**AB TEST Report**

**Author: Rocha**

**Date: April 2017**

**Experiment Design**

**Metric Choice**

 **Number of cookies:**  Invariant Metrics. This is our unit of diversion, in the two groups, we want to see them even distributed.

 **Number of user-ids:** Not chosen. This is neither Invariant metrics nor evaluation metrics. User-id only can be tracked when students enrolled, so we can not track those who don't enrolled.

 **Number of clicks:**Invariant Metrics. We don’t want to see any impact on our “start free button”. So we want them even distributed in our two groups, thus I chose it for our Invariant.

 **Click-through-probability:** Invariant Metrics.  Unique cookies to click the "start free trial" button per unique cookies to view the course overview page.This variable is the same as the number of clicks, so we just need to keep it invariant.

 **Gross conversion:**Evaluation Metrics. This one is a measurement of number of enrollment after completing the checkout.

 **Retention:**Evaluation Metrics. This is per number of user-id which paid at least once after the checkout.

 **Net conversion:**Evaluation Metrics. This is also an evaluation metrics after the checkout, the number of user-id at least paid once.

We would want to see a statistically and practically significant **decrease** in gross conversion and we would want to be confident that net conversion **did not decrease** by the practical significance boundary or more.

**Measuring Standard Deviation**

For the sheet of baseline is 40,000 unique cookies to view the page per day. And in our question it’s 5,000 cookies. First, we assume our two evaluation metrics follows binomial distribution. Then the standard deviation is . In 5000 samples, we get our number of enrollment is 82.5. So our **Gross conversion’s analytic variance** is 0.0202, the unit of diversion is cookie and so the analytic variance should match the empirical variance. The **Retention’s analytic variance** is 0.0549 its unit of analysis is user-id, not matches the unit of diversion, so it won’t match the empirical variance. The **Net conversion’s analytic variance** is 0.0156, its unit of analysis is cookie, matches our unit of diversion.

**Sizing**

**Number of Samples vs. Power**

I don’t use the Bonferroni correction. is 5% , is 20%.

**Gross Conversion**: Baseline conversion rate is 20.625% and the minimum detectable effect is 1% .Our sample size needed is 25,835 / (320/40,000) = 645,875

**Retention**: Baseline conversion rate is 53% and the minimum detectable effect is 1% . Our sample size needed is 39,115 / (660/40,000) \* 2 = 4,741,212.

**Net Conversion**: Baseline conversion rate is 20.625% and the minimum detectable effect is 1% . Our sample size needed is 27,413 / (3200/40,000) \* 2 = 685,325

And my number of pageview is 4,741,212. Because of this number is so large, so we don’t choose the Retention as our evaluation metrics. So our sample size we used is 685,325.

**Duration vs. Exposure**

From above, our sample size is 685,325. Because divert the traffic doesn’t have moral effect so we can divert 100% of the traffic to our experiment group. In this situation, I can't see any risk in this experiment. When students choose that they don’t have more than 5 hours per week to study, they will be oriented to free courses. They could enroll in free trial whenever they feel like to do so. There is nothing changed related database, so we don’t need to worry about missing any important data. Also, no personal privacy information will be required in the experiment, therefore there is no ethical consideration. It will be ok to run on all traffic if this is the only experiment Udacity wants to run, especially given that the experiment will take a non-trivial amount of time even on all traffic. Our daily number of pageviews per cookie is 40,000. So we collect our data just need 18 days.

**Experiment Analysis**

**Sanity Checks**

In the sanity check, we should check our invariant, namely number of cookies and number of click on “Start free trial”. In our check, equals 0.05

**number of cookies**: [0.4988, 0.5012] and our observation is 0.5006. So it’s pass our check.

**number of click on “Start free trial”:** [0.4959, 0.5041] and our observation is 0.5005. Pass.

**Click-through-probability:** [0.0812, 0.0830] and our observation is 0.0821.Pass.

**Result Analysis**

**Effect Size Tests**

**Gross Conversion**: [ -0.0291, -0.0120]. Statistical significant. Practical significant.

**Net Conversion**:[ -0.0116, 0.0019]. Both not significant.

**Sign Tests**

We assume our evaluation metrics follows binomial distribution.

**Gross Conversion:** We have 4/23 times, so p = 0.0026. That’s statistically significant.

**Net Conversion:** We have 10/23 times, so p = 0.6776. That’s not statistically significant.

**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 don’t use the Bonferroni correction because if we use that we will need more time to collect our sample. In this case, two metrics are measured. Because they are highly related, it is inappropriate to advocate launching, so Bonferroni is overly conservative. Even though Bonferroni didn't affect the statistical significance results here, it's important not to use the Bonferroni correction unless you need it, since otherwise you're going to miss out on launching a lot of experiments. There’s discrepancy between hypothesis tests and sign tests in Gross Conversion rate. Maybe because the time affect, like it concluding holiday traffic.

**Recommendation**

So in the end, we can conclude that our test would reduce the click on “free start screenshot”, however it won’t show a significant increase in the net conversion. So we didn’t reach our goal, so we shouldn’t launch the experiment.

**Follow-Up Experiment**

In the future work, if we wanted to reduce the number of frustrated students who cancel early in the course, we can do things such as: 1. Have more specific instruction before they start the course. 2. Enhance the free trial courses, the same as improve the course’s quality. 3. Have stricter and clarified prerequisite to filter the students. We can use user-id as our unit of diversion. We could measure the same net conversion rate, use similar invariant like number of pagesviews, number of clicks on “free start trial.”

As for the last method, having stricter and clarified prerequisite:

**Unit Of Diversion:** We can choose cookie, the same as before.

**Null Hypothesis:** We assume our gross conversion rate reduced both statistical significant and practical significant and our retention or net conversion rate don’t decreased significantly.

**Invariant Metrics:** Number of cookies and number of clicks or click-through-rate are good choice.

**Evaluation Metrics:** The same as our test before. We can still use gross conversion ,retention, net conversion as our evaluation metrics.

**Reference**

Discussion Forum

1V1 Mentor