# P7: Design an A/B Test by Prem Mithilesh Morampudi

# Experiment Design

## Metric Choice

* **Invariant Metrics:** Number of Cookies, Number of Clicks.
* **Evaluation Metrics:** Gross Conversion, Retention, Net conversion.
* **Number of Cookies:** This happens even before the user see’s the Experiment so this is a good Invariant Metric and is Independent.
* **Number of user Id’s:** As this is dependent on the experiment, this is not good Invariant Metric. This is also not a good Evaluation metric as it is just a Numerical Representation and I believe the Gross conversion does a better job at counting the number of users.
* **Number of Clicks:** This happens even before the user see’s the Experiment so this is a good Invariant Metric and is Independent.
* **Click-through Probability:** This happens even before the user see’s the Experiment so this is a good Invariant Metric and is Independent. The Click-Through Probability is not considered as an Invariant Metric as this metric is just a ratio between the other invariant metrics. The Click through probability would make it a better invariant metric compared to the number of clicks, as it normalizes to the size of the control and experiment group.
* **Gross Conversion:** Not a good Invariant metric as it is dependent on the Experiment and a good Evaluation metric as it is Influenced by the effect of the Experiment. The experiment should have a negative impact on the Gross Conversion as the experiment is trying to eliminate the student who can’t afford enough time for the course later.
* **Retention:** Not a good Invariant metric as it is dependent on the Experiment and a good Evaluation metric as it is Influenced by the effect of the Experiment. The experiment should have a Positive impact on the student Retention as they at least have one lesser reason to quit the enrollment.
* **Net Conversion:** Not a good Invariant metric as it is dependent on the Experiment and a good Evaluation metric as it is Influenced by the effect of the Experiment. The experiment should not have a negative impact on the Net Conversion as they at least have one lesser reason to quit enrollment.

## Measuring Standard Deviation

|  |  |
| --- | --- |
| Gross Conversion | 0.0202 |
| Retention | 0.0549 |
| Net Conversion | 0.0156 |

For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the empirical variability, or whether you expect them to be different. Briefly give your reasoning in each case.

Cookie has been the unit of evaluation and the unit of diversion for both Gross Conversion and Net conversion and so, the analytical Estimate should match the empirical standard deviation of the experiment.

The unit of diversion and unit of analysis for the Net Retention are Different and so the Analytical Estimate and the Empirical Standard Deviation will differ for Retention.

## Sizing

### Number of Samples vs. Power

No, I will not be using the Bonferroni Correction as the three evaluation Metrics are likely to be covariant.

The calculations were done with an Alpha = 0.05, Beta= 0.2 and

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metric | Dmin | Baseline Estimates | Samples Needed | Pageviews | Required Page Views |
| Gross Conversion Rate | 0.01 | 0.20625 | 25,835 | 322,937.5 | 645,875 |
| Retention | 0.01 | 0.53 | 39,087 | 2,370,606 | 4,741,212 |
| Net Conversion | 0.0075 | 0.1093125 | 27,413 | 342,662.5 | 685,325 |

Considering the Retention rate as an Evaluation Metric will significantly take a longer period as 4.75 Million Pageview will require a lot time. Therefore, dropping Retention, we consider only Gross Conversion Rate and Net Conversion as Evaluation Metrics.

The Total Number of Required Pageviews are: 685,325.

### Duration vs. Exposure

Exposure:

100% of the traffic will be exposed towards the new experiment as there is not any risk involved in this experiment.

Duration:

Running the experiment on 100% of the traffic would take around 18 days to reach 685,325 Pageviews.

The change is not Risky when executed correctly, As There is neither any chance of students being Hurt nor Dealing with sensitive Data of the students. Thus, I recommend the experiment to be run on 100% of the traffic as it has no risk involved.

# Experiment Analysis

## Sanity Checks

**Number of Cookies:**

|  |  |
| --- | --- |
| **Type** | **Data** |
| Standard Deviation | 0.0006018 |
| Margin of Error | 0.0011796 |
| Control Group | 345543 |
| Experimental Group | 344660 |
| Lower Bound | 0.4988 |
| Upper Bound | 0.5012 |
| Observed Value | 0.5006 |

The Observed value is within the 95% Confidence Interval. Sanity check Passed.

**Number of Clicks:**

|  |  |
| --- | --- |
| **Type** | **Data** |
| Standard Deviation | 0.021 |
| Margin of Error | 0.0041 |
| Control Group | 28378 |
| Experimental Group | 28325 |
| Lower Bound | 0.4959 |
| Upper Bound | 0.5041 |
| Observed Value | 0.5004 |

The Observed value is within the 95% Confidence Interval. Sanity check Passed.

## Result Analysis

### Effect Size Tests & Sign Tests

The Evaluation Metrics were as the follows:

**Gross Conversion:**

|  |  |
| --- | --- |
| Confidence interval | [-0.0291, -0.0120] |
| Statistically Significant | Yes |
| Practically Significant | Yes |
| P Value (2 Tailed) | 0.0026 |

Since the P- Value of Gross Conversion is less that alpha value (0.0026<0.05), The Difference is Statistically Significant.

**Net Conversion:**

|  |  |
| --- | --- |
| Confidence interval | [-0.0116, -0.0018] |
| Statistically Significant | No |
| Practically Significant | No |
| P Value (2 Tailed) | 0.6776 |

Since the P- Value of Net Conversion is Greater than the alpha value (0.6776>0.05), The Difference is Not Statistically Significant.

### Summary

I chose not to use the Bonferroni correction because the Decision is based on both the Gross and Net Conversion metrics as the Hypothesis requires that both the metrics to Significant. We need both the metrics to be significant to launch the metric.

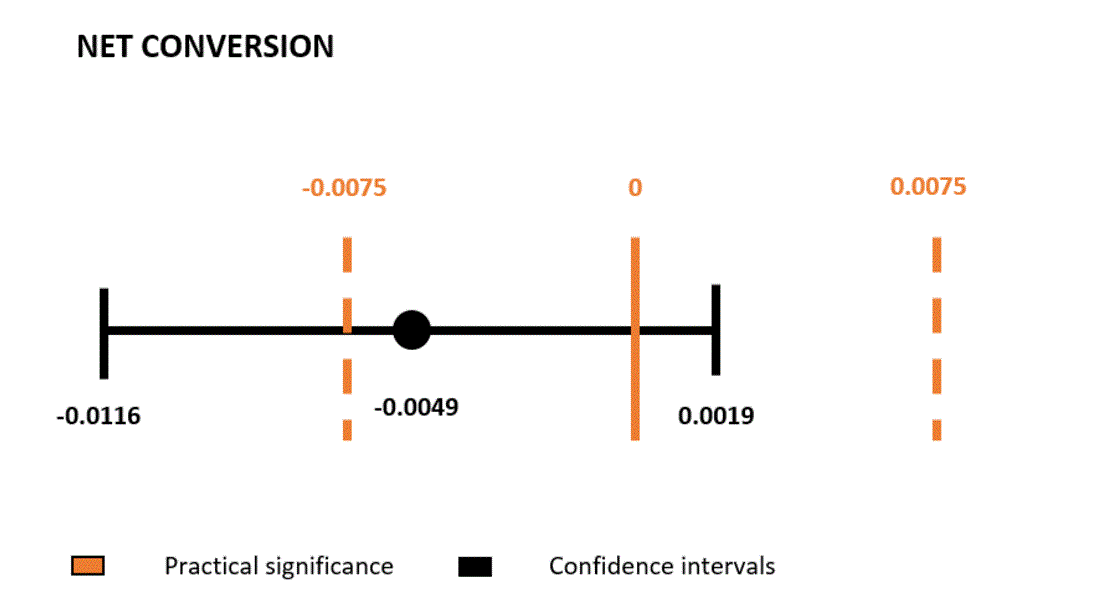
## Recommendation

The hypothesis was that this 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 this hypothesis held true, Udacity could improve the overall student experience and improve coach’s capacity to support students who are likely to complete the course.

The Significance of the Experiment is as it can be clearly understood and Predicted from the hypothesis that the Gross Conversion of Students will be considerably reduced which was observed to be Practically and Statistically significant.

The Gross Conversion is found to be Practically and Statistically Significant which is a good go and Discourages Student who couldn’t spend enough time to sign up so Udacity will have less Frustrated Students who cancel the Subscription later and Thus Increase the Student Retention Rate.

The Net conversion is surprisingly found Not be either Statistically or Practically Significant and the C.I confidence interval for net conversion contains the negative of practically significant value -0.0075. The net conversion **might** have diminished by an amount that matters to the business.



I **do not** suggest to launch the experiment based on the above-mentioned factors.

# Follow-Up Experiment

The follow up experiment is required to propose a Single practical change to be tested to reduce the number of frustrated students that abandon early. The following might be some potential reasons that leads to frustration amongst the free trail Students:

* Lack of knowledge about the course structure, Projects involved.
* Not being able to commit enough time to the course.
* Not enough zeal or motivation to progress regularly.
* Lack of Competitive Sprit/ Environment.

I might not be able to address all the problems potentially faced by the user, but for the follow up experiment Let us consider one of the problem. I suggest we introduce a Video in place of this experiment which gives a brief introduction about the Pre-Requisites, Course structure, Commitments, forums, slack, One-on-one Appointments, Submissions, Reviews etc., which I think would help a lot of people in getting to know what is required to successfully complete the course.

The experiment is Launched as a Pop up Video after the user click’s the “Start free trail” button and before they check out to enroll into the course.

My Hypothesis here to Introduce a Video which gives a Brief Introduction about the Course, thus creating awareness amongst the students reducing the number of frustrated students who left the free trail early because they Didn’t have enough time without Significantly Reducing the Number of Students to continue past the free trail and eventually complete the course.

**Unit of Diversion:** Cookies, As the Experiment starts after a user enrolls into the course and before the free trial ends.

**Invariant Metrics:**

* **Number of Cookies:** This happens even before the user see’s the Experiment so this is a good Invariant Metric and is Independent.
* **Number of Clicks:** This happens even before the user see’s the Experiment so this is a good Invariant Metric and is Independent.

**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.
* **Net conversion:**That is, number of user-ids to remain enrolled past the 14-day boundary divided by the number of unique cookies to click the "Start free trial" button.

If the Gross conversion is negative/Reduced due to the experiment and the Net conversion is not Decreased, we can go for the Launch of the new experiment.

# References

* <http://www.evanmiller.org/ab-testing/sample-size.html>
* <http://www.utdallas.edu/~herve/Abdi-Bonferroni2007-pretty.pdf>
* <http://www.stat.berkeley.edu/~mgoldman/Section0402.pdf>
* <https://review.udacity.com/#!/reviews/404418>