# **Design and Analyze A/B Testing - Udacity Free Trial Screener**

# Experiment Design

# Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here:

**Invariant Metrics:**

* Number of services: Number of cookies to view the course overview page
* Number of clicks: Number of unique cookies to click the “start free trial” button (happens before the free trial screener is trigger)
* Click through probability: Number of unique cookies to click the “start free trial” button (which happens before the free trial screener is trigger)

**Evaluation Metrics:**

* Gross conversion: Number of user-ids to complete checkout and enroll in “the free trial button” divided by number of unique cookies to view the course overview page.
* Net Conversion: 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.
* Retention

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 services: Number of cookies to view the course overview page
* Number of clicks: Number of unique cookies to click the “start free trial” button (happens before the free trial screener is trigger)
* Click through probability: Number of unique cookies to click the “start free trial” button (which happens before the free trial screener is trigger)

**Evaluation Metrics:**

* Gross conversion: Number of user-ids to complete checkout and enroll in “the free trial button” divided by number of unique cookies to view the course overview page.
* Net Conversion: 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.

# Measuring Standard Deviation

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.

# 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.)

# 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?

# 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.**

# 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.)

# 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.