

# Experiment Design

## Metric Choice

### Invariant

- Number of cookies
- Number of clicks
- Click-through-probability

### Evaluation

- Gross Conversion
- Net Conversion

### Explanation:

- Number of cookies
  - This is the number of cookies viewing the course overview page. This happens before the user would see any change that was made due to the test, so it should be invariant and it would not be useful for evaluation.
- Number of user-ids
  - This isn't good as an invariant metric because the number of users enrolling in the free trial may change due to the extra step introduced by the test. However, it isn't very useful as an evaluation metric since we are more interested in increasing the rate of users completing the free trial and remaining enrolled, rather than just a count of those starting the free trial.
- Number of clicks
  - This is another metric that would be measured before the change being tested, making it an invariant metric and not an evaluation metric.
- Click-through-probability
  - This is also another metric that would be measured before the change being tested, making it an invariant metric and not an evaluation metric.
- Gross Conversion
  - This is a good evaluation metric. It isn't the goal of the test (we care about users remaining enrolled and thus making payments rather than just starting the free trial) but it will show us how the change may affect the rate of users enrolling in the free trial. Ideally it is not a significantly negative change, but we may ignore such a change.
- Retention
  - This appears to be a good evaluation metric- we are hoping that the change being tested will increase this ratio. However, sizing calculations indicate that it would require over 4.5 million pageviews to test with sufficient power and that would make the experiment take far too long.
- Net conversion

- This is also a good evaluation metric- the goal of the change is to increase the rate of users that make a payment and making this into a ratio between payment and clicking makes sense since the experiment change occurs in between those two steps. It should tell us almost the same information as retention would, but it won't require nearly as many pageviews.

In order to launch the change I would like to see a practically significant change in net conversion. It would also be good to check the results of gross conversion, but whether or not we detect a change in that metric doesn't actually matter if the net conversion increases enough.

## Measuring Standard Deviation

- Standard Deviation of Gross Conversion = 0.0202
- Standard Deviation of Net conversion = 0.0156

It's possible that the analytic estimate underestimates the true variability in both metrics. We are dealing with ratios rather than individual counts in a distribution, so the actual distribution may be more complex than a normal distribution. Additionally, in both metrics the unit of analysis (cookies) is different from the unit of diversion (user id). One user could result in multiple cookies (on different devices or in different browsers), invalidating the independence assumption.

## Sizing

### Number of Samples vs. Power

The Bonferroni correction will not be used since I am not looking for one or the other metric to indicate a change (meaning there is no multiple testing problem)- I only actually care about net conversion changing. I will need 685,275 total pageviews for the experiment in order to have enough power to detect a 0.75% increase in net conversion. Detecting a 1.0% change in gross conversion requires slightly fewer pageviews (645,875), but I will need to go with the larger of the two numbers.

### Duration vs. Exposure

I will divert 100% of traffic for 18 days.

I don't think this will be a very risky experiment. The hypothesis is that the change will not significantly reduce the number of users that continue past the free trial. It may even encourage some users to stick around longer (make more payments) since coaches may have more time. I think it is better to finish the experiment sooner, so diverting all traffic is fine.

## Experiment Analysis

### Sanity Checks

| Metric                    | Lower Bound<br>95% CI | Upper Bound<br>95% CI | Actual | Passed? |
|---------------------------|-----------------------|-----------------------|--------|---------|
| Number of cookies         | .4988                 | .5012                 | .5006  | Yes     |
| Number of clicks          | .4959                 | .5041                 | .5005  | Yes     |
| Click Through Probability | -.0013                | .0013                 | .0001  | Yes     |

## Result Analysis

### Effect Size Tests

| Metric           | Lower Bound<br>95% CI | Upper Bound<br>95% CI | Statistically<br>Significant? | Practically<br>Significant? |
|------------------|-----------------------|-----------------------|-------------------------------|-----------------------------|
| Gross Conversion | -.0291                | -.0120                | Yes                           | Yes                         |
| Net Conversion   | -.0116                | .0019                 | No                            | No                          |

### Sign Tests

| Metric           | P value | Statistically Significant |
|------------------|---------|---------------------------|
| Gross Conversion | .0026   | Yes                       |
| Net Conversion   | .6776   | No                        |

## Summary

I did use the Bonferroni correction since I was running two tests and looking for significance of either/both of them (for comparison to the other significance results). In this case a statistically significant result would have a p value of less than 0.025. There does not appear to be any discrepancy- the gross conversion shows a statistically significant decrease while the net conversion difference is not statistically significant.

## **Recommendation**

I would recommend against launching this change since we have had a negative effect on one metric (although this was expected) and no statistically significant effect on the metric that we were hoping to influence.

We expected that the gross conversion would decrease since there is now an additional step between clicking the button and enrolling which may discourage some potential users. We did detect a statistically and practically significant decrease for that metric. This could be excused if we improved the net conversion but we did not find a statistically significant change for that metric.

It is possible that the experiment change is a good idea, but the specifics are wrong. It may be worth running follow up experiment showing the message at different rates. For example, asking the user how many hours a month or how many days a week they can dedicate some time. It may be better to ask how many hours the user plans on dedicating during the free trial period specifically, and to encourage them to delay enrolling in the free trial if needed.

## **Follow-Up Experiment**

### **Experimental change:**

I think that some portion of students may sign up for the free trial with good intentions but get busy at some point during the two weeks and procrastinate. They don't use the website enough to become committed and cancel before they get charged. I think sending one or more motivational reminder emails to users that have enrolled in the free trial may motivate them to continue. For a specific experiment, there could be an email sent one week into the free trial with stories from successful graduates about how Udacity helped them in their careers.

### **Hypothesis:**

Sending a motivational reminder email to users enrolled in the free trial will help some of them avoid procrastination, become engaged in using the site, and eventually become paying users.

### **Metrics:**

Retention is a great metric to evaluate. However, we should increase the limit for a practical change in order to lower the amount of pageviews we need- 2 or 3 % would work much better than 1%.

### **Unit of Diversion:**

The unit of diversion will have to be user ID since the change will only affect users enrolled in the free trial.