Template Format

This template can be used to organize your answers to the final project. Items that should be copied from your answers to the quizzes should be given in blue.

Experiment Design

**Metric Choice**

The invariant metrics chosen for this test were:

* Number of cookies (The number of unique cookies to view the course overview page)
* Number of clicks (The number of unique cookies to click the “Start free trial” button)
* Click-through-probability (Number of Clicks/Number of cookies)

All three of the invariant metrics were chosen because they are recorded prior to the testing protocol – they could not have been affected by the test procedure.

The evaluation metrics chosen for this test were:

* Gross conversion (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
* Retention (Number of user-ids to remain enrolled past the 14-day boundary (and make at least one payment) divided by number of user-ids to complete checkout)
* Net Conversion (Number of user-ids to remain enrolled past the 14-day boundary (and make at least one payment) divided by number of unique cookies to click “Start free trial” button.

Gross conversion was chosen because it is first point of diversion – were some users swayed to attempt free classes rather than start the trial. If the change is working, and users with less than five hours per week are being diverted to free class rather than the free trial, we would expect this metric to decrease in our test group vs the control group.

Retention was chosen as it is the primary metric we are attempting to increase. The assumption of this test is that if people who do not have the time to make it through the material every week will no longer choose to start the free trial. Thus, those that do begin the trial are controlled for personal time availability, and should be retained at a higher level, all else being equal.

Net conversion was chosen because it is expected to vary with the addition of the time estimate prompt. My assumption is that Net conversion will be reduced if the change is working as expected.

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.

**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 |
| Gross Conversion | 0.0202 |
| Retention | 0.0549 |
| 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.

I believe the analytical estimate of the standard deviation for gross conversion and net conversion is adequate due to the underlying sample sizes. Given time, I’d like to empirically calculate the standard deviation for retention, as the enrollments over our 40000 page view sample was only 660.

**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.)

I will not be using the Bonferroni correction, as the three evaluation metrics I chose are likely covariant.

This experiment will 685,325 pageviews.

**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.)

I will need to divert 50% of Udacity’s traffic for this experiment for a duration of 35 days. I don’t believe diverting more traffic would be harmful for Udacity, but I wanted to gather data over a full month to control for potential Simpson effects while maintaining a compact experimental period.

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**

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.)

|  |  |  |
| --- | --- | --- |
|  | Pageviews | Clicks |
| probability | 0.5 | 0.5 |
| SE | 0.0006019 | 0.002100 |
| ME | 0.0011797 | 0.004116 |
| low CI bound | 0.4988 | 0.4959 |
| Upper CI bound | 0.5012 | 0.5041 |
| Observed | 0.5007 | 0.5005 |

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.**

**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.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Result Analysis |  |  |  |  |  |
|  |  |  |  |  |  |
|  | Clicks | Enrollments | Gross conversion | Payments | Net conversion |
| Control | 17293 | 3785 | 0.21887 | 2033 | 0.11756 |
| Experiment | 17260 | 3423 | 0.19832 | 1945 | 0.11269 |
| p |  |  | 0.20860 |  | 0.11513 |
| d |  |  | -0.02055 |  | -0.00487 |
| Empirical SE |  |  | 0.02020 |  | 0.01560 |
| SE |  |  | 0.00437 |  | 0.00343 |
| m |  |  | 0.00857 |  | 0.00673 |
| Upper CI |  |  | -0.01199 |  | 0.00186 |
| Lower CI |  |  | -0.02912 |  | -0.01160 |

**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.)

**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.