

Experiment Design

Metric Choice

I chose number of cookies and number of clicks as invariant metrics. The unique cookies to view the site should not change because of the experiment. Also, the number of clicks on the start free trial button should not change, because the experimental change only comes up after that button is clicked. The number of user-ids, gross conversion, retention, and net conversion all might change because of the experiment, so are not good choices for invariant metrics.

For evaluation metrics, I chose gross conversion and net conversion. Gross conversion is the number of users to complete checkout divided by the number of cookies to click the start free trial button, which could change with the pop-up that appears after clicking the button. Net conversion is the users to remain past the first payment divided by number of cookies to click, which also could change because of the experiment. I originally wanted to use retention as well, but found that required an impractically large/long experiment, so I decided not to use it. Number of users who enroll might change, but we really care about keeping the users past the trial period, so I did not use this as an evaluation metric either. The other metrics should be invariant.

I am expecting a decrease in gross conversion in the experimental group with respect to the control. I would think the extra pop-up would discourage some people who don't have enough time from signing up, while I can't see why it would encourage anyone who would have otherwise not enrolled. For net conversion, I am less certain what to expect. If gross conversion goes down, there are less enrollers per click, which may mean less payers per click. But if the pop-up is having the intended effect, it should only weed out those who would not have made it though the free trial, and should therefore not affect the net conversion rate much. To launch the experiment, I would want to see a drop in the gross conversion, while the net conversion remains the same. This would correspond to less people enrolling per click, but the same amount paying per click, and therefore less people who enrolled but did not end up paying.

Measuring Standard Deviation

The gross conversion rate standard deviation is 0.0202 and the net conversion rate standard deviation is 0.0156. I expect these analytical estimates to be fairly close to empirical variability because the metrics are straightforward and the underlying distributions are probably very close to the binomial distribution.

Sizing

Number of Samples vs. Power

I chose not to use the Bonferroni correction because I am only concerned with two evaluation metrics and I think it would be unnecessarily conservative in evaluating those metrics, especially because they are not independent metrics. At $\alpha = 0.05$ and $\beta = 0.2$, I calculated I would need 685,325 page views to get an appropriate power for this experiment.

Duration vs. Exposure

If I divert half my traffic (fraction = 0.5) to this experiment, I can get it done in 35 days. I think this experiment is very low risk to udacity, because pop-up should only discourage people who weren't going to be able to spend enough time on the courses. It is also a relatively easy hurdle for people to ignore if they want and it does not affect the user experience drastically. It also only affects new users starting a free trial, so would not affect anyone currently taking udacity courses or just browsing the course list. Diverting half the traffic to this experiment would mean getting it done in a little over a month and being able to run other experiments with the other half of traffic at the same time if desired.

Experiment Analysis

Sanity Checks

The 95% confidence interval for proportion of cookies diverted to each group (control vs. experiment) was 0.4988 to 0.5012. The actual proportion diverted to the control group was 0.5006, which passes the sanity check.

The 95% confidence interval for proportion of 'start free trial' clicks diverted to each group was 0.4959 to 0.5041 and the actual proportion diverted to the control group was 0.5005, which passes the sanity check.

Result Analysis

Effect Size Tests

For gross conversion, the 95% confidence interval of effect size (difference between experimental and control group) was -0.0291 to -0.0120. This is statistically significant because the confidence interval does not include zero. It is also practically significant because the effect size is larger (in magnitude) than the practical significance boundary of 0.01.

For net conversion, the 95% confidence interval was -0.0116 to 0.0018. This interval includes zero, so the effect was neither statistically nor practically significant.

Sign Tests

A sign test for the gross conversion rate gave a statistically significant p-value of 0.0026, while a sign test for net conversion gave a p-value of 0.6776, which is not statistically significant. Both results agree with the effect size tests above.

Summary

I did not use the Bonferroni correction because the metrics are not independent and having two metrics only slightly raises the chance of getting a false positive. Both metrics use number of clicks in the denominator, and the numerators are related to users who enroll and get past the free trial, which are certainly somewhat correlated. The Bonferroni correction would be unnecessarily conservative and might lead to rejecting the significance of the change in gross conversion.

Recommendation

Based on these results, I would recommend going ahead with the experimental change in the website. We saw a significant decrease in gross conversion, which means fewer students are enrolling overall, and it was practically significant, affecting over 1% of cookies. This may seem bad until we also consider that we saw no significant change in net conversion, which means the same number of students are continuing past the free trial and paying, regardless of the experiment. This means that some students who would have enrolled in the free trial, but not continued to make a payment, are now not enrolling, probably because they realize they won't have the time to devote to the course(s). This means an overall decrease in unsatisfied/frustrated students who have to quit the free trial, which means a happy user base for the website. Based on this knowledge I would recommend changing the website to include the pop-up about time needed to complete these courses.

Follow-Up Experiment

A follow-up experiment I would run would be to require students (in the experiment group) to watch a brief overview video of the course before enrolling. My aim here would be to target those students who get frustrated because the course is not exactly what they were expecting based on the description. I think a brief introductory video by the instructor would give students a feel for how the instructor teaches, what topics will be discussed, and the general flow of the lectures. Hopefully the student can use that information to better decide if this is the right course for them. In terms of how I would logistically run this experiment, the video would show up as a pop-up after clicking on the button to enroll in a course but before completing enrollment. My hypothesis would be that the experimental group (which views the video) would have a slightly lower rate of completing the enrollment, but a similar rate of continuing past the free trial to a payment. I would set it up very similarly to the above experiment, using gross conversion and net conversion as evaluation metrics. The only difference is I would probably change the denominator to the clicks on the enroll button on the course description page, instead of clicks on the 'start free trial' button. I would use cookie as a diversion metric and then track by user-id once students were enrolled. For invariants, I would want to make sure the page views of the course description page was consistent, as well as clicks on the enroll button. This test could be run on a few different courses, but would not need to be implemented across all courses. If I saw that less users enrolled in a given course, but there was no decrease in paying students, then I would probably recommend this change.

Resources

Sample size

<http://www.evanmiller.org/ab-testing/sample-size.html>

Bonferroni Correction

https://en.wikipedia.org/wiki/Bonferroni_correction

sign tests

<http://graphpad.com/quickcalcs/binomial1/>