

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.)

- Invariant Metrics: Number of Cookies, Number of Clicks
- Evaluation Metrics: Gross Conversion, Net Conversion

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.

- **Number of cookies:** This is a good invariant metrics as this is a population sizing metric. Since the unit of diversion is cookies, number of cookies should be split evenly in control and experiment group.
- **Number of user-ids:** Per [instructions](#), "the unit of diversion is a cookie, although if a student enrolls in a free trial, they are tracked by user-id from that point of forward". We expect student enrolls in experiment group are going to be different from student enrolls in control group, meaning user-ids are expected to vary too, thus not good invariant metrics. However, without benchmark, it's hard to compare number of user-ids in different groups, also not good evaluation metric.
- **Number of clicks:** Number of clicks should stay unchanged in control and experiment group as the experiment that we are doing happens after user clicks, thus this is a good invariant metric.
- **Click-through-probability:** Since click-through-probability is number of clicks divided by number of cookies, as number of cookies and number of clicks are invariant metrics, this one would also be a good invariant metric.
- **Gross conversion:** gross conversion is number of user-ids get enrolled divided by number of unique cookies that click. We expect user-ids to differ in two groups, but number of clicks stay the same, so the ratio is a good metric as evaluation metric.
- **Retention:** This could be a good evaluation metric in a business sense. However, it may take too long to run the experiment and requires too many page views, so it may not be worth it.
- **Net conversion:** net conversion is a good evaluation metric, because user-ids remain enrolled past 14 days are expected to change in two groups, but number of unique cookies are expected to stay the same.

In order to launch the change, Gross conversion should significantly decrease, as people who enroll the free trial would decrease; Net conversion should not significantly decrease or even show a significant increase, as people who pay should not decrease, it should rather increase.

Measuring Standard Deviation

List the standard deviation of each of your evaluation metrics. (These should be the answers from the "Calculating standard deviation" quiz.)

- Gross conversion: 0.0202
- Net conversion: 0.0156

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.

For Gross conversion and Net conversion, analytic estimate should be comparable to the empirical variability. Because the unit of diversion and unit of analysis are the same -- cookies, so we are more confident that samples are independent.

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.)

No Bonferroni correction.

Number of Pageviews: 685325

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.)

I would divert 0.6 of the traffic, and I would need 29 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?

If I divert 0.5 of the traffic, then 35 days are needed. If I divert 0.6 of the traffic, then only 29 days are needed, which meet the goal of 'a few weeks' (≤ 30 days)

The experiment should not be risky for Udacity. It doesn't ask for sensitive information from users or limit their actions, and also don't change the content of the course. Meanwhile, diverting only 0.6 of the traffic also prevent small bugs that may occur in the process.

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.)

	Lower bound	Upper bound	Observed	Pass
Number of Cookies	0.4988	0.4988	0.5006	Yes
Number of clicks on 'Start free trial'	0.4959	0.5042	0.5005	Yes

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.**

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.)

	Lower bound	Upper bound	Statistical Significance	Practical Significance	α level	Bonferroni
Gross conversion	-0.0291	-0.012	Yes	Yes	0.05	0.025
Net conversion	-0.0116	0.0018	No	No	0.05	0.025

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.)

	p-value	Statistic significance	α level	Bonferroni
Gross conversion	0.0026	Yes	0.05	0.025
Net conversion	0.6776	No	0.05	0.025

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.

The two metrics we are tracking are highly correlated and using Bonferroni corrections would be too conservative. As α / n would be smaller than α , metrics are less likely to be significant than not using Bonferroni. In this experiment, we require ALL metrics to be significant in order to launch. Using Bonferroni correction will control false positive in trade of false negatives. In the ALL case, one false negative result will trigger wrong decision. Thus, we should not use Bonferroni correction.

Recommendation

Make a recommendation and briefly describe your reasoning.

I do not recommend launching. Although 'Gross conversion' is negatively practically significant, which is what we expected, 'Net conversion' is not statistically significant. Furthermore, confidence interval of 'net conversion' could be even negative, which means decrease of students who proceeds with payment. The launch would not benefit company.

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.

Early cancellations may also have direct relationship with the experience of students completing the first project. If students find the first project too difficult to complete, they may feel discouraged and frustrating, on the other hand, if students can complete first project within deadline or earlier, they could find it very encouraging. Since I don't have the data, I don't know what is a good time frame to draw of completing the first project. Personally I think I completed it within a month, and find it well structured.

Null hypothesis: Completing the first project within 2 months will not increase retention.

Unit of Diversion: unit of diversion should be user-ids. As soon as users enrolled in the free trial, they each have user-id that could be tracked.

Invariant Metrics: Number of user-ids enrolled in the course.

Evaluation Metric: Retention. After people who have been enrolled after 2 months, people who stay in the course divided by people who enrolled in the course.

If retention rate is statistically and practically significant, then Udacity may need to look into ways make project one more supportive and encouraging for students to continue.