# Udacity - Data Analyst Nanodegree

# Project 7: A/B Testing

# Created by: Layne Newhouse

# Submitted: March 16, 2017

# Experiment Design

## Metric Choice

***Number of cookies*** (number of unique cookies to view the course overview page)

* Number of cookies is a metric that should not change within the experiment because it precedes any changes that we present to the user i.e. the free trail screener. For this reason, it was chosen as an **invariant metric** and not an evaluation metric.

***Number of user-ids*** (number of users who enroll in the free trial)

* Number of user-ids is a metric that we expect see some change in. This is because it is dependent on the number of people that choose to continue with the enrollment process after observing the free trail screener. For this reason, it does not qualify as an invariant metric. This metric could be used as an evaluation metric but I have chosen to use the gross conversion rate instead which purveys the same information but in a slightly different form.

***Number of clicks*** (number of unique cookies to click the “Start free trail” button)

* Number of clicks is another metric that, like the number of cookies, precedes the experiment and thus we can use it as an **invariant metric** and not an evaluation metric for the same reason.

***Click-through-probability*** (number of user-ids to complete checkout and enroll in free trial divided by number of unique cookies to click the “Start free trail” button)

* Click-through-probability is a function of the number of clicks and the number of cookies and thus precedes the experimental change. For this reason, it is not a good evaluation metric but could be used as an invariant metric but I have opted towards using clicks and cookies instead as this would simplify further calculations

***Gross conversion*** (number of user-ids to complete checkout and enroll in the free trial divided by the number of unique cookies to click the “Start free trail” button)

* Gross conversion will provide insight on the effect that the free trail screener has on the number of students that enroll in the free trial. Since this metric is expected to change after implementing the experiment, it cannot be used as an invariant metric. This metric is analogous to the number of user-ids and will be used as an **evaluation metric**.

***Retention*** (number of user-ids to remain enrolled past the 14-day boundary divided by the number of user-ids to complete checkout)

* Retention is a metric that would provide us insight on whether student that have enrolled are less frustrated with the course-work and thus have stayed with the program past the free trail period. Because this metric is a function of our experiment it can not be used as an invariant metric. The metric itself could be used as an evaluation metric but since the unit of analysis is user-ids, a calculation regarding the necessary size of the experiment shows us that this metric is un-feasible as it would take too long to collect the data.

***Net conversion*** (number of user-ids to remain enrolled past the 14-day boundary divided by the number of unique cookies to click the “Start free trail” button)

* Net conversion provides us with similar information as retention but with the benefit of having cookies as the unit of analysis. This allows us to gain insight on the effect of the free trail screener with regards to the number of students that continue past the free trail period. Because of these reasons, net conversion is a good **evaluation metric** for our experiment and thus cannot be used as an invariant metric.

In order to recommend launching the experiment, we would like to see a statistically significant decrease in the gross conversion rate, which equates to a reduction in students that enroll in the free trail, while at the same time observing no change, or a positive change in the retention rate, thus not decreasing the number of students that will continue on to make payments.

## Measuring Standard Deviation

***Gross conversion***

***Net conversion***

For the evaluation metrics, I expect the analytical estimate to be comparable to the empirical variability since the unit of analysis is the same as the unit of divergence, which in both cases is a cookie.

## 

## Sizing

### Number of Samples vs. Power

Using <http://www.evanmiller.org/ab-testing/sample-size.html> the following sample sizes were obtained. The Bonderroni correction was not used during this analysis.

***Gross conversion***

***Net conversion***

\*Note: we multiply the total pageviews by two to obtain adequate pageviews for both the control and the experiment groups

Therefore, in order to power the experiment properly we will need 685,276 pageviews.

### Duration vs. Exposure

Since this experiment is low risk, being we are only introducing a pop-up when the ‘Start free trail’ button is being clicked, I would expose all the traffic to the experiment, 1.0, which would result in the experiment being completed in 18 days. Given the ability to ramp up the exposure to the project I would consider beginning with diverting 0.2 of the traffic and increasing this number to 1 in increments such that the project would be completed in 28 days. By analysing the invariant metrics while they initially came in, you would be able to catch any early bugs in the experiment.

# Experiment Analysis

## Sanity Checks

***Number of cookies (pageviews)***

Since is within the confidence interval, the sanity check for this metric passes.

***Number of clicks***

Since is within the confidence interval, the sanity check for this metric passes.

## Result Analysis

### Effect Size Tests

***Gross conversion***

Statistically significance is achieved in this metric since the confidence interval does not include 0. It is also practically significant since .

***Net conversion***

Neither statistically significance or practical significance is achieved in this metric since the confidence interval includes 0.

### Sign Tests

Using <https://graphpad.com/quickcalcs/binomial1.cfm> the following p-values were obtained.

***Gross conversion***

Using graphpad, we obtain a two-tail p-value of 0.0026. Since this is less than α = 0.05 the Gross conversion sign test is statistically significant.

***Net conversion***

Using graphpad, we obtain a two-tail p-value of 0.6776. Since this is greater than α = 0.05 the Net conversion sign test is not statistically significant.

### Summary

I did not use the Bonferroni correction because I was only using two metrics and I choose to accept the risk that there is approximately a ten percent chance that one of the metrics showed statistical significance when there was not, and vice versa. There was no discrepancy between the hypothesis test and the sign test.

## Recommendation

I recommend based on the observed data that Udacity goes ahead with implementing the free trial screen on the basis that it should reduce the number of students that enroll in the 14-day free trial while not having a significant effect on the number of students that stay enrolled past this trial period.

# Follow-Up Experiment

A potential follow up experiment could be as follows:

Test a change where when a student first enrolls in the 14-day free trail, they are taken to an overview page of their mentors and coaches with pictures and background summaries. We could test to see if this leads to more users using the forums to ask questions as well as a higher conversion rate.

The unit of diversion would be user-id, since we are tracking whether the user uses the forums and whether the user continues past the free trial period.

An invariant metric could be the number of enrollments (number of new user-ids) since this step happened prior to showing the mentors and coaches. The evaluation metrics would be the number of times users visit the forums as well net conversion.