Data Analyst

Data Analyst Nano Degree

P7 – DESIGN AN A/B TEST

Harshit Trivedi | Data Analyst Nano Degree | July 23, 2016

A/B Testing

EXPERIMENT

Udacity tested a change where if the student clicked 'Start Free Trial', they were asked how much time they had available to devote to the course. If the student 5 or more hours/week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours/week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free.

INITIAL UNIT OF DIVERSION

Cookies

Experiment Design

METRIC CHOICE

Invariant Metrics: Number Of Cookies, Number Of Clicks, Click-through Probability

Evaluation Metrics: Gross Conversion, Retention, Net Conversion

Number Of Cookies:

Number of unique users to visit the course overview page. Good invariant metric because the visits happen before the user sees the experiment, and are thus independent from it.

Number of User IDs:

Number of users who enroll in the free trial. Not a good invariant metric because the number of users who enroll in the free trial is dependent on the experiment. Not an ideal evaluation metric because the number of visitors may be different between the experiment and control groups, which would skew the results.

Number Of Clicks:

Number of unique cookies to click the start free trial button. Good invariant metric because the clicks happen before the user sees the experiment, and are thus independent from it.

Click-through Probability:

Number of unique cookies to click the 'Start Free Trial' button divided by the number of unique cookies to view the course overview page. Good invariant metric because the clicks happen before the user sees the experiment, and are thus independent from it.

Gross Conversion:

Number of users who enrolled in the free trial/Number of users who clicked the Start Free Trial button. Not a good invariant metric because the number of users who enroll in the free trial is dependent on

the experiment. Good evaluation metric because it is directly dependent on the effect of the experiment, and also shows positive financial outcome of the change.

Retention:

Number of user-ids to remain enrolled for 14 days trial period and make their first payment/Number of users who enrolled in the free trial. Not a good invariant metric because the number of users who enroll in the free trial is dependent on the experiment. Good evaluation metric because it is directly dependent on the effect of the experiment, and also shows positive financial outcome of the change.

Net Conversion:

Number of user-ids remained enrolled for 14 days trial and at least make their first payment/Number of users clicked the Start Free Trial button. Not a good invariant metric because the number of users who enroll in the free trial is dependent on the experiment. Good evaluation metric because it is directly dependent on the effect of the experiment, and also shows positive financial outcome of the change.

I will look at **Gross Conversion** and **Net Conversion**. The first metric will show us whether we lower our costs by introducing the screener. The second metric will show how the change affects our revenues.

To launch the experiment, I will require **Gross Conversion** to have practically **significant decrease**, and **Net Conversion** to have a statistically **no harm to revenue (increase or stay the same i.e. not decrease)**.

MEASURING STANDARD DEVIATION

Evaluation Metric	Standard Deviation
Gross Conversion	0.0202
Net Conversion	0.0156

The Unit of Diversion is equal to the Unit Of Analysis for Gross Conversion and Net Conversion, i.e. both have number of cookies in their denominator. We can therefore proceed using an analytical estimate of the variance.

SIZING

Number Of Samples v/s Power

I did not use Bonferroni correction during the analysis phase.

The evaluation metrics I selected to proceed with are **Gross Conversion** and **Net Conversion**.

I will need **685,324** pageviews to power the experiment with these metrics. That is, 2x (control + experiment groups) the number of samples required for the more demanding of the two metrics, i.e. **Net Conversion**.

Duration v/s Exposure

I would divert 70% of the traffic of Udacity to this experiment given that the experiment will take 25 days, which is a reasonable time for our needs.

The experiment is not extremely risky given that it does not affect existing paying customers, and is simple enough that there is a low chance of bugs occurring in the process. Nevertheless, it may have a substantial impact on new enrollments, and diverting 100% of the traffic may thus not be advisable.

Experiment Analysis

Sanity Checks

1. Number of Cookies:

Number Of Cookies = **[0.4988, 0.5012]**

Observed Value = 344,660 / 690,203 = **0.5006**

which **passes the Sanity Check**.

2. Number of Clicks:

Number Of Clicks on 'Start Free Trial' = **[0.4959, 0.5041]**

Observed Value = 344,660 / 690,203 = **0.5005**

which **passes the Sanity Check**.

3. Click-through Probability:

Click-through Probability on 'Start Free Trial' = **[0.0812, 0.0830]**

Observed Value = **0.0822**

which **passes the Sanity Check**.

Result Analysis

EFFECTIVE SIZE TESTS

1. Gross Conversion:

d_min : minimum practical significance = +/- 0.01

Confidence Interval	[-0.0291, -0.0120]
Statiscally Significant	Yes, since CI doesn't contain zero.
Practically Significant	Yes, since CI doesn't contain d_min value.

2. Net Conversion:

d_min : minimum practical significance = +/- 0.0075

Confidence Interval	[-0.0116, 0.0019]
Statiscally Significant	No
Practically Significant	No

SIGN TESTS

1. Gross Conversion:

Number Of Successes	4
Number Of Trials	23
Probability	0.5
Two-tailed p-value	0.0026

Since, it is less than the given alpha level, it is statistically significant. It agrees with the Hypothesis Test for Gross Conversion which was statistically and practically significant.

2. Net Conversion:

Number Of Successes	10
Number Of Trials	23
Probability	0.5
Two-tailed p-value	0.6776

As we can see the result is not less than the given alpha_level and hence is statistically insignificant. This is in agreement with that of the Hypothesis Test for Net Conversion which was also statistically insignificant.

SUMMARY

The null hypotheses is that there is no difference in the evaluation metrics between the groups, furthermore, a practical significance threshold. Because our acceptance criteria requires statistically significant differences for ALL evaluation metrics, the use of Bonferroni correction is not appropriate. The Bonferroni correction is a method for controlling for type I errors (false positives) when using multiple metrics in which relevance of ANY of the metrics matches the hypothesis. In this case the risk of type I errors increases as the number of metrics increases (significance by random chance). In our case in which ALL metrics must be relevant to launch, the risk of type II errors (false negatives) increases as the number of metrics increases, so it stands to reason that controlling for false positives is not consistent with our acceptance criteria.

Recommendation

The metrics I was interested in were **Gross Conversion** and **Net Conversion**.

Gross Conversion turned out to be negative and practically significant. This is a good outcome because we lower our costs by discouraging trial signups that are unlikely to convert. Net conversion unfortunately ended up being statistically and practically insignificant and the confidence interval includes negative numbers. Therefore, there is a risk that the introduction of the trial screener may lead to a decrease in revenue.

We should therefore, consider testing other designs of the screener before we decide whether to release the feature, or abandon the idea entirely.

Follow-Up Experiment

From your favorite Genius handling your case in the Apple Store, to concierge services of AmEx Centurion – we love when one, real person is assigned to us in times of need, and especially when we pay a lot of money for a product of service of the company.

Udacity does a great job at offering a variety of ways for students to get help, be it through the discussion forum, office hours, or final project reviews. All of these however fail in two important aspects:

1. They require the students to make the first move. (As an aside, in case of the Discussion forum and Office hours, the user interface of the site makes it very challenging to find how to make that move, even if one is determined to seek help. Why is there no Discussion forum link in the navbar at the top of the site?)
2. They tend to expose the student to numerous people, with no one clear point of contact throughout the Nanodegree. (In the real world, this person would be either your Study Advisor, Faculty Advisor or assigned senior in your freshman year at college.)

Udacity could consider implementing a similar system for the Nanodegree program. When a user joins the trial program, he or she will receive an on-site message and a subsequent email from a randomly assigned member of the Udacity team. This message will introduce them to their concierge/guru/mentor and encourage them to reach out to this team member whenever they need help.

Periodic check-up emails should be set up, coming from the assigned team member's email address. He or she should also be notified of new discussion forum posts made by the user in case they happen to have the expertise to answer the question(s).

My **Null Hypothesis** is that assigning a single point of contact to new trial signups will not increase retention by a practically significant amount.

New free trial signups will randomly be assigned to a Control and an Experiment group. The experience for users in the Control group will remain unchanged. Users in the Experiment group will be assigned a random member of the Udacity team and receive an on-site onboarding message and one email follow up from that person.

The **Unit of Diversion** will be the **User IDs** as this change only impacts what happens after a free trial account is created.

The **Invariant Metric** will be the **number of User IDs** because the users sign up for the free trial before they are assigned a point of contact and are exposed to the new onboarding messages.

The **Evaluation Metric** will be **Retention**, which, if positive and practically significant, will show an increase in revenue resulting from this change.

If Retention is positive and practically significant at the end of the experiment, we can launch the new feature, and expand it with more regular follow up emails and personalized on-site messages throughout the Nanodegree program.