# A/B Testing Report

- Rahul Ravindran

## Experiment Design

### Metric Choice

### Invariant Metrics Selected:

Number of cookies - Invariant Metric
Number of clicks - Invariant Metric

### Evaluation Metrics Selected:

Gross Conversion - Evaluation Metric
Net Conversion - Evaluation Metric

**Explanation for invariant Metrics:**

Before discussing the reasons for selecting the following invariant metrics, it is important to understand the goal of the study. The business goal here is to ensure that students are not frustrated after enrolling in the free trial and end up leaving the course. Thus, it would allow not only for the students's experience with audacity but also ensure that each enrolled student gets proper attention from the coaches.

Keeping in mind these goals I selected the above metrics as invariant that is to say they won't change over the control or experiment group.

- I selected number of cookies, number of clicks as invariant simply because the change we are testing does not really affect the usability of free trial button, the users who view the course overview page or the click through probability of free trial button.
- Simply, put the experiment is designed to look at the overall conversion from starting free trial to proceeding to checkout and later completing the course.
- If the number of users who went on from going to free trial to checkout to course completion is more after the change is introduced, the experiment can be said to be successful.
- I did not go with click through probability because I felt it would capture outliers in our experiment and not it might not be true invariant in this experiment.

**Explanation for Evaluation Metrics:**

- Gross conversion is used to check how many clicks are converted into enrolment. This is important because the experiment setup mainly tries to encourage only students who can commit a certain amount of hours.  Thus, since the experiment can cause this value to change over the two groups, it is selected as evaluation metric.
- Gross conversion simply tracks a user to the checkout, but the net conversion looks at the entire set of operation, which includes from clicking on free trial, checkout,completing the trial period and finally making the first payment. As a business objective, you would want to know that your experiment has successfully caused a change in the number of people who aren't frustrated and leave the course in the free trial itself. Since, net conversion tracks the result

from clicking on free trial to payment, it actually computes the overall success rate of the experiment. Ideally, you would want all your enrolled students to make the first payment, thus suggesting that you have successfully captured the students who are committed to give their effort to the course.

- I did not go ahead with retention because it is more important to know the payment or enrolment done given a click on free trial. This measurement would not directly have any effect on our experiment.

All the above have a direct relation to the feature we are testing and hence become candidates for evaluation metric.

## Measuring Standard Deviation

Standard deviation for my metrics are
Gross Conversion:0.0202
Net Conversion:0.0156

Since both the evaluation metric are probabilities, they can be assumed to have a binomial distribution, thus not needing ay kind of empirical estimate. Hence, I feel it won't be necessary to an empirical analysis.

## Sizing

### Number of Samples vs. Power

I decided to not use Bonferroni's correction, because I felt that the events have a correlation between them. I will explain my views in the section which needs justification for this decision.

Therefore, since I have not used bonferoni's correction, I have gone ahead with using alpha=0.05 as the power for each metric.

Required Pageviews: 679620 across both branches

I used the r code to got the required result as shown in lectures.

This was done due to the fact that unit of diversion in our study is a cookie. the analytical estimates of both gross conversion and net conversion were made based on 5000 page views. It cannot be assumed that cookie and page views have a 1:1 correspondence.

### Duration vs. Exposure

#### Distribution of traffic

I would divert around 50% of the traffic overall.

#### Exposure of the experiment

In order to get the required number of page views which is around 339810 pageviews, it would require approximately 34 days.

#### Risks Involved

As far as the risk is concerned, Udacity could lose potential candidates who could perform well with less commitment to the course. Not all students would be able to commit as much time as people who do some kind of work. Therefore, a major pitfall could be that Udacity could lose potential customers.

# Experiment Analysis

## Sanity Checks

| | Lower Bound | Upper Bound | Observed | Passes |
|---|---|---|---|---|
| **Number of Cookies** | 0.4988 | 0.5012 | 0.5006 | Yes |
| **Number of clicks on start free trial** | 0.4959 | 0.5041 | 0.5005 | Yes |

I used the method as described in lesson 5. I assumed a 50% split in experiment and control group. using which I computed my standard deviation and my margin of error. Observed calculations were done by using the values from the control group.

Observed: Ncntrl/(Ncntrl+Nexp)

Both the metrics pass the sanity test, hence suggesting to move ahead with the experiment.

## Result Analysis

### Effect Size Tests

| | Lower Bound | Upper Bound | Statistical Significance | Practical Significance |
|---|---|---|---|---|
| **Gross Conversion** | -0.0291 | -0.0120 | Yes | Yes |
| **Net Conversion** | -0.0116 | 0.0018 | No | No |

### Sign Tests

| | p value | statistical significance |
|---|---|---|
| **Gross Conversion** | 0.0026 | Yes |
| **Net Conversion** | 0.6776 | No |

### Summary

I did not go ahead with Bonferroni's correction, simply due to the fact that my evaluation metrics were depended on each other. Both gross conversion and net conversion move together thus suggesting that Bonferroni's correction would not be significant. Only after an enrolment can a payment take place, thus both metrics were connected, substantiating my choice of not using Bonferroni's correction.

I observed that both the effect test size results and sign test results coincided. Gross conversion is statistically significant in case of both the tests, whereas net conversion wasn't.

## Recommendation

I would not go ahead with launching the experiment. So here are the reasons:

- Firstly, Gross conversion has a negative impact as seen from the effect test size result. Thus, there were significantly less enrolment. But this should be obvious as we are filtering out candidates so that we would want he coaches to be less taxed and ensure that students enrolling are ready to give their commitment to the course.
- Now, the above point does not raise any flag. You can ideally think of launching given a positive change from Net conversion. This is because it actually measures how many make the successful transition from enrolling to first payment. Since there is not much change in this metric, it could not provide any positive feedback for launching this experiment. This change could also be risky, since you are assuming commitment hours to be a significant to your objective, when in fact these course are usually taken up by people who are also involved in other activities. Thus, a person with a significant background could complete the course well within the threshold of the number of hours needed to commit per week. Hence, a certain risk is involved in launching the experiment.

If a drastic positive change would've been observed in net conversion, it would have given a clear indication to launch this change. However, the values don't support this decision in any way and with the risk involved, it would be better to not launch this experiment.

Thus, using both the above observation, I would recommend not to launch.

# Follow-Up Experiment

I think more than the number of hours, a background in that specific field would be more necessary to avoid students from being frustrated. Thus, instead of asking the number of hours a student can commit, it would be better to do either one of the following:

- A small questionnaire that can judge the perquisites a person needs. this could be simple statistic or python quiz.
- Another way would be rate their own knowledge with respect to categories like python, statistics etc. on a scale of beginner to experienced.

I think this would be more meaningful and will help improve with both net conversion and gross conversion.

The metrics and unit of diversion will be the same as the above experiment.

Null hypothesis: Prior background does not have any effect on early cancellation.