Experiment Design (A/B test)

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.)

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.

Answer:

Invariant metrics:

- Number of cookies - This should be the same in both control and experiment group because experiment is after click.
- Number of clicks - This should be the same in both control and experiment group because experiment is after click.
- Click-through-probability - Same reason as above.

Evaluation metrics:

- Gross conversion - Number of enrollment is expected to decrease in the experiment group because the experiment is adding one more step before the enrollment. We need to see how much this change with the experiment.
- Retention - This is expected to increase in the experiment group because this experiment should reduce the number of students who leave after 14 days. However, due to the requirement for a large number of pages, this is excluded in the evaluation.
- Net conversion - Because this experiment reduces the number of enrollments, it is important to know whether the net conversion is affected (reduced) by this experiment.

Not selected as invariant or evaluation metrics.

- Number of user-ids: This won't be the same in both control and experiment group because # of enrollments is likely to be affected by the free-trial screeners. Therefore, it can't be used as an invariant metric. This might provide some information about the difference from the experiment, but it makes more sense to use gross conversion as an evaluation metric, which is normalized by the number of clicks.

In this analysis, we want to test the hypothesis that has two parts

- Free trial screener reduces the number of frustrated students who left the free trial
- Free trial screener doesn't significantly reduce the number of students who eventually continue past 14 days.

To test the first part, I can evaluate gross conversion and see if the free trial screener reduces the number of enrollment. To test the second part, I can evaluate net conversion and see if the free trial screener doesn't significantly reduce the number of students who eventually stay past 14 days. Retention can be also used, but it requires too many pages to have enough power, so I exclude the retention in this analysis.

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.

Answer:

- Gross conversion - 0.0202

Unit of analysis is the cookie, and unit of diversion is the cookie. Because they are the same, it is expected that analytical estimation can be comparable to empirical variability.

- Retention - 0.0549

Unit of analysis is userid and unit of diversion is cookie. Because they are different, it is expected that analytical estimation can be different from empirical variability, and it is worth doing an empirical estimate.

- Net conversion - 0.0156

Unit of analysis is the cookie, and unit of diversion is the cookie. Because they are the same, it is expected that analytical estimation can be comparable to empirical variability.

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.)

Answer:

No, I won't use Bonferroni correction because we will launch the experiment only when all metrics are significant.

The number of page views that I need is 685275. Note that gross conversion and net co versions are considered as evaluation metrics, retention is not considered.

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.)

Answer:

I would divert 100% of traffic to this experiment, and it would require 18 days to run the experiment.

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

I chose to diver 100% of traffic because this experiment would not be risky for Udacity considering that most people who are willing to enroll should expect to spend more than 5 hours per week, and this experiment won't prevent those users from starting a free trial.

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.)

Answer:

Number of cookies - CI: (0.4988, 0.5012), Observed: 0.5006, pass

Number of clicks - CI: (0.4959, 0.5041), Observed: 0.5005, pass

Click-through-probability - CI: (-0.0013, 0.0013), Observed: 0 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.**

Answer:

All sanity check passed.

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.)

Answer:

Gross conversion - CI: (-0.0291, -0.012). Statistically and practically significant

Net conversion - CI: (-0.0116, 0.0019). Statistically and practically insignificant

Sign Tests

For each of your evaluation metrics, do a sign test using the daybyday data, and report the pvalue of the sign test and whether the result is statistically significant. (These should be the answers from the "Sign Tests" quiz.)

Answer:

Gross conversion - p-value: 0.0026, Statistically significant

Net conversion - p-value: 0.6776, Statistically insignificant

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.

Answer:

In this analysis, I didn't use Bonferroni correction. Because we launch the experiment only when all the metrics are significant, we don't need to use Bonferroni correction. Bonferroni collection is used to reduce false positive or the risk to launch the experiment when any metric is significant.

Both effect size test and sign tests are consistent - the difference in gross conversion is statistically significant in both tests, and the difference in net conversion is statistically insignificant.

Recommendation

Make a recommendation and briefly describe your reasoning.

Answer:

Based on the test, this free trial screener improves the gross conversion with practical significance. Net conversion was reduced without statistical significance. However, it (the difference) was in a negative range, which means that it may affect the business even though it is not statically significant. Therefore, the risk is not acceptable in order to launch the experiment.

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.

Answer:

1. Proposed experiment:

At the end of 14 day free trials, students should decide whether to proceed or cancel the enrollments. Some people will have a clear idea about whether they should do or not, but some people may need more time before they make a decision. In this experiment, we will give one more chance to extend the free trial by 7 days when they click "cancel" their free trial. Some students still continue to cancel, but some students who didn't

have enough time to evaluate the degree program, or who are not very sure about enrolling or cancelling may take the opportunities to extend the free trial.

For the students who decide to enroll, we don't need to make any changes.

2. Metric choices:

Invariant metrics:

- Number of cookies - # of unique cookies to view the course overview page.
- Number of user-ids - # of users who enroll in a free trial
- Number of clicks to start a free trial- # of unique cookies to click the start free trial
- Number of clicks to cancel: The number of user-ids who click "cancel" button after 14 days of trial.
- Gross conversion - # of user-ids who complete checkout and enroll / # of users who click "start a free trial"

Evaluation metrics:

- Net conversion (21 days) - # of user-ids who remained enrolled past 21 days / # of unique cookies to click the start free trial

3. Hypothesis:

The hypothesis was that this would give a second chance to students who couldn't make a decision within 14 days, and some students will remain past the 21 days boundary, thus increasing the overall enrollment (net conversion). Because this additional 7 days is only given to students who click "cancel" button, it shouldn't affect the students who already decided to enroll past 14 days after the free trial.

If this hypothesis is true, we can increase the number of enrollment and improve the overall businesses

The unit of diversion is a cookie in this analysis.