Problem Statement

The neighboring cities of Gotham and Metropolis have complementary circadian rhythms: on weekdays, Ultimate Gotham is most active at night, and Ultimate Metropolis is most active during the day. On weekends, there is reasonable activity in both cities.

However, a toll bridge, with a two way toll, between the two cities causes driver partners to tend to be exclusive to each city. The Ultimate managers of city operations for the two cities have proposed an experiment to encourage driver partners to be available in both cities, by reimbursing all toll costs.

1. *What would you choose as the key measure of success of this experiment in encouraging driver partners to serve both cities, and why would you choose this metric?*

2. *Describe a practical experiment you would design to compare the effectiveness of the proposed change in relation to the key measure of success. Please provide details on:*
    a. *how you will implement the experiment*
    b. *what statistical test(s) you will conduct to verify the significance of the observation*
    c. *how you would interpret the results and provide recommendations to the city operations team along with any caveats*

Solution

1 Key Measure:

Definition

Base city: the primary city served by the driver prior to experiment (i.e., the one in which most of the driver's trips are conducted).

Options:

Option zero:

I would use the difference (before and after the incentive) in percentage of trips per driver that are from the non-base city for the driver and compare the means for the control vs treatment group

(before / after experiment with control group)

Other options are:

These are side-by-side experiments as opposed to before/after.

1. The aggregate percentage of trips that are from the non-base city for drivers.
2. The mean of the percentage of each driver's trips that are from the non-base city.

Option zero measure the effects of individual behavior, before and after the incentive. There would still be a control group as a comparison. It also measures changes in either direction for a driver (i.e., we would see both negative and positive effects, rather than an overall average; in an aggregated measure negative and positive effects would not be captured.

Option 1 measures aggregate effect but could be influenced by a small number of drivers with a large number of trips.

Options 2 measures individual behavior, rather than aggregate behavior, but could be affected by drivers with a small number of trips.

2 Experiment Design:

a) Experiment implementation

- Use two cohorts:
  - Base-city = Gotham

    - Base city = Metropolis
- Determine significance level (e.g., alpha = .05)
- Determine driver sample size
- Determine length of time (i.e., number of samples) needed to measure results
- For each cohort
    - Randomly assign drivers into a control and test group:
        - Test group: gets reimbursed for tolls
        - Control group: status quo: no toll reimbursement
    - NOTE: we are using a non-treatment control group over the same time period and measuring before/after differences for the control and treatment groups rather than have the 'before' behavior be the control; this mitigates for other factors such as time of year, special events, weather conditions, etc.

b) Statistical test

Use a two-sample t-statistic to determine if there is a difference in the means of the test group vs the control group.

NOTE: this assumes the percentages are normally distributed.

If the sample values are not normally distributed, then can use hacker statistics to compare the results.

X = sample statistic: change in percent of trips that are from non-base city before and after the incentive

Xbar-1 : mean percentage change of control group

Xbar-2: mean percentage change of test group

Hypotheses:

H0: u2 – u1 = 0

H1: u2 – u1 <> 0

In general, the population standard deviations are not known, and are estimated by the calculated values s1 and s2. In this case, the test statistic is defined by the two-sample t statistic

From <http://www.stat.yale.edu/Courses/1997-98/101/meancomp.htm>

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

---

c) Result interpretation and recommendation

Present the change observed, and if it were significant.

Explain that policy makers should consider cost/benefit for the observed amount of change.

Caveats:

- experiment is not independent events and does not account for potential interactions when more drivers are offered the incentive.
-
- Could also have unknown side effects.

- Other factors may influence actual results…experiment was done over a limited timeframe that may not be representative of overall conditions, such as time of year, weather, etc
- Before/after behavior could be affected by time of year
    - This is mitigated by the control group having the same circumstances