

Project 07: A/B Testing Final Project

Experiment Design

Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here.

- Invariant metrics
 - Number of cookies
 - Number of clicks
 - Click-through-probability
- Evaluation metrics
 - Gross conversion
 - Retention
 - Net conversion

For each metric, explain both why you did or did not use it as an invariant metric and why you did or did not use it as an evaluation metric. Also, state what results you will look for in your evaluation metrics in order to launch the experiment.

- Number of cookies: Number of users that view the website should not change as we change the way rendering of start free trial page, because the users haven't seen the page before deciding to visit the page. Which is the reason it is an invariant metric.
- Number of clicks: This is similar to number of cookies; it is an invariant metric because the click happens before user sees the experiment, and they are independent from it.
- Number of user-ids: This is not a good invariant because the number of users who enroll in the trial is dependent on the experiment and evaluation metric is also not good because number of visitors may be different between the experiments, this could change the results.
- Click-through-probability: It is an invariant metric because the click happen before the user sees the experiment, and they are independent from it.
- Gross conversion: This is not an invariant metric because the number of users who enroll in the free trial is dependent on the experiment. It is an evaluation metric because it is directly dependent on the effect of the experiment. It shows whether we were able to manage in decreasing the cost of enrollments that aren't like to become paying customers.
- Retention: This is not an invariant metric because the number of users who enroll in the free trial is dependent on the experiment. This is an evaluation metric because it is directly dependent on the effect of the experiment. This will show positive financial outcome.
- Net conversion: This is not an invariant metric because number of user who enroll in the free trial is dependent on the experiment. It is an evaluation metric because it directly dependent on the effect of the experiment. This will show positive financial outcome.

We will look at Gross conversion and Net conversion. Gross conversion will show us whether we lower our costs or not. Net conversion will show how it will affect the revenue. In order to launch we require Gross conversion to have a practically significant decrease and Net conversion can go up or stay the same but not go down.

Measuring Standard Deviation

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

Gross conversion: $\sqrt{p * (1-p) / N}$
 $\text{round}(\text{np.sqrt}((.206250*(1-.206250))/(5000*3200/40000)),4)$
0.0202

Retention: $\sqrt{p * (1-p) / N}$
 $\text{round}(\text{np.sqrt}((.53*(1-.53))/(5000 * 0.08 * 0.20625)),4)$
0.0549

Net Conversion: $\sqrt{p * (1-p) / N}$
 $\text{round}(\text{np.sqrt}((.109313*(1-.109313))/(5000*3200/40000)),4)$
0.0156

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.

Gross conversion:

This is using number of cookies as denominator, which is also unit of diversion. Unit of diversion is equal to unit of analysis. This indicates the analytical estimate would be comparable to the empirical variability.

Net conversion:

Similar to Gross conversion this is using number of cookies as denominator, which is also unit of diversion. Unit of diversion is equal to unit of analysis. This indicates the analytical estimate would be comparable to the empirical variability.

Retention:

The denominator is number of users enrolled, which is not same as unit of diversion. The unit of analysis and the unit of diversion are not the same therefore the analytical and empirical estimates are different.

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 you experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

Bonferroni correction was not used in analysis phase. Metrics in the test has high correlation and Bonferroni correction is too conservative. I have iterated my evaluation metric and used gross conversion and net conversion as evaluation metrics. We need 685324 page views to power the

experiment with these metrics. If we use the retention as a evaluation metric, Number of required pages will be too large. Which is the reason to drop the retention as a evaluation metric.

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.)

We would divert 70% of the traffic to the experiment. Given the experiment will take 25 days which is achievable time.

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

This experiment is not a risky one because it doesn't affect the existing paying users, and it is simple It may impact new enrollments, diverting 100% of the traffic may not be advisable. If we reduce the diversion traffic to half, then the length of the experiment will be doubled. Risk of diverting the traffic would be that users would be seeing the feature before evaluation. There is no chance that anyone getting hurt because of the duration of our experiment. If we have time, collecting more data to verify will not hurt. As per this experiment we are not collecting any sensitive data.

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.)

Number of cookies:

0.4988, 0.5012 – observed 0.5006 – Pass

Number of clicks:

0.4959, 0.5041 – observed 0.5005 – Pass

Click through probability:

0.0821, 0.0830 – observed 0.822 – Pass

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.

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, -0.0120 – Statistically significant, Practically significant

Net conversion:

-0.0116, 0.0019 – not Statistically significant, Not practically significant

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.)

Gross conversion:

0.0026 – Statistically significant

0.6776 – Not statistically significant

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.

I didn't use a Bonferroni correction. It might be useful to apply the Bonferroni correction if we have multiple metrics but we need only one of them to meet our criteria in order to launch. If we have multiple metrics, but would need all of them to meet some criteria in order to launch. In most of the cases we need more metrics to meet some criteria in order to launch. If a single metric demonstrated a false positive, it would not be enough to launch. Our actual risk is that we might reject some metrics by mistake, and that is a type II error. The Bonferroni correction is designed to avoid type I errors when, out of many, it is sufficient that one metric meets the criteria in order to launch. It is useless in cases like ours where, out of many, we need all the metrics to meet criteria in order to launch.

Recommendation

Make a recommendation and briefly describe your reasoning.

Gross conversion turned out to be negative and practically significant. This is a good outcome because we lower our costs by discouraging trial signups that are unlikely to convert. Net conversion unfortunately ended up being statistically and practically insignificant and the confidence interval includes negative numbers. Therefore, there is a risk that the introduction of the trial screener may lead to a decrease in revenue. We should therefore consider test other designs of the screener before we decide whether to release the feature, or abandon the idea entirely.

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.

Add a new point based incentive program where users would be given points in each and every step of their learning journey. Like completing a quiz or completing a lesson, etc. This feature will help in reducing the early cancellations. It creates more interest and keeps them motivated. This concept is similar to gaming where lessons are like levels.

The hypothesis is that by providing this option, Number of cancellations would be less based on more number of points. We would like to measure evaluation metric retention and Invariant metric user_id. This will test whether the additional option helped to reduce the overall cancellation or not

Unit of diversion would be user_id, because this can be evaluated only after enrollment and after signing in.