# Experiment Design

## Metric Choice

## Invariant metrics:

**Number of cookies: (**That is, number of unique cookies to view the course overview page**)**
The unit of diversion is a cookie. Thus when we do this experiment, when should assign the number of cookies randomly. This metrics occurred before experiment. It would not change before experiment or after it.

**Number of clicks: (**That is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is trigger)**)**
The reason we choose the metrics is we want the data in both groups be comparable. Besides, we also want it be random since the unit of diversion is a cookie.

**Click-through-probability: (**That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page**)**
Click-through-probability is good invariant metric too. This event completes before main experiment. We hope it can be comparable between two groups.

All of these metrics occurs before experiment so they are not suitable as evaluation metrics.

## Evaluation metrics:

**Gross conversion: (**That is, 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).
Since the difference between two groups is a question which occurs after clicking "Start free trial" button, we want to know if this change will affect the number of students enrolling a free trial. In this case, gross conversion is a great metrics to see the effect.

**Net conversion: (**That is, 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**)**
The unit of diversion is a cookie. We also need to see the overall percentage that how many people stay in the course after clicking the button. (some of them may click the button more than once in different day and decide to enroll in the end.)

All of these metrics can be variant depends on people in this experiment which is not qualified as invariant metrics.

## Dropped metrics:

**Number of user-ids: (**That is, number of users who enroll in the free trial**)**
We don't use it as invariant metrics since the result will be affected by "time request message". It's can be different between two groups.
Considering our unit of diversion is a cookie which are measured by day and we separate our data by assigning cookies randomly, number of user-ids can be more likely affected than gross conversion by this factor as evaluation metrics.
Thus we decide to drop this metrics.

**Retention: (**That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout).
This metrics seems highly correlated with our hypothesis. But as we mentioned before, the way we separate our data will affect the accuracy of this metrics. Thus we decide to not use this metrics.

The hypothesis is: this change might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have enough time—without significantly reducing the number of students to continue past the free trial and eventually complete the course.
If the hypothesis is true, we will see the number of students who enrolling free trial decrease and the number of students who pay for class don't change a lot. If gross conversion decrease significantly with net conversion don't change a lot or gross conversion don't change a lot with net conversion increase significantly, we will launch the experiment.

## Measuring Standard Deviation

| | |
|---|---|
| **Gross conversion** | 0.0202 |
| **Net conversion** | 0.0156 |

We know our unit of diversion is a cookie and we also find both gross conversion and net conversion's denominator are the number of unique cookies to click the "Start free trial" button. In this case, we can say both metrics have independently drawn samples. Based on this, analytical and empirical probability of these metrics should match.

## Sizing

### Number of Samples vs. Power

Since our goal can be interpreted as AND problem, we shouldn't use Bonferroni correction. I choose gross conversion and net conversion as my metrics. In this case, we need approximate 685325 pageviews to complete experiment.

### Duration vs. Exposure

| | |
|---|---|
| **Pageviews per day** | 40000 |
| **Number of Pageviews** | 685325 |
| **Fraction of Traffic Exposed** | 0.9 |
| **Length of Experiment** | 19 |

We add an extra information of "time request" in this experiment. This change is a low risk test since we will not hurt anyone and there is no important information involved in this experiment. In this case, we can set a high fraction of traffic exposed to shorten our length of experiment. Our final length of experiment is 19 days.

# Experiment Analysis

## Sanity Checks

| | Lower Bound | Upper Bound | Observed | Passes |
|---|---|---|---|---|
| **Number of cookies** | 0.4988 | 0.5012 | 0.5006 | Yes |
| **Number of clicks** | 0.4959 | 0.5041 | 0.5005 | Yes |
| **Click-through-probability** | 0.08121 | 0.08304 | 0.08218 | Yes |

All of metrics pass sanity check.

## Result Analysis

### Effect Size Tests

| Using Bonferroni correction | Lower Bound | Upper Bound | Statistical significance | Practical significance |
|---|---|---|---|---|
| **Gross conversion** | -0.0291 | -0.0119 | Yes | Yes |
| **Net conversion** | -0.0116 | 0.0018 | No | No |

For gross conversion, lower bound and upper bound don't include 0, thus it is statistically significant. Since confidence interval doesn't include the practical significance boundary, we find it is practical significant.

For Net conversion, lower bound and upper bound include 0, thus it is statistically significant. . Since confidence interval include the practical significance boundary, thus it is not practical significant.

### Sign Tests

| Using Bonferroni correction | P-value | Statistical significance |
|---|---|---|
| Gross conversion | 0.0026 | Yes |
| Net conversion | 0.6776 | No |

For gross conversion, p-value is less than 0.05, thus we think it is statistically significant.
For Net conversion, p-value is greater than 0.05, thus it is not statistically significant.

### Summary

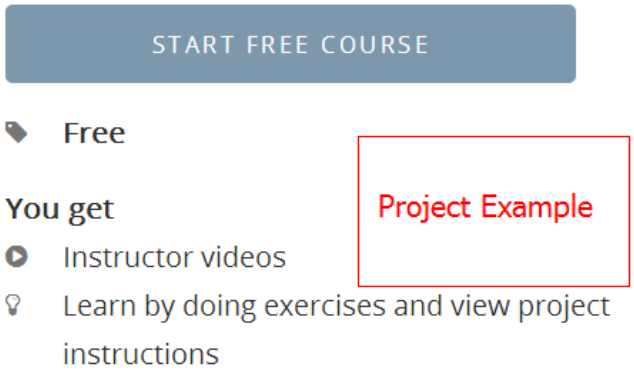
We have two metrics in out experiment and if we want to prove our hypothesis, we want one of them change significantly and another don't change a lot. In other words, our test can be interpreted as "AND" problem. We tend to have a less conservative result.
Thus we decide to not to apply Bonferroni correction in our experiment.

## Recommendation

First of all, our recommendation should be made if and only if we don't harm revenue. Net conversion can reflect our revenue and its practical significant boundary is (-0.0075, 0.0075). Our result shows that its CI range from -0.0126 to 0.0028. The change of net conversion is more likely be negative. In negative side, it's clearly significant compared with practical boundary and this is not significant for positive side.
This shows if we launch our experiment, we have high risk of harming revenue. Thus even though the result match our hypothesis, I don't recommend launching this experiment.

# Follow-Up Experiment

For reducing early cancellations, I think we can add an extra button which can show an example of awesome final project in enrolling page of particular class. Compared with others, an actual final project can tell students what they need to learn to get the credit more clearly. In the other hand, it's also a great method of making students be interested in course. Like this:




The hypothesis is when we add this button of final project, there will be less students who enroll the free trial and the number of students who decide to pay for the class will not change or increase.

We will use cookies as unit of diversion.

Our invariant metrics include:

1. Number of cookies: number of unique cookies to view the course overview page.
2. Number of class clicks: number of unique cookies to click particular class to see class description. (enrolling page)
3. Class-click-through-probability: Number of class clicks divided by Number of cookies.
4. Number of example clicks: number of unique cookies to click particular class to see class description
5. Project-click-through-probability: Number of example clicks divided by Number of cookies.

Our evaluation metrics include:

1. Probability of enrolling: number of students who enrolling a particular class divided by number of unique cookies to click particular class.
2. Probability of paying: number of students who paying for a particular class divided by number of unique cookies to click particular class.

If probability of enrolling don't change with increase of probability of paying probability of paying don't change with probability of enrolling decrease, we can launch this experiment.