Michael Collison
Udacity A/B Test

Experiment Design

Metric Choice

Number of Cookies - Invariant Metric

The number of cookies that view the course overview page should not change between the experimental and control groups, because the experiment only affects users that click the “start free trial” button.

Number of Clicks - Invariant Metric

Like the number of cookies, the number of clicks should not change because clicking on the “start free trial” button comes before the experimental diversion.

Gross Conversion - Evaluation Metric

The gross conversion measures the number of users who enroll divided by the number who click the “start free trial” button. This is useful for measuring whether the disclaimer message had an impact on the number of users who enrolled for classes. If the gross conversion rate is much lower in the experimental group than the control, the change might not be recommendable as fewer users are enrolling, which might offset gains in retention.

Net Conversion - Evaluation Metric

The net conversion measures the number of users who stay past the free trial divided by the number who click the “start free trial” button. If the hypothesis is true, then the experimental group should have a higher net conversion than the control, as the number of users who click the “start free trial” button will stay the same but the number of users who stay after the free trial will increase.

Click-Through-Probability - Not used as invariant

The click-through-probability on the “start free trial” button should not be affected by the experiment because there are no changes being made to the course overview page between the experimental and control groups. However, number of cookies and number of clicks are already being tested as invariant metrics, and the click-through-probability simply combines these. If the cookies and clicks pass the sanity checks, then the click-through-probability will as well.

Number of User Ids - Not used as invariant

Not useful as an invariant because the number of users who enroll in the free trial is expected to change in the experiment. Not useful as an evaluation metric because the information is already included in the other evaluation metrics chosen.

Retention - Not used as evaluation

Not used because net and gross conversion are sufficient metrics to test the hypothesis. Adding retention would not give additional useful data. Between gross conversion and retention, gross conversion is more appropriate as an evaluation metric because it directly measures the degree to which the disclaimer message affects enrollment, which is essential for making a recommendation with this data. Retention can be used for this in conjunction with net conversion, but in a less direct way.

Measuring Standard Deviation

Gross conversion - 0.0202

Net conversion - 0.0156

The empirical variability does not need to be computed for these metrics because the unit of analysis and unit of diversion are the same in both cases.

Sizing

Number of Samples vs. Power

The bonferroni correction will not be used during the analysis phase. 685325 pageviews will be necessary to power the experiment appropriately.

Duration vs. Exposure

Given the number of pageviews necessary for the test to yield results, all the traffic from the Udacity website will be diverted for this experiment. With all the traffic, the experiment will take 18 days to run. All the traffic is being diverted because any less will extend the experiment to a longer than desirable length. The experiment is relatively low risk as it is unlikely to have repercussions that would damage Udacity financially.

Experiment Analysis

Sanity Checks

Number of Cookies - confidence interval [0.4988, 0.5012]

Observed value - 0.5006 - passes

Number of Clicks - confidence interval [0.4959, 0.5041]

Observed value - 0.5005 - passes

Result Analysis

Effect Size Tests

Gross Conversion - confidence interval [-0.0291, -0.0120]

Practical significance level = 0.01

Statistically and practically significant

Net Conversion - confidence interval [-0.0116, 0.0019]
Practical significance level = 0.0075
Neither statistically or practically significant

Sign Tests

Gross Conversion - p-value = 0.0026 - statistically significant

Net Conversion- p-value = 0.6776 - not statistically significant

Summary

I decided not to use the Bonferroni correction because with only 2 evaluation metrics a false positive by chance is unlikely. In this case the Bonferroni correction is too conservative. There are no discrepancies between the effect size hypothesis tests and the the sign tests. Both were significant for Gross Conversion and neither were significant for the Net Conversion.

Recommendation

Given the above results I would recommend not making the change to include the disclaimer. The hypothesis of this experiment was to see whether the disclaimer would increase the number of users who made it to the 1st payment. The net conversion which measures this shows no significant change, meaning that the disclaimer does not have a positive impact on the number of users who made it to the 1st payment. Furthermore, the gross conversion shows that the disclaimer significantly decreased the number of users who enrolled for classes. Implementing this change is not in Udacity's best interest.

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.

One of the strongest incentives for going through a full nanodegree at Udacity is the offer of getting half the money back upon completion. The follow-up experiment I'm proposing is to add an incentive onto the individual classes as well. In order to reduce early cancellations, the experiment would be to offer an additional 2 weeks free on top of the given 2 week free trial, given that the user completes some amount of the course within the first 2 weeks. The format of the test is similar to the prior experiment. A pop-up window would appear after clicking the start free trial button, explaining the possibility of gaining an extra 2 free weeks of classes dependent on completing 2 lessons within the first 2 weeks. The hypothesis is that fewer people will drop out of class early when incentivized by an extension.

Number of cookies that view the course page and number of clicks on the start free trial button would be the invariant metrics. Net conversion and gross conversion would be used as the evaluation metrics, with net conversion being the number of users who make it to the 1st payment divided by the number of clicks on the start free trial button, and gross conversion being the number of users who enroll divided by the number of clicks on the start button. The unit of diversion is a unique cookie until enrollment, after which users are tracked by userid. A positive result for the experiment would be a significant increase in the number of users who make the 1st payment, and a positive change in the number of users who enroll for courses. Limitations of this experiment include the extended time range (4 weeks until the 1st payment for some users) as well as the variable length of time for some users to get to the 1st payment.

Resources

http://www.evanmiller.org/ab-testing/sample-size.html
http://graphpad.com/quickcalcs/binomial2/
http://stats.stackexchange.com/questions/65942/how-and-when-to-use-the-bonferroni-adjustment