Introduction

In this project I designed an experiment and analyzed its results to check whether to implement a new feature on Udacity's website.

At the moment of the experiment Udacity's homepage has two options, "start free trial" and "access course materials", if a user clicks on "start free trial" they would be asked for their credit card information and then they would be enrolled on the free trial for the paid version of the course, otherwise they would still be able to view the videos and take the quizzes, but they would not get coaching support nor a verified certificate and they would not submit their final project for feedback.

The change that is being tested consists on asking the user how much time they have available to devote to the course prior to ask the credit card information, and in case the user indicates less than 5 hours per week, a message would show up saying that usually courses require a higher time commitment for successful completion, and at this point the user would have the option to continue enrolling on the free trial or access the content for free.

The hypothesis is that this change might set clearer expectations for students, thus reducing the number of students that drop out during the first week trial because they don't have enough time - without significantly reducing the number of students to continue past the free trial.

Experiment Design

Metric Choice

The available metrics are the following:

Number of cookies:	number of unique cookies to visit the course overview page. (d_{min}=3000)
Number of IDs:	number of users who enroll in free trial. (d_{min}=50)
Number of clicks:	number of unique cookies to click the "start free trial" button. (d_{min}=240)
Click-Through-Probability:	number of clicks divided by number of cookies. (d_{min}=0.01)
Gross conversion:	number of users IDs to complete checkout and enroll in free trial divided by the number of clicks. (d_{min}=0.01)
Retention:	number of users IDs to remain enrolled past the 14 days boundary (and thus make at least one payment) divided by number of IDs to complete checkout. (d_{min}=0.01)
Net conversion:	number of users IDs to remain enrolled past the 14 days boundary (and thus make at least one payment) divided by number of clicks. (d_{min}=0.0075)

I chose "number of cookies", "number of clicks" and "click-through-probability" as invariant metrics because those are the only metrics that (in case there's no problem with the experiment) shouldn't change since they're not affected in any way by the new feature. The "number of IDs" and the last 3 metrics (which depend on "number of IDs") are affected by the change we are testing, so they wouldn't be good candidates as invariant metrics.

As of evaluation metrics I chose “gross conversion” and “net conversion”, because they best measure what we want to achieve, namely reduce the rate of users enrolling while keeping the rate of users continuing past the free trial the same. I could have chose “retention” too, but it depends on the other two metrics already chosen, and the sample size required to power the experiment using this metric would make the experiment unfeasible.

In order to launch the experiment I will need a significantly lower value of “gross conversion” in the experiment group, while the “net conversion” should not decrease significantly in the experiment group.

Measuring Standard Deviation

The analytic estimates for the standard deviation are 0.0202 for “gross conversion” and 0.0156 for “net conversion”. In both cases being the unit of analysis the same as the unit of diversion (cookies) the analytic and empirical estimates should have similar values.

Sizing

Number of Samples vs. Power

Using an alpha value of 0.05, a beta value of 0.2 and not using the Bonferroni correction in my analysis, I would need 685,325 page views to power my experiment appropriately.

Duration vs. Exposure

Regarding the traffic fraction to divert to this experiment I could use 50% of the traffic to make the change visible just to a part of the users (approximately 25% since the traffic for the experiment is divided into a control and an experiment group), but that would require the experiment to run for over a month (35 days), so in this case, assuming that there’s no other experiment running and that the engineers tested the change, it wouldn’t be risky to run the experiment on the whole traffic so that it would require just 18 days.

Experiment Analysis

Sanity Checks

Below a table containing the 95% confidence interval and the observed value for each invariant metric. It can be noticed that all sanity checks passed and this allows us to continue to the next step of the analysis.

Metric	Confidence Interval	Observed Value	Passes
number of pageviews	(0.4988, 0.5012)	0.5006	yes
number of clicks	(0.4959, 0.5041)	0.5004	yes
click-through-probability	(-0.0013, 0.0013)	0.0001	yes

Result Analysis

Effect Size Tests

The confidence intervals built around the difference between the experiment and the control group shows that the gross conversion (95% CI [-0.0291, -0.0120]) is statistically and practically

significant, while the net conversion (95% CI [-0.0116, 0.0019]) does not show any significant difference, and this results correspond to the change we wanted to obtain with the new feature, namely reduce the rate of students enrolling while not decreasing the rate of completion of the free trial.

Sign Tests

The sign test agrees with the effect size test, so even here the results for gross conversion (p=0.0026) are statistically significant, while the results for net conversion (p=0.6776) are not.

Summary

I didn't use the Bonferroni correction because in this case it wouldn't be useful. The reason for this is that the Bonferroni correction aims at reducing the chance of a type I error, while since we need all metrics to meet our requirements, we're already being conservative and our main concern is the chance of a type II error.

Recommendation

The effects of the change would result in a decrease of users enrolling, users that we can assume would have low probability of completing the course. Regarding the rate of users completing the first week trial, the results didn't show any statistically significant difference, but the confidence interval included the negative of the practical significance boundary which means the change could actually negatively affect the net conversion.
In this case my recommendation is to not launch the change as it could be risky, meaning that the change could significantly decrease the number of students continuing past the free trial.

Follow-Up Experiment

A different experiment, with the same purpose would be adding a quiz section before the enrollment to check whether a user meets the prerequisites, and in case they don't they would be suggested to review the prerequisites before starting the course (with eventual links to useful resources).

The hypothesis here is be that this change would make the students that don't meet the prerequisites aware that it could be more difficult and frustrating for them to proceed in the course without first reviewing the suggested topics, so it might reduce the number of students that drop out because they think that the course it's too difficult.

Even in this case I would use cookies as unit of diversion and after users enroll I would track them using user-ids. I would measure the number of cookies clicking the "start now" button, the number of ids enrolling and the number of ids completing the free trial. I would then use those metrics to calculate gross and net conversion and I'd expect a significantly lower gross conversion on the experiment group, while depending on the length of the experiment I'd expect the net conversion to not differ significantly or to be significantly higher, due to the users that after not enrolling review the suggested topics and then come back to starts the course.

Resources

- Sample Size Calculator

- Sign and Binomial Test