Christian Rivera

Udacity DAND Project 7, Design an A/B Test

Experiment Design

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.

Invariant Metrics:

Number of Cookies - The population sizes for the experimental group and control group should be equal or proportional with significant enough sample sizes for each.

Number of Clicks - Since there is nothing different between the first pages, the experimental and control groups should have the same number of clicks or a proportional amount of clicks if their sample sizes are different.

Click Through Probability - As the first pages are the same between the experiment and control groups, the click through probability should be the same.

Evaluation Metrics:

Gross Conversion - Since gross conversion shows the proportional number of user-id's created created from the total number of unique cookies, the control group should have a higher gross conversion as it has no deterring survey unlike the experiment group.

Net Conversion - Since the experimental group, unlike the control group, has a survey meant to prepare students for the workload ahead, the experimental group should have a higher net conversion rate than the control group as both should have the same or proportional number of unique cookies clicking the 'start free trial button'.

To justify launching the experiment, the Gross Conversion of the experimental group must be statistically less than the Gross Conversion of the control group, and the Net Conversion rate for the experimental group must be statistically greater than or equal to the Net Conversion of the control group.

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 a sample size of 5000 cookies, the standard deviations for the following evaluation metrics are shown below

Standard Deviation Gross Conversion = 0.0202

=*sqrt((0.20625*(1-0.20625))/(5000*(3200/40000)))*

Standard Deviation Net Conversion = 0.0156

=sqrt((0.1093125(1-0.1093125)/(5000*(3200/40000))))*

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.

Since both Net Conversion and Gross conversion both have the same unit of analysis, *unique cookies to click button,* and since the unit of analysis is also a unit of diversion (cookie), there is no need to test for 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.)

I will not be using a Bonferroni correction for any of the metrics.

	Gross Conversion	Net Conversion
Alpha	0.05	0.05
Beta	0.2	0.2
Baseline conversion rate	0.20625	0.1093125
Minimum Detectable Effect	0.01	0.0075
Sample size	322938	342662.5
Larger Value	-	342662.5
Control Group + Exp Group		**685325**

Number of pageviews needed to power the experiment: 685325

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.)

Fraction of traffic diverted to the experiment: **60%**

Length of experiment: **29 days**
*685325 / (40,000 *0.5) = 28.55, rounded up to 29*

I chose to divert 60% of all traffic to the experiment because the experiment in question does not drastically change the website's interface or flow of traffic. The addition of a simple survey, while potentially positive, probably does not have too much of an effect one way or another. As such, there is less of a concern of alienating the user base with this experiment. Therefore it is a low risk venture for the company. Also, for the user test subjects, no high risk data is being collected for the experiment as the experiment only requires cookie and user-id data. Since the experiment, is just the addition of a simple survey to better inform new students, it can also be reasonably assumed that any students in the experiment will not be harmed by participating in the experiment. As the experiment poses a low risk to both the company and the test subjects, a larger proportion of company's website traffic can be diverted towards the experiment which, for this experiment is 60%.

By diverting 60% of all traffic into this experiment, the experiment can be completed in a reasonable amount of time (29 days) which and therefore satisfy the time constraints the company works under.

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:
Passed? Yes

Pageviews Control	Pageviews Experiment	Probability	Standard Error	Marginal Error	Upper Bound	Lower Bound	Observed

345543	344660	0.5	0.00060184074	0.0011796078 5	0.5012	0.4988	0.5006

Number of Clicks on “start free trial”:
Passed? Yes

Clicks Control	Clicks Experiment	Probability	Standard Error	Marginal Error	Upper Bound	Lower Bound	Observed
28378	28325	0.5	0.00209974708	0.0041155042 76	0.5041	0.4959	0.5004

Click Through Probability on “start free trial”:
Passed? Yes

Clicks % Control	Clicks % Experiment	Probability	Standard Error	Marginal Error	Upper Bound	Lower Bound	Observed
28378	28325	0.08213	0.00047	0.00092	0.083	0.0812	0.0822

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

P_hat_cont	0.21887
P_hat_exp	0.19832
D	-0.02055
Probability	0.20861
Standard Error	0.00437
Marginal Error	0.00857
Upper Bound	-0.012

Lower Bound	-0.0291
dmin	0.01
Statistically Significant	Yes
Practically Significant	Yes

Net Conversion

P_hat_cont	0.11756
P_hat_exp	0.11269
D	-0.00487
Probability	0.11513
Standard Error	0.00343
Marginal Error	0.00673
Upper Bound	0.0018
Lower Bound	-0.0116
dmin	0.0075
Statistically Significant	No
Practically Significant	No

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:

P value = 0.0026

Statistically Significant? Yes

Practically Significant? Yes

Net Conversion:

P value = 0.6676

Statistically Significant? No

Practically Significant? No

Summary

I did not use a Bonferroni correction because, in order for the experiment to be statistically significant, both Gross Conversion and Net Conversion results need to be statistically significant. Since, both metrics are required, a Bonferroni correction is not needed.

Recommendation

The hypothesis, that a survey after the “start free trial” button would set clearer expectations for prospective students without reducing the number of paying students, appears to be shown true by the data. The experimental group had a statistically and practically significant drop in its gross conversion rate compared to the rate of the control group. However, the results for the net conversions between the experimental and control groups were statistically and practically insignificant. Even so, the results from the net conversion had a confidence interval that overlapped the negative practical significance boundary of -0.0075. This indicates that, had more net conversion data been collected to be practically significant, it would have more likely shown results that the experimental group’s net conversion rate would be lower than the rate of the control group. Since the hypothesis required that the experimental group’s net conversion rate not be lower than the net conversion rate of the control group, the experiment failed. Because of this, I recommend to not 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.

Since the previous experiment identified that the less dedicated students do not stay enrolled long enough in the course to become paying students, I believe a good follow up experiment is to identify if specific sections in the early course work are disproportionately impacting the current dropout rate among less dedicated enrollees in the trial period. If these specific sections exist and are identified, then Udacity could redesign those sections of a course to create a more gradual ramp up in difficulty which in turn would no longer discourage less dedicated students from dropping out.

Hypothesis: There are specific sections of the early course work that cause a higher dropout rate among less dedicated student than the more dedicated ones.

Parameters: This experiment will have the same parameters as the previous one with an experimental group that receives a survey after clicking the “start free trial” button, which has shown to weed out less dedicated students, and a control group that doesn’t receive the survey.

Unit of Diversion: The Unit of Diversion for this experiment will be the user-id. Since this experiment is tracking only the enrolled students, and each student has a unique user-id, these user-id’s can be tracked to see which pages they have visited which show how far the students have progressed in a course.

Invariant Metrics: User-id’s would also be the invariant metric for this experiment as there should be an equal number of student enrollees in both the experimental and control groups.

Evaluation Metrics: There will be multiple conversion rate evaluations metrics in the experiment that show the percentage of user-id’s that have visited specific pages (end of section 1 of the course, end of section 2, etc) out of the total number of user-id’s.