

A/B Test – Vishnu Devarakonda.

About the Experiment: Udacity page has two options “Start Free Trial” and “Access Course Materials”. Udacity chose to introduce a new change, at the time of clicking Start Free Trial, that the user can spend more than 5 hours per week or not. Based on the input provided by user, Udacity redirects the pages. If the user enters more than 5 hours per week, then Udacity has a chance to take care of such potential users by providing co-ordination between coaches and them.

Experiment Design

Unit of Diversion is cookie as it is unique and once the student enrolls, then tracking is done by user-id.

Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here. (These should be the same metrics you chose in the "Choosing Invariant Metrics" and "Choosing Evaluation Metrics" quizzes.)

Invariant Metrics: [Number of cookies](#), [Number of clicks](#), [Click-through-probability](#).

Number of cookies – As this is unit of diversion and is distributed equally among two groups, this is appropriate for invariant metric. Also it is taken well before the experiment is launched.

Number of Clicks – This is also taken before the experiment and would not impact clicking the “Start Free Trial” button.

CTP is nothing but a reflection of clicks and pageviews.

Evaluation Metrics: [Gross Conversion](#) and [Net Conversion](#).

The above two are chosen as they have strong relation with each other.

Gross Conversion – showing 5 hours per week pop-up influences the student and this is what we want to observe in this experiment. Hence it is considered as good evaluation metric.

Net Conversion – The ratio of students what pay over students seeing start free trial is dependent on the pop-up “5 hours per week time allocation” and this is what we want to observe in this experiment. Hence it is considered as a good evaluation metric.

Measuring Standard Deviation

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

For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the the empirical variability, or whether you expect them to be different (in which case it might be worth doing an empirical estimate if there is time). Briefly give your reasoning in each case.

Evaluation Metric	Standard Deviation
Gross Conversion	0.0202
Net Conversion	0.0156

As both the above evaluation metrics have number of cookies, which is our unit of diversion in the experiment, we can safely proceed using this analytical estimate of variance.

Sizing

Number of Samples vs. Power

Indicate whether you will use the Bonferroni correction during your analysis phase, and give the number of pageviews you will need to power your experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

- *I did not use the Bonferroni correlation.*
- *Number of pageviews needed is 685325.*

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.)

- *Fraction of traffic exposed is 0.7 – i.e., 70% traffic exposed.*
- *Number of days needed to run the experiment – 25*

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?

Choosing 100% of the traffic for this experiment is good and dividing 50% to control and experiment groups for more accurate results. If we are going to decrease the traffic, we have to increase the length of the experiment which is not a desirable constraint in A/B testing.

Experiment Analysis

Sanity Checks

For each of your invariant metrics, give the 95% confidence interval for the value you expect to observe, the actual observed value, and whether the metric passes your sanity check. (These should be the answers from the "Sanity Checks" quiz.)

For any sanity check that did not pass, explain your best guess as to what went wrong based on the day-by-day data. **Do not proceed to the rest of the analysis unless all sanity checks pass.**

- Number of cookies – 0.5006 with bounds [0.4988, 0.5011]
- Number of clicks – 0.5004 with bounds [0.4958, 0.5041]

As the observed values are within the bounds, we can safely assume that sanity check is passed with 95% confidence interval.

Result Analysis

Effect Size Tests

For each of your evaluation metrics, give a 95% confidence interval around the difference between the experiment and control groups. Indicate whether each metric is statistically and practically significant. (These should be the answers from the "Effect Size Tests" quiz.)

Gross conversion – [-0.0291, -0.0119]

This result is statistically significant because this interval/range doesn't include zero. Also it is practically significant

Net conversion – [-0.0116, 0.0018]

This result is not statistically significant and not practically significant as well because the range includes zero.

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.)

I have used the online calculator given in the instructor notes to get p-value.

	Number of Days	p-value	Statistically significant (< alpha)
Gross Conversion	4	0.0026	Yes
Net Conversion	10	0.6776	No

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.

Bonferroni correction is not used because ideally it is used to check for Type 1 Errors. Here the null hypothesis is there is no difference in evaluation metrics in both the groups which is not the

case when we see the results statistically. The requirement for launching the experiment is that the null hypothesis must be rejected for all the evaluation metrics and that the difference between branches must meet or exceed the practical significance threshold. Because our acceptance criteria requires statistically significant differences for all the evaluation metrics, the use of the Bonferroni correction is not appropriate.

Based on the results above gross conversion will decrease and net conversion will not be impacted significantly.

Recommendation

Make a recommendation and briefly describe your reasoning.

The experiment is to determine whether “5-hour study time per week” pop-up should be launched or not. From the above results we could see that there is decrease in gross conversions and there is no statistically significant impact on net conversion, launch of this change is not recommended. Before abandoning this idea, we can consider other options/designs.

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.

Description: The idea is to increase the conversion rate of students and decrease the number of frustrated students. Incorporating one on one meeting or coach session with the users who clicked on start free trial is likely to increase the net conversion rate.

Null Hypothesis: arranging coach session or one on one meeting will not increase net conversion rate.

Unit of Diversion: user-id.

Invariant metrics: user-id as it will be equally distributed among control and experiment groups and considered before the experiment.

Evaluation metrics: net conversion rate. A statistically significant increase in this metric shows us the successful outcome of the experiment and we can launch this experiment.