# A/B Testing Project

*Prasun Srivastava*

*2 December 2016*

## Experiment Design

### Metric Design

- Invariant Metrics:
    1. Number of Cookies - Since the change that we are testing doesn't occur until the user reaches the course overview page, it should not affect how many users reach the course overview page.
    2. Number of clicks - Since, the change that we are testing does not occur until the user clicks on the free trial button, it should not affect how many users click on the free trial button.
    3. Click Through Probability - Since, the change that we are testing does not occur until the user clicks on the free trial button, it should not affect how many users click on the free trial button.
- Evaluation Metrics:
    1. Gross Conversion - Since we are interested in evaluating the change in the number of users enrolling in the course after a question is being asked about the time commitment, this metric should be affected. As per the hypothesis, it is expected that the gross conversion will go down after the launch of the experiment since users will have a better understanding of time commitment.
    2. Net Conversion - This metric measures how many users continue past the free trial period. Since, we expect the enrollment to be affected by the question about the time commitment, this metric should also be affected. As per the hypothesis, it is expected that net conversion will go up as informing users about time commitment prior to enrolling will set an expectation beforehand.
- Unused Metrics:
    1. Number of User-ids - This is the number of users who enroll in the free trial. Since, the experiment is going to test the effect of additional information about time commitment at the tme of enrollment, this metric is not going to have same distribution for experiment and control groups. Hence, it can not be used as an invariant metric. Also, since it has not been normalized and the number of visitors between the experiment and control groups can be different, it has not been chosen as an evaluation metric.
    2. Retention - It is not a good invariant metric as the number of users who enroll in the free trial is dependent on the experiment. It was an excellent choice for an evaluation metric as it is directly dependent on the effect of the experiment. However, it was not chosen as a metric because its unit of analysis is user-id which is different from the unit of diversion for the experiment (cookies). Also, in order to run the experiment with the given significance level and power, we would require over 4M pageviews and the experiment would have to run for a long time rather than a couple of weeks.

### Measuring Standard Deviations

- Standard Deviations
    1. Gross Conversion:
        - N = 400

        - $SE_{GrossConversion} = 0.0202306$

    2. Net Conversion:
        - $SE_{NetConversion} = 0.0156015$

3. Since, the unit of diversion is a cookie and both Gross Conversion and Net Conversion have number of cookies as the denominator, the unit of analysis is also cookie. Hence, we do not expect any difference between the analytical and empirical variance.

## Sizing

### Number of Samples vs Power

- Bonferroni Correction : No, Bonferroni correction is not used during the analysis as we want to evaluate changes in multiple metrics and bonferroni correction becomes conservative in case of multiple metrics.
- Pageviews:
  1. Gross Conversion:
     - Baseline Conversion Rate = 20.625%
     - Minimum Detectable Effect(Absolute) = 1%
     - Free Trials per Group(calculated using online calc) = 25835
     - Pageviews per Group = $\frac{25835}{CTR} = \frac{25835}{0.08} = 322937.5$
     - Total Pageviews = 645876
  2. Net Conversion:
     - Baseline Conversion Rate = 10.93125%
     - Minimum Detectable Effect(Absolute) = 0.75%
     - Free Trials Per Group(calculated using online calc) = 27413
     - Pageviews per Group = $\frac{27413}{CTR} = \frac{27413}{0.08} = 342662.5$
     - Total Pageviews = 685325
       Since, we have to evaluate both gross conversion and net conversion, we will take maximum among the two. Hence, total number of pageviews required is 685325.

### Duration vs Exposure

- Fraction of Traffic - We will divert 100% of the traffic for the test as the change is not risky because we are providing additional information to the users in order to provide a better user experience.
- Days Experiment should be run:
  - Pageviews per Day = 40000
  - Number of Days = $\frac{TotalPageviews}{PageviewsperDay}$ = 68325 / 40000 = 17.13 = 18 Days

## Experiment Analysis

### Sanity Checks

1. Number of Cookies:
   - p = 0.5
   - Standard Error = 0.0006
   - Confidence Interval = (0.4988, 0.5012)
   - $\hat{p} = 0.5006$
   - Since, estimate belongs to the confidence interval, sanity check is passed.
2. Number of clicks on "Start free Trial":
   - p = 0.5
   - Standard Error = 0.0021
   - Confidence Interval = (0.4959, 0.5041)
   - $\hat{p} = 0.5005$
   - Since, estimate belongs to the confidence interval, sanity check is passed.
3. Click Through probability:
   - p = 0.0821
   - Standard Error = 0.0005

- Confidence Interval = (0.0812, 0.0830)
- $\hat{p} = 0.0822$
- Since, estimate belongs to the confidence interval, sanity check is passed.

**Result Analysis**

**Effect Size Tests(All calculations available in spreadsheet)**

1. Gross Conversion:
   - Confidence Interval = (-0.0291, -0.0120)
   - Statistically Significant
   - Practically Significant
2. Net Conversion:
   - Confidence Interval = (-0.0116, 0.0019)
   - Not Statistically Significant
   - Not Practically Significant

**Sign Test(Calculated from online calculator)**

1. Gross Conversion:
   - Success = 19, Total = 23
   - p value = 0.0026
   - Statistically Significant
2. Net Conversion:
   - Success = 13, Total = 23
   - p value = 0.6776
   - Not Statistically Significant

## Summary

Bonferroni correction is not used in the analysis as we wanted to evaluate significant change in two metrics ie. Gross Conversion and Net Conversion. Bonferroni correction is used to adjust for scenarios where any change is significant.
There were no discrepancies between the effect size tests and sign tests.

## Recommendation

We were expecting that the change will lead to decreased enrollments after the message as users will be better informed if they could not commit the time required for the course. Since, the change in gross conversion is statistically significant, the test is working as expected. On the other hand, we wanted that students who enroll go on to complete the course which we measured using a proxy (Net Conversion ie. people who continue after free trial might also finish the course). Although the change is not statistically significant, the confidence interval lower bound is negative indicating that there is a decrease in the net conversion. Also, the confidence interval of the net conversion does include the negative of the practical significance boundary. So, there is a chance that this change might lead to a decrease in net conversion and hence, revenue. Hence, I do not recommend this change to be implemented.

## Follow Up Experiment

Sometimes, students might get frustrated because they might not have the essential prerequisites and thus quit the course. If we add a lesson covering the prerequisites as a review, students might be able to catch up

and more likely to complete the course.

**Hypothesis**

If a student is given a review of the prerequisites, he might be better equipped to take the course and hence motivated to complete the course.

**Unit of Diversion**

User ID

**Evaluation Metric**

- Drop out Rate after 1 month - $\frac{Number of Users who leave course after a month}{Number of Users who join the Course}$
- Idea is if the student is going to pay for the second month, this means that he is motivated to complete the course. Also, the unit of analysis and unit of diversion will be same in this case.