A/B Testing
By Rehab Fathi Ali

Experiment Design
Metric Choice
Invariant Metrics: Number of cookies, Number of clicks & Click-through-probability
Evaluation Metrics: Gross conversion, Retention & 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 (Invariant): The unit of diversion is cookies, so cookies are population sizing metrics and are split evenly between control and experiment groups.
- Number of clicks (Invariant): This event happens before the proposed change so it does not affect our experiment.
- Click-through-probability (Invariant): This event also occurs before the proposed change and will not affect the experiment.
- Number of User-IDS (Not used): It could be used as an evaluation metric but since it is not normalised and net conversion is way better as an evaluation metric, we will not use it.
- Gross conversion (Evaluation): This metric is directly dependent on the proposed change. It will tell us the proportion of users who click on the button and do enroll in the trial.
- Retention (Evaluation): This is a good evaluation metric. The purpose of the experiment is to increase the percentage of students who do not drop out during the free trial and that's exactly what this metric measures.
- Net conversion (Evaluation): This is the combination of Gross conversion & Retention and will show the proportion of students who click the button and complete the process till the first payment. We would expect that the added feature will increase this metrics.

The expectation of the experiment is that the gross conversion will decrease significantly without a significant decrease in the net conversion or simply we decrease the number of students who enroll in the free trial and drop out before the first payment.

Measuring Standard Deviation

Evaluation Metrics	Standard Deviation
Gross conversion	0.0202
Retention	0.0549

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.

- Gross conversion: The unit of diversion is the unit of analysis (# cookies). So the analytic estimate would be comparable to the empirical variability
- Retention: The unit of diversion and the unit of analysis are different and I expect the empirical variability to be different.
- Net conversion:The unit of diversion is the unit of analysis. So the analytic estimate would be comparable to the empirical variability

Sizing
Number of Samples vs. Power
The Bonferroni test will not be used since we only have 3 independent metrics

Evaluation Metrics	Number of pages
Gross conversion	646450
Retention	4741212 (Dismiss as an evaluation metric for taking too long)
Net conversion	685325

Duration vs. Exposure
Percentage of the traffic: 100% of the traffic
Time: 685325/(40000) = 17.13 days

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?
- The experiment is not risky. It does not involve sensitive data and no one will get hurt by a simple question survey in the enrollment process. So we will use all traffic to complete the experiment in shorter time as this feature might have educational and financial benefits.

Experiment Analysis

Sanity Checks

Invariant metric	Lower bound	Upper bound	Observed	Pass
Number of Cookies	.4988	.5012	.5006	Pass
Number of Clicks	.4959	.5041	.5005	Pass
Click-through Probability	.0812	.0830	.0822	Pass

Result Analysis

Effect Size Tests

Evaluation Metric	Lower Bound	Upper Bound	Significance
Gross Conversion	-0.0291	-0.012	Yes (statisticially and Practically)
Net Conversion	-0.0116	.0019	No (statistically and Practically)

Sign Tests

Evaluation Metric	p-value	Statistically Significant
Gross Conversion	0.0026	Yes
Net Conversion	0.6776	No

Summary

No, the Bonferroni correction will be very conservative for our experiment. The two metrics are highly correlated and both will be used to take our decision not just one of them so the bonferroni correction will not be appropriate.

Recommendation

I recommend against the launch of this new feature. Even though the gross conversion statistics showed that this feature will decrease the number of students who will enroll in the free trial and drop out, the results hint out for a decrease in net conversion.
The confidence interval of the net conversion does include the negative of the practical significance boundary. That is, it's possible that this number went down by an amount that would matter to the business.

Follow-Up Experiment

A different approach to the previous experiment is to encourage students enrolling in the free trial to complete the course instead of discouraging them from enrolling in the free trial.
Experiment: Send a reminder every couple of days to the student with his learning report, suggestions to proceed learning and a success story of alumni.
Hypothesis: More students will complete the free trial and proceed to the first payment.
The unit of diversion: user_id
Invariant metrics: The user_id of all students enrolling in the free trial
Evaluation metrics: The number of students enrolling in the free trial/ number of students proceeding to first payment.