the free trail.

Invariant metric: number of user ids.

Evaluation metric: retention. If the difference of retention between the control and experiment groups are both practically and significantly significant, it means the email helps student to stay in the nanodegree.