# Introduction

This is an Udacity Data Analyst Nanodegree.project. The experiment is described as follow. Udacity courses currently have two options on the home page: "start free trial", and "access course materials". If the student clicks "start free trial", they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid version of the course. After 14 days, they will automatically be charged unless they cancel first. If the student clicks "access course materials", they will be able to view the videos and take the quizzes for free, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.

In the experiment, Udacity tested a change where if the student clicked "start free trial", they were asked how much time they had available to devote to the course. If the student indicated 5 or more hours per week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours per week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free.

# Experiment Design

## Metric Choice

Invariant metrics I choose are: Number of cookies, Number of clicks, Click through probability.

Number of cookies mean number of unique cookies to view the course overview page (dmin=3000). It should remain the same during the experiment. So it’s an invariant metric as its value is expected to be the same in the experiment and the control group.

Number of clicks will also stay constant, as users haven’t seen the experiment when they click the button. So it can be an invariant metric.

Since the is a ratio of number of clicks and number of cookies, it will not change neither. Hence, it is invariant too.

Evaluation metrics I choose are: Gross Conversion, Net Conversion.

Gross conversion is the number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button (dmin= 0.01). It’s expected to decrease as the experiment carries out. Because the experiment should reduce the number of students enrolling who can’t make the required time commitment.

Net conversion is the number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button (dmin= 0.0075). It should not change as the students remain enrolled after “free trial” and the clicks will not be affected by the experiment.

## Measuring Standard Deviation

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

The standard deviation of 3 evaluation metrics I chose is listed in the Table1 below.

Table. 1 The standard deviation of evaluation metrics

|  |  |
| --- | --- |
| **Evaluation metrics** | **Standard deviation** |
| Gross conversion | 0.0202 |
| Net conversion | 0.0156 |

Gross Conversion and Net Conversion, their empirical variance should approximate analytical variance, because the unit of analysis and unit of diversion is the same, user-ids/cookie-ids. On the other hand, the empirical variability should be calculated for Retention as it’s not likely to match the empirical standard deviation.

## Sizing

### Number of Samples vs. Power

Bonferroni correction will not be uesd in the analysis phase. After feeding parameters to sample size calculator (<http://www.evanmiller.org/ab-testing/sample-size.html>), I got 25835 samples needed for Gross conversion and 27413 samples for Net conversion. And after scaling from the given unit to pageviews, eventually I will need 685325 page views to run the experiment for both the metrics.

### 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.)

Udacity receives 40000 unique cookie views per day. This experiment isn't risky and participants will not suffer any except a warning. So I set a hight fraction of experiment exposure to Udacity visitors as 80%, the duration will be 22 days.

# Experiment Analysis

## Sanity Checks

Sanity check result are listed in table2.

Table 2. Sanity checks (alpha=0.05)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Invariant metric | Lower bound | Upper bound | Observed | Result |
| Number of cookies | 0.4988 | 0.5012 | 0.5006 | Pass |
| Number of clicks on “Start free trial” | 0.4959 | 0.5041 | 0.5005 | Pass |
| Click-through-probability on “Start free trial” | -0.0012 | 0.0013 | 0.0001 | Pass |

All invariant metrics passed snity check.

## Result Analysis

### Effect Size Tests

Effect size tests result is listed in table3.

Table.3 Effect size tests result (alpha=0.05)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Evaluation metric | dmin | Observed difference | Lower bound | Upper bound | Statistical significance | Practical significance |
| Gross conversion | 0.01 | -0.0205 | -0.0291 | -0.0120 | Yes | Yes |
| Net conversion | 0.0075 | -0.0048 | -0.0116 | 0.00186 | No | No |

As shown in Table3, the observed different and minimum detectable effect of Gross conversion are not in the confidence interval, it’s both statistical and practical significant. On the other hand, net conversion is neither statistically nor practically significant.

### 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.)

### 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.

## Recommendation

Make a recommendation and briefly describe your reasoning.

# 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.