Project 7-Design A/B Testing

Data Analyst Nanodegree

Udacity

Rahel Ghebrekidan

September 6, 2016

A/B Testing - Experiment Design

Introduction

Udacity is an educational organization which provides online courses. Udacity's course home page has two option during the time of experiment. These are "start free trial" and "access course material. If a user clicks on the "start free trial", the system will request the user to fill his/her credit card information. After filling the credit card information, the user will be enrolled in the free trial for 14 days. After 14 days, the user will be automatically charged if he/she did not cancel the program. On the other hand, if a user clicks the "access course materials" button, the system will not ask for credit card information. The user has access to the lecture videos and take quizzes for free. Of course, the user will not have coaching support service, access to submitting their projects and get feedback and at last cannot get certificate.

Udacity courses needs some devotion. Udacity (for this experiment) estimates the time a student needs to complete successful is at least 5 hours per week. Udacity wants to add a time screener (click here) in the home page in order to reduce the number of frustrated students who drops out in the 14 days of free trial because they do not have enough time without reducing the number of students who continue after the 14 days of free trial and complete the course successfully. The experiment will ask the user how many time will they allocate for the course immediately after the user clicks the "free trial button". If the user mentioned more than 5 hours, it will ask the user to fill his/her credit card information and the user will be enrolled in the free trial. However, if the user, mentioned less than 5 hours a week, the test will forward him that the course needs at least 5 hours devotion and if the person cannot allocate at least 5 hours a week, it will suggest the uses to use the "access course materials" instead. In this project, I will evaluate if introducing this test will meet the expectation or hypothesis of Udacity mention above.

Metric Choice

The list of Metrics given are number of cookies, number of user-ids, number of clicks, click-through-probability, gross conversion, retention and net conversion. The metrics which I have chosen as invariant metrics and evaluation metrics are:-

Invariant Metrics: Number of cookies, Number of clicks and Click-through-probability

Evaluation Metrics: Gross conversion, Retention and Net conversion

Reason for Choosing a Metric as Invariant Metrics or Evaluation Metrics or not choosing at all

Number of cookies: is number of unique cookies to view the course overview page (home page holding the free trial and the course material buttons). This is the same for the control and experiment group because it comes before the test. It is also the unit of Diversion. So I have chosen it as invariant Metrics.

Number of user-ids: is number of users who enroll in the free trial. For a user to enroll in the free trial, he/she will first go through the test and mentioned that he/she will commit more than 5 hours a week for the course. So this is not invariant metrics. I have neither chosen this as an evaluation metrics although it is evaluation metric. It is not a good evaluation metric as the number of visitors may not be the same in the control and experiment group. It is better to choose Gross conversion because it adjusts size difference in experiment and control group.

Number of clicks: is number of unique cookies to click the "start free trial" button. This comes before test so it will not be affected by the test. Hence, it is invariant metrics.

Click-through-probability: is derived from two invariant metrics, number of clicks divided to number of cookies. As both number of clicks and number of cookies are independent of the test, click-through-probability is also independent of the test and is invariant metric.

Gross conversion: is derived when number of user-id to complete checkout and enroll in the free trial is divided by number of clicks. The numerator (completing checkout and enrolling in the free trail) is dependent on the experiment, it is not invariant metric. Since this is dependent on the effect of the test filtering out users who have the potential of frustration because they could not allocate the required time so it is an evaluation metric.

Retention: is number of user-ids to remain enrolled past the 14- day boundary divided by number of user-ids to complete checkout. Both the numerator and denominator are dependent of the test. So retention is not invariant metrics. It is an evaluation metric because this helps us measure the effect the test on the ratio of users who remain enrolled after the free trial (make at least one payment) to the number of user who complete checkout, it is an evaluation metric. As retention requires too many page views, I have drop it later in my analysis (see sizing).

Net conversion: is the number of user-ids to remain enrolled past the 14-day boundry divided by the number of clicks. It is not invariant matric because the number of user-ids to remain enrolled past the 14-day is dependent on the experiment. It is not the same for the control and experiment group. But it is a good evaluation metrics because it will clearly evaluate the effect of the experiment on having users who will at least make one payment.

Expected outcome to launch the experiment

I will recommend to launch the experiment if the number of enrollments are reduced but the number of users who continue to enroll after the 14 days of free trial increases or remain the same. That is if the Gross conversion decrease but the net conversion either increase or remain the same.

Measuring Standard Deviation

I will calculate the standard deviation of Gross Conversion, Retention and Net Conversion for sample size of 5000 cookies visiting the course overview page.

Rough estimate of base line value of metrics given as below.

Unique cookies to view page per day:	40000
Unique cookies to click "Start free trial" per day:	3200
Enrollments per day:	660
Click-through-probability on "Start free trial":	0.08
Probability of enrolling, given click:	0.20625
Probability of payment, given enroll:	0.53
Probability of payment, given click	0.1093125

Gross Conversion

I will need number of clicks (Number of unique cookies to click the "start free trial" button) out of the 5000 sample size given 0.08 click through probability

Number of click(N) = 0.08*5000 = 400

$$\text{Std} = \sqrt{\left(\frac{P(1-P)}{N}\right)} = \sqrt{\left(\frac{0.20625(1-20625)}{400}\right)} = 0.0202$$

Retention

I will need number of enrollment out of the 5000 sample size given 0.20625 probability of enrolling and 400 number of clicks, based on gross conversion definition number of enrollment will be:

Number of enrollment (N) = 400 * 0.20625 = 82.5

$$\text{Std} = \sqrt{\left(\frac{P(1-P)}{N}\right)} = \sqrt{\frac{0.53(1-0.53)}{82.5}} = 0.0549$$

Net Conversion

$$\text{Std} = \sqrt{\left(\frac{P(1-P)}{N}\right)} = \sqrt{\frac{0.1093125(1-0.1093125)}{400}} = 0.0156$$

Both Gross conversion and Net conversion use the unit of diversion (number of cookies) as their denominator so their analytic variance and empirical variance are likely to be similar however Retention uses number of enrolled users as its denominator, so its analytical variance and empirical variance are different.

Sizing

Number of Samples vs. Power

I have decided not to use Bonferroni correction, because to launch the experiment, I will need Gross conversion to decrease and Net conversion to be increase or remain the same.

Based on the baseline information given and 5% alpha, and 20% Beta, I have used the online calculator to calculate the sample size needed for each metric of my choice.

Evaluation Metrics	Baseline conversion rate	Minimum Detectable Effect	Sample Size (per group)
Gross Conversion	20.625%	1%	25,835
Retention	53%	1%	39,115
Net Conversion	10.93129%	0.75%	27,413

The above table shows us that we will need 25,835 number of clicks in each groups for Gross conversion, 39,115 number of enrollment in each group for Retention and 27, 413 number of click in each group for Net conversion.

There are 3200 number of clicks in 40,000 page views, to have 25,835 number of clicks for Gross Conversion evaluation metrics we need 322,938(25,835*40,000/3200) number of page views.

There are 660 enrollments in 40,000 page views, in order to have 39,115 page views we will need 2,370,606(39,115*40,000/660) page views.

We will need 342,662.5(27,413*40,000/3200) page views to have 27,413 number of clicks for Net Conversion.

Those numbers of page views are only for one group. So I have to multiply all by two to get the total number of page views needed for both the experiment and control groups. That is:-

Gross conversion:- 322,938 *2 = 645,876

Retention:- 2,370,606*2 =4,741,212

Net Conversion:- 342,662.5*2=685,325

According to the above calculation, Retention evaluation metric needs too many page views which is 4,741,212 which will require too many days. In addition to this it uses number of users-id as its denominator which is not our unit of diversion so I will drop it and continue with Gross conversion and Net conversion.

Duration vs. Exposure

I have decided to divert 100% of the traffic to the experiment. Because the experiment is not risky for the users (viewers) as well us for udacity. Users who did not enroll because they did not have enough time, can access the course material and can enroll any time when they are ready. They still have access to material through "access course materials". In addition to this we are not asking any confidential information at this point so adding this experiment is not violating any ethics. Similarly, adding the time screener will not affect Udacity's website. Udacity is not making major change to include this test.

Taking the higher number of page views (page views required by Net conversion), the duration will be;

Duration(days) = 685,325 /1/40,000 = 17.13

It will approximately take 18 days.

Experiment Analysis

Sanity Checks

Here I will check if my invariant metrics are equivalent for the control and experiment groups. I will use binomial test and take probability 0.5 for the number of clicks and number of cookies. Please click here for the experiment and control data.

Number of cookies

Total number of page views of the control group(n_C) = 345,543

Total number of page views of the experiment group(n_e) = 344,660

Total page view ($n = n_C + n_e$) = 690,203

SE = $\sqrt{\left(\frac{P(1-P)}{n}\right)}$ = $\sqrt{(\frac{0.5(1-0.5)}{690,203})}$ = 0.0006018

Margin of Error (m) = Z* S.E = 1.96 * 0.0006018 = 0.0011795

Confidence Interval = (0.5-0.0011795, 0.5 + 0.0011795) = (0.4989,0.5012)

Observed fraction (P) = 344660/690203 = 0.5006

The observed fraction falls within the confidence interval, so the number of cookies has passed sanity check.

Number of clicks

Total number of clicks of the control group(n_c) = 28,378

Total number of clicks of the experiment group(n_e) = 28,325

Total number of clicks ($n = n_c + n_e$) = 28,378 + 28,325 = 56,703

SE = $\sqrt{\left(\frac{P(1-P)}{n}\right)}$ = $\sqrt{(\frac{0.5(1-0.5)}{56,703})}$ = 0.0021

Margin of Error (m) = Z* S.E = 1.96 * 0.0021 = 0.004116

Confidence Interval = (0.5-0.004116, 0.5 + 0.004116) = (0.4959,0.5041)

Observed fraction (p) = 28,378/56,703 = 0.5005

The observed fraction falls within the confidence interval, so the number of clicks has passed sanity check.

Click through probability

Click through probability is Number of cookies divided by number of clicks. As both number of cookies and number of clicks have passed sanity check, click through probability will also passed the sanity check.

Control value probability(P_C)= Number of clicks of the control group/Number of page views of the control group

= 28,378/345,543 = 0.082126

Total number of page views of the experiment group(n_e) = 344,660

SE = $\sqrt{Pc\ (1 - PC)/ne}$ = $\sqrt{0.082126\ (1 - 0.082126)/344,660}$ = 0.000468

Margin of Error (m) = Z* S.E = 1.96 * 0.000468 = 0.000912

Confidence Interval = (0.082126 - 0.000912, 0.082126 + 0.000912) = (0.0812,0.0830)

Experiment Value = 28,325/344,660 = 0.0822

The experiment value is within the confidence interval, so it has passed sanity check.

Result Analysis

Effect Size Tests

Given 95% confidence of interval, I will check if Gross conversion and Net conversion are statistically and practically significant.

Gross Conversion

Number of clicks of the control group (Nc) = 17,293

Number of enrollment of the control group(X_c) = 3,785

Number of clicks of the Experiment group($N_{e)}$ = 17,260

Number of enrollment of the Experiment group($X_{e)}$ = 3,423

$P_{pool} = (Xc + X_e)/(N_C + N_e) = (3,785 + 3,423)/(17,260 + 17,293) = 0.2086$

$$SE_{pool} = \sqrt{\left(Ppool(1 - Ppool)\left(\frac{1}{Nc} + \frac{1}{Ne}\right)\right)} = \sqrt{(0.2086(1 - 0.2086)(1/17{,}260 + 1/17{,}293)} = 0.00437$$

$d = X_E/ N_E – X_C/ N_C = 3,423/17,260 – 3,785/17,293 = -0.02055$

$m= SE_{pool} * 1.96 = 0.00437 * 1.96 = 0.00857$

Confidence Interval = (d-m, d+m) = (-0.02055 – 0.00857, -0.02055 + 0.00857)

= (-0.0291, -0.0120)

The confidence interval does not include 0 and is greater than the Practical significant boundary ($d_{min),}$ Gross conversion is statistical and practically significant.

Net Conversion

Number of clicks of the control group (Nc) = 17,293

Number of payment of the control group(X_c) = 2,033

Number of clicks of the Experiment group($N_{e)}$ = 17,260

Number of payment of the Experiment group($X_{e)}$ = 1,945

$P_{pool} = (Xc + X_e)/(N_C + N_e) = (2,033 + 1,945)/(17,293 + 17,260) = 0.1151$

$$SE_{pool} = \sqrt{\left(Ppool(1 - Ppool)\left(\frac{1}{Nc} + \frac{1}{Ne}\right)\right)} = \sqrt{(0.1151(1 - 0.1151)(1/17{,}260 + 1/17{,}293)}$$

$= 0.0034$

$d = Xe/ Ne – X_C/ N_C = 1,945/17,260 – 2,033/17,293 = -0.00487$

$m = SE_{pool} * 1.96 = 0.00437 * 1.96 = 0.0067$

Confidence Interval = (d-m, d+m) = (-0.00487 – 0.0067, -0.00487 + 0.0067)

= (-0.0116, 0.0019)

The confidence interval includes 0 and the lower boundary is less than the practical significant, hence it can be concluded that Net conversion is not statistical and practical significant.

Sign Tests

I have done the day by day sign test test using the sign and binomial test calculator.

Gross Conversion

Number of successes = 4 …. Click here for the calculation

Probability of success = 0.5

Number of trials per experiment = 23

Two-tail P value is 0.0026

Since the two- tail P-value is less than α (0.025), it can be concluded that Gross conversion is statistical significant.

Net Conversion

Number of successes = 4 …. Click here for the calculation

Number of trials per experiment = 23

Probability of success = 0.5

The two-tail P value is 0.6776

The p- value is greater than α (0.025).So similar to the conclusion derived from effective size tests, Net conversion is not statistically significant.

Summary

As mentioned earlier, I have not used the Bonferroni correction. In our experiment we want Gross conversion to decrease and the Net conversion to increase. Bonferroni correction would have been good to control type I error if we were performing multiple tests.

The conclusions drawn from both the effect size hypothesis tests and the sign tests is the same. In both tests, Gross conversion is statistically and practically significant but Net conversion is neither statistical significant nor practically significant. Although Gross conversion decreases, the confidence

Recommendation

Based on the analysis done, I would not launch the experiment. We want to decrease the number of students who might withdraw during the 14 free trial days. At the same time, we want to ensure the number of students who accomplish the course successfully or who remained enrolled after the 14 days of free trial to increase or remain the same. In this analysis, the Gross conversion and Net conversion are decreased. That is implementing the screener might reduce the number of users who might withdraw due to frustration of not getting enough time. However, it is also decrease the number of total students who remained enrolled after the 14 days of free trial. Therefore, I do not recommend launching this experiment as it is not beneficiary for users, and it is decreasing Udacity's revenue.

Fellow- Up Experiment

From the above experiment, the reason for the decrease in the Net Conversion could be missing users who could have remained enrolled after the 14 free trial days but were scared of the time commitment. Some of those who changed their mind when they saw the time screener could have devoted the required time after they started the free trial. I would suggest if Udacity includes some incentives for users who remained enrolled after the free trial and complete the course successfully. For instance, when I first started the program, it was mentioned that I need to devote at least 40 hours a week. The required time was too much for me when I first read it. However, I do not want to miss the opportunity given. I joined and practically I am allocating even more than 40 hours a week for the course by arranging and given priority to the course. One of the reasons is if I finish the program within the three months, I am not going to pay for it. Similarly, users will be encouraged to complete the program successfully if there are some incentives like discount.

Therefore, I would like to suggest adding incentive(discount) to the time screener. That is, instead of asking "How many hours you will spend?", it will be good if it says, ***"Students who completed the course successfully will get X% discount and the course requires at least 5 hours' devotion per week".*** The user will see it after clicking the "start free trial" the same as the above experiment done. The experiment is informing users need to devote at least 5 hours a week and at the same time encouraging users to enroll and devote at least 5 hours for the course and get some discount. To estimate the possible discount to be given, we need to consult the finance department to come up with visible rate which increases the profit with increase number of enrollments and which gives good deal to users.

The hypothesis is to reduce students who cannot make any change to their timetable to devote at least 5 hours a week for the course and who have the potential of frustration and withdrawing during the free trial days. However, there might be some users who can make some arrangements and give priority to the course to devote at least 5 hours when they saw the discount. Therefore, my expectation is the Gross conversion will decrease but with small margin. The Net conversion will increase because users will remain enrolled and complete the course motivated by the discount.

The unit of diversion and metrics will be the same as the above. Except I will not put Retention as evaluation metrics and I will make some change on the definition of Net Gross.

The unit of diversion will be a cookie.

Invariant Metrics: -

Number of clicks*:* the number of unique cookies to click the "Start free trial" button. This is independent of the experiment as the user will not see the experiment before clicking the "Start free trial" button.

Number of cookies: is number of unique cookies to view the course home page.

Click- through- probability: is number of cookies dived by number of clicks

Evaluation Metrics:-

Gross Conversion**:** The number of users enrolled in the free trial divided by the number of clicks

Net Conversion**:** The number of users remain enrolled after the 14 free trial days and complete the course successfully divided by the number of clicks. Here I am focusing on students who completed the course because the advantage of the experiment will be optimal if students complete the course.

I think with this experiment, students will be encouraged to complete the course that they are enrolled in and the total revenue will increase.

Reference:

1. A/B Testing Metrics: Retrieved from http://napitupulu-jon.appspot.com/posts/metrics-abtesting-udacity.html