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, Number of Clicks

Evaluation Metrics: Gross Conversion, Retention, Net Conversion

Number of cookies:

Represents the number of unique cookies to view the course overview page.

Chosen as invariant metric because the number of cookies are independent from the experiments since the visits happen before the user sees the experiment.

Number of clicks:

Represents the number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is trigger).

Chosen as an invariant metric for the same reason as the number of cookies. It happens before the experiment, so it will be the same for control and experiment group (equal probability of clicking the Start Free Trial button by both groups)

Click-through-probability:

Represents the number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page.

Chosen as an invariant metric for the same reason as both metrics above (happens before the free trial screen).

Gross conversion:

Represents the number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button.

Chosen as an evaluation metric because it directly depends on the effect of the experiment

Retention:

Represents the number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout.

Chosen as an evaluation metric because it directly depends on the effect of the experiment and indicates financial gain resulted from the change

Net conversion:

Represents the number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button.

Chosen as an evaluation metric because it directly depends on the effect of the experiment and indicates financial gain resulted from the change

Number of user-ids:

Represents the number of users who enrol in the free trial. Not chosen as an invariant metric because the number of users who enroll in the free trial is dependent on the experiment. Neither chosen as an ideal evaluation metric because the number of visitor might reasonably be different between the control and experiment groups,

**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 each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to 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.

Standard deviation will be calculated using the formula:   
σ = (*p* (1 − *p*)/*n*) 1/2

That's the standard deviation of a binomial distribution

40000 unique cookies view the page per day. From those, 3200 will click on the "Start free trial" button (0.08% of them). That’s the click-through-probability on "Start free trial" button;

On the current format, from all the 3200 clicks, 660 will enrol on the course, which gives a probability of enrolling of 0.20625 %.

Again, on the current format, the probability of payment is 0.53%, which results on the probability of payment of 0.1093125 %

|  |  |
| --- | --- |
| Unique cookies to view page per day: | 40000 |
| Unique cookies to click "Start free trial" per day: | 3200 |
| Click-through-probability on "Start free trial": | 0.08 |
| Enrolments per day: | 660 |

|  |  |  |
| --- | --- | --- |
| Gross Conversion | Probability of enrolling, given click: | 0.20625 |
| Retention | Probability of payment, given enroll: | 0.53 |
| Net Conversion | Probability of payment, given click | 0.1093125 |

Given a Sample size of 500:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Numerator | Denominator | Division | Square Root (Standard Deviation) |
| *p*(1 − *p*) | *n* |
| Gross Conversion | 0.163710938 | 400 | 0.000409 | 0.02023 |
| Retention | 0.2491 | 82.5 | 0.003019 | 0.05495 |
| Net Conversion | 0.097363277 | 400 | 0.000243 | 0.01560 |

**Sizing**

**Number of Samples vs. Power**

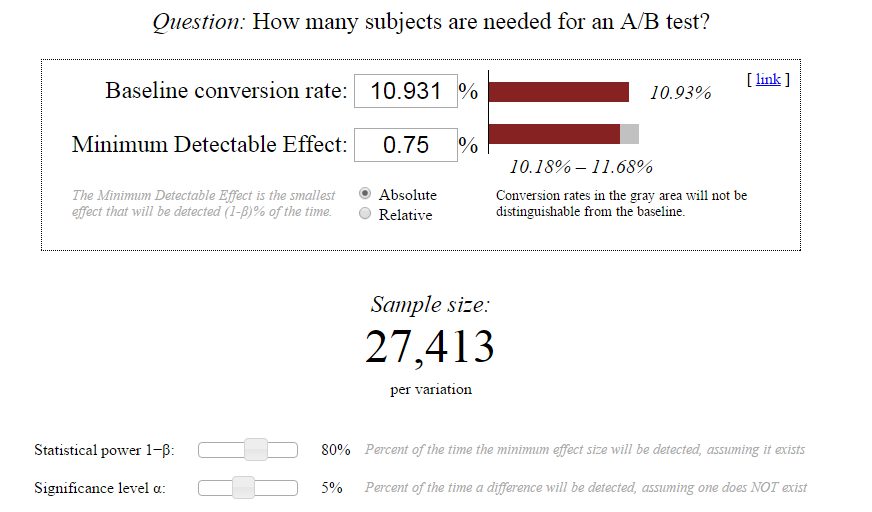
Indicate whether you will use the Bonferroni correction during your analysis phase, and give the number of pageviews you will need to power you experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

Base Line Conversion rate: click-through-probability before making the change

Minimal detectable Effect: practical significance level

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Baseline conversion rate | dmin | Samples Needed | Page Views | times 2 (Experiment and Control groups) |
| Gross Conversion | 20.625% | 1.00% | 25,835 | 322,938 | 645,875 |
| Retention | 53.000% | 1.00% | 39,115 | 2,370,606 | 4,741,212 |
| Net Conversion | 10.931% | 0.75% | 27,413 | 342,663 | 685,325 |

<http://www.evanmiller.org/ab-testing/sample-size.html>



I did not use the Bonferroni correction.

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

Since to evaluate retention would need over 8M page views, I’ve decide to use Gorss Conversion and Net Conversion so 685 324 pageviews will be necessary

**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.)

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?

100% of traffic, 18 days

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.)

For any sanity check that did not pass, explain your best guess as to what went wrong based on the day-by-day data. **Do not proceed to the rest of the analysis unless all sanity checks pass.**

Checking if the numbers on the control group is the same as the numbers on the experiment group:

Page views:

Control: 345543

Experiment: 344660

Random event with exactly two outcomes (probability of assignment = 0.5) – binomial distribution.

So we’ll need to calculate the binomial confidence interval.

Total sample size: 690203 which is enough to assume a normal distribution. (The binomial distribution assumes a normal distribution when n is large)

1. Compute the standard deviation of a binomial distribution with probability of 0.5 of success:

SD = 0.0006

1. Calculate the margin of error with a 95% confidence interval:

m = SD \* 1.96  
m = 0.001180

1. Calculate the confidence Interval (5% of the time the observed values should fall into this range):

CI = [0.5 –m, 0.5 +m]  
CI = [0.4988, 0.5012]

1. Check if the observed values is within the interval:

Values on the control group:

1 – (344,660 / 690,203) = 0.5006

Here’s the summary:



**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.)



We can only understand whether or not a user have gone through the full enrolment process until the 2nd of November, that’s why the total clicks on the formulas will only include that interval



For Gross Conversion, we calculate the “p – experiment” by dividing number of enrolments by number of clicks on the experiment totals and the “p – control” by doing the same calculation on the control Group. The Net Conversion follows the same logic but it uses the number of payments on the numerator:



Then we calculate the pooled probability by using the same formula on the total values and the standard error is calculated by the formula:

σ = (*p* (1 − *p*)/*n*) ½

n in this case being:

*(1/n-control + 1/n-experiment)*

Because the sample size on the control is different than the sample size on the experiment.

Next we calculate the margin of error by multiplying the SE by the z-score for 95% confidence (1.96) and apply the margin of error to the “d” to get the Lower and Upper CI boundaries:



Considering that the practical significance (d\_min) for gross conversion is +/- 0.01 and for Net Conversion is 0.0075:

**Gross conversion: Is statistically significant because the CI range doesn’t contain zero and practically significant because the CI doesn’t contain the d\_min value.**

**Net conversion: Is not statistically significant because the CI range contains zero and is not practically significant because the CI contains the d\_min value.**

**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.)

To do a sign test using the day-by-day data, we need to know how many successes we’ve for each measure and the total number of trials. Success can be considered when the metric was bigger on the experiment than on the control.



Looking at the table above we can see that there were 4 successes for Gross Conversion and 10 successes for Net conversion. The number of trials is 23 in both cases.

Using an online calculator like [this](http://graphpad.com/quickcalcs/binomial1.cfm), we can plug in the values and get the two-tailed p value, which represents the chance of observing either X or fewer successes, or Y-X or more successes, in Y trials.



**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.

**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.

<https://en.wikipedia.org/wiki/Binomial_distribution>

<https://en.wikipedia.org/wiki/Bernoulli_distribution>