Final Project of Design an A/B Test Course

Udacity Data Analyst Nanodegree Project 7

Experiment Design

**Metric Choice**

*Invariant metrics*: Number of Cookies and Number of clicks

I chose the number of cookies and the number of clicks as invariant metrics because these two values will be consistent throughout the experiment. These two values can be calculated before users decide to enroll so they won’t be affected by the experiment.

*Evaluation metrics*: Gross Conversion and Net Conversion

From the experiment, we want to know whether pop-up message asking the possible commitment to the degree affects students to enroll the free trial and make payments. These two metrics provide the values to analyze the results of the experiment. I will calculate if the experiment group has lower Gross Conversion and Net Conversion than 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.)

Standard Deviation of Gross Conversion: 0.0202

Standard Deviation of Net Conversion: 0.0156

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.

**Sizing**

**Number of Samples vs. Power**

**Will you use the Bonferroni correction in your analysis phase?**

No

**The number of pageviews I need to power the experiment:**

I calculated the pageviews from the calculator from <http://www.evanmiller.org/ab-testing/sample-size.html> with a baseline conversion rate = 10.931% (Net Conversion rate) and minimum detectable effect = 0.75%. As a result, I got 27,413. However, this number is the size we need to power the experiment for net conversion rate. To find the pageviews we need, we need to take account that among 40,000 people viewing the page, only 3,200 people clicks the “Start free trial”. Thus, 27,413 \* 40,000 / 3,200 = 685,325 pageviews are needed.

**Duration vs. Exposure**

The fraction of traffic I would divert to this experiment is 1 (from 0 to 1). Given this fraction, I need 685,325 / 40,000 = 17.13 -> 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?

Experiment Analysis

**Sanity Checks**

Number of cookies: Lower bound = 0.4988, Upper bound = 0.5012, Observed = 0.5006: Passed

SE = sqrt( p \* (1-p) \* (1/Number of Control Pageviews + 1/Number of Experiment Pageviews)

= sqrt( 0.5 \* 0.5 \* (1/344660 + 1/345543) = 0.0006

Margin of error = 0.0006 \* 1.96 = 0.0012

Lower bound = 0.5 – 0.0012 = 0.4988, Upper bound = 0.5 + 0.0012 = 0.5012

Observed = 345543 / (345543 + 344660) = 0.5006

Number of clicks: Lower bound = 0.4959, Upper bound = 0.5041, Observed = 0.5005: Passed

SE = sqrt( p \* (1-p) \* (1/Number of Control Clicks + 1/Number of Experiment Clicks)

= sqrt( 0.5 \* 0.5 \* (1/28325 + 1/28378) = 0.0021

Margin of error = 0.0021 \* 1.96 = 0.0041

Lower bound = 0.5 – 0.0041 = 0.4959, Upper bound = 0.5 + 0.0041 = 0.5041

Observed = 28378 / (28378 + 28325) = 0.5005

**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: Lower bound = -0.0291, Upper bound = -0.0120

Statistical Significance: Yes, Practical Significance: Yes

Net Conversion: Lower bound = -0.0116, Upper bound = 0.0019

Statistical Significance: No, Practical Significance: No

**Sign Tests**

Gross Conversion: p-value = 0.0026, Statistical Significance = Yes

Net Conversion: p-value = 0.6776, Statistical Significance = No

**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.

I didn’t use Bonferroni correction because …

**Recommendation**

Make a recommendation and briefly describe your reasoning.

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.