Udacity, A/B Testing

July 8, 2016

1 Metrics Choice

Invariant Metrics:

- Number of Cookies
- Number of Clicks

Evaluation Metrics:

- Gross Conversion
- Net Conversion
- Retention

2 Variability

Given the following data about the experiment:

- Unique cookies to view page per day $N_{view} = 40000$
- Unique cookies to click "Start free trial" per day $N_{click} = 3200$
- Enrolments per day $E = 660$
- Click-through probability on "Start free trial" $P(click) = 0.08$
- $P(E|click) = 0.20625$
- $P(Pay|E) = 0.53$
- $P(Pay|Click) = 0.1093125$

If we assume a sample size $S = 5000$ unique cookies visiting the page, based on the original data, the number of clicks would be:

$$\hat{N}_{click} = S * P(click) = 400 \tag{1}$$

and the number of enrolments:

$$\hat{N}_{enroll} = S * P(click) * P(E|click) = 82.5 \tag{2}$$

Given that all the chosen metrics can be assumed to follow a Binomial distribution, the estimate of Standard Error follows:

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} \tag{3}$$

so for each metric, the SE is:

- Gross Conversion: 0.020231
- Retention: 0.0549
- Net conversion: 0.015602

3 Sizing

3.1 Number of Samples

With a type I error rate of $\alpha = 0.05$ and a type II error of $\beta = 0.2$, the minimum detectable effects are $d_{gc,min} = 0.01$, $d_{r,min} = 0.01$, $d_{nc,min} = 0.0075$. I will not be using the Bonferroni correction, because the measures are covariant.

The ratios of clicks to pageviews is $\frac{N_{clicks}}{N_{views}} = 0.08$ and the ration of conversions to pageviews is $\frac{N_{enroll}}{N_{views}} = 0.0165$, which will be used for corrections.

The non-corrected experiment sizes are as follows:

- Gross Conversion: 25835
- Net Conversion: 27413

After the corrections they are as follows:

- Gross Conversion: 322938
- Net Conversion: 342662.5

Doubling the maximum because we need this number of pageviews for each hypothesis, we get a total of 685325 pageviews required.

3.2 Duration vs Exposure

Given the relatively low risk nature of the experiment[1], it makes sense to divert a large amount of traffic to it to reach a conclusion quickly. However, it also makes sense to keep a small amount of traffic out of the experiment, in case there was an error or bug in the experiment, so that it can be detected quickly. Therefore I would say that diverting 90% of Udacity traffic makes sense. Give a daily traffic of 40000 pageviews, this means there would be 36000 pageviews dedicated to the experiment. The total duration of the experiment would be of 19.03 days, which will be rounded to 20. This is an acceptable time.

4 Sanity Checks

5 Effect Size Tests

6 Sign Tests

7 Results Summary

8 Recommendation

9 Follow-up Experiment

[1]It does not affect any of the pages with content, so it does not affect existing users, and it is only adding a small prompt that wouldn't be considered annoying by most people