

## Experiment Design

### Metric Choice

Given that the experiment changes take place after clicking on the “Start Free Trial” button, any information that can be collected before that interaction is a good candidate for an invariant metric. Therefore, only 3 of the 7 metrics fit into that category. They are:

- Number of cookies
- Number of clicks
- Click-through-probability (number of clicks/number of cookies)

One can argue about using just the click-through-probability or just number of cookies and clicks, although, the first option can be misleading because it doesn't display the magnitude of the other two metrics. The second option does not have information about the interaction of the two metrics.

On the other hand, the other 4 metrics represents events that occurs after changes proposed on the experiment. Is crucial to notice that one of them has a big problem. The number of user-ids can be heavily affected by the site utilization which can lead to inaccurate conclusions. Gross conversion deal with this problem by calculating the probability of a cookies to enroll on the free trial given that he/she clicked on the “Start Free Trial” button. Retention and Net conversion also are probabilities which helps the comparison between different situations.

Lastly, I would look for:

- Decrease in Gross conversion bigger than 1%
- Increase in Retention bigger than 1%
- Maximum decrease in Net conversion of 0,75%

### Measuring Standard Deviation

- Gross conversion = 0.0202
- Retention = 0.0549
- Net conversion = 0.0156

Given that all 3 evaluation metrics are probabilities and their unit of analysis is the same of the experiment unit of diversion (cookies), both analytic and empirical variance are expected to be comparable.

### Sizing

#### Number of Samples vs. Power

Based on all 3 evaluation metrics description, is very likely to find some sort of correlation between them. With that in mind, using Bonferroni correction is not the best option in this case given the conservative nature of the correction.

The strategy is to calculate the number of pageviews needed for each evaluation metric and choose the bigger value. The results are:

- Gross conversion = 322938
- **Retention = 2370606**
- Net conversion = 342662.5

The experiment also needs a control group with the same size so the final number of pageviews needed is 4741212.

Update: After calculating the number of days needed to run the experiment using those 3 evaluation metrics, I've decided to remove retention from the set of evaluation metrics in order to lower the duration of the experiment. The new number of pageviews needed is 683625.

## Duration vs. Exposure

The fastest way to do the experiment is to divert 100% of the traffic which will need 18 days to complete. However, the "Start Free Trial" process can be a critical point of the site which makes the decision of diverting 100% of the traffic not so good. Also, from my point of view the experiment will test a change that is not crucial for the site, we can call it a quality of life change. This means that increasing the duration of the experiment shouldn't be a big problem. Thus, my decision is to run the experiment over 35 days by diverting only 50% of the daily traffic.

## Experiment Analysis

### Sanity Checks

- Number of cookies
  - CI = [0.4988, 0.5012]
  - Observed = 0.5006
  - Passes
- Number of clicks
  - CI = [0.4959, 0.5042]
  - Observed = 0.5005
  - Passes
- Click-through-probability
  - CI = [-0.0013, 0.0013]
  - Observed = 0.0001
  - Passes

## Result Analysis

### Effect Size Tests

- Gross conversion ( $d_{min} = 0.01$ )
  - $CI = [-0.0291, -0.0120]$
  - The confidence interval does not include 0 = Statistically significant
  - The absolute value of both bounds are bigger than  $d_{min}$  = Practically significant
- Net conversion ( $d_{min} = 0.0075$ )
  - $CI = [-0.0116, 0.0018]$
  - The confidence interval includes 0 = Not statistically significant
  - The absolute value of the upper bound is lower than  $d_{min}$  = Not practically significant

## Sign Tests

With the p-value of 0.0026 the sign test agree with the effect size test that the change in Gross conversion is statistically significant. The sign test also confirms that the difference in Net conversion is not statistically significant with p-value of 0.6776.

## Summary

As discussed before, both evaluation metrics can have some level of correlation because the payment (net conversion) requires the user to enroll (Gross conversion). Another important aspect is the condition to launch the experiment, we need to achieve the expected results on both evaluation metrics. Because of that, the impact of reducing false positive rate is lower than the impact of reducing false negative rate. Therefore, no correction was used.

Both effect size and sign test agreed on the results that Gross conversion is statistically and practically significant and Net conversion don't.

## Recommendation

Based on the initial hypothesis, the main goal is to reduce the number of users who enroll on the free trial (Gross conversion) without reducing the number of users that continue the course after the free trial (Net conversion). The tests results has shown that both objectives have been met regarding the minimum significance thresholds. Although, the net conversion confidence interval includes the practical significance boundary which shows the possibility of undesirable results. My recommendation is to tune the experiment in order to get more certainty that a possible change in net conversion is not a issue.

## Follow-Up Experiment

- Experiment: How to reduce early cancellations
  - Consider that an early cancellation can happen in the first month of the course, which can include the free trial.
- Change: Every course will have a Lesson 0 that gives a simple and fast example of what the student will learn in the course, including a project. Maybe this new lesson could be a good fit for the 2 weeks free trial.

- Hypothesis: There is the possibility that some of the subjects of a course are less appealing for the student. Also, some of the first classes are not directly related with the product that will be achieved in the end of the course (maybe more theoretical classes). So with this new lesson 0 the student will have a glimpse of the future in the course which can keep the student engaged.
- Metrics:
  - Invariant:
    - Number of users who enroll in the free trial is the only invariant metric that I can think of.
  - Evaluation:
    - Probability of 1 month engagement, given enroll: that is, the number of users who continued in the course divided by the number of users who started a course.
- Duration: The experiment would apply the change during a period of 15-20 days. After that all users that started a course would be tracked for 30 days or until a cancellation. The experiment should last 1 month and 20 days at maximum. Given the frequency of a enrollment, I think this duration is decent.
- Unit of diversion: The best unit of diversion for the experiment should be the user id. It's also a good idea to use only users who are starting a course. So each time a user start a course it will be chosen randomly which group to participate (control or experiment).