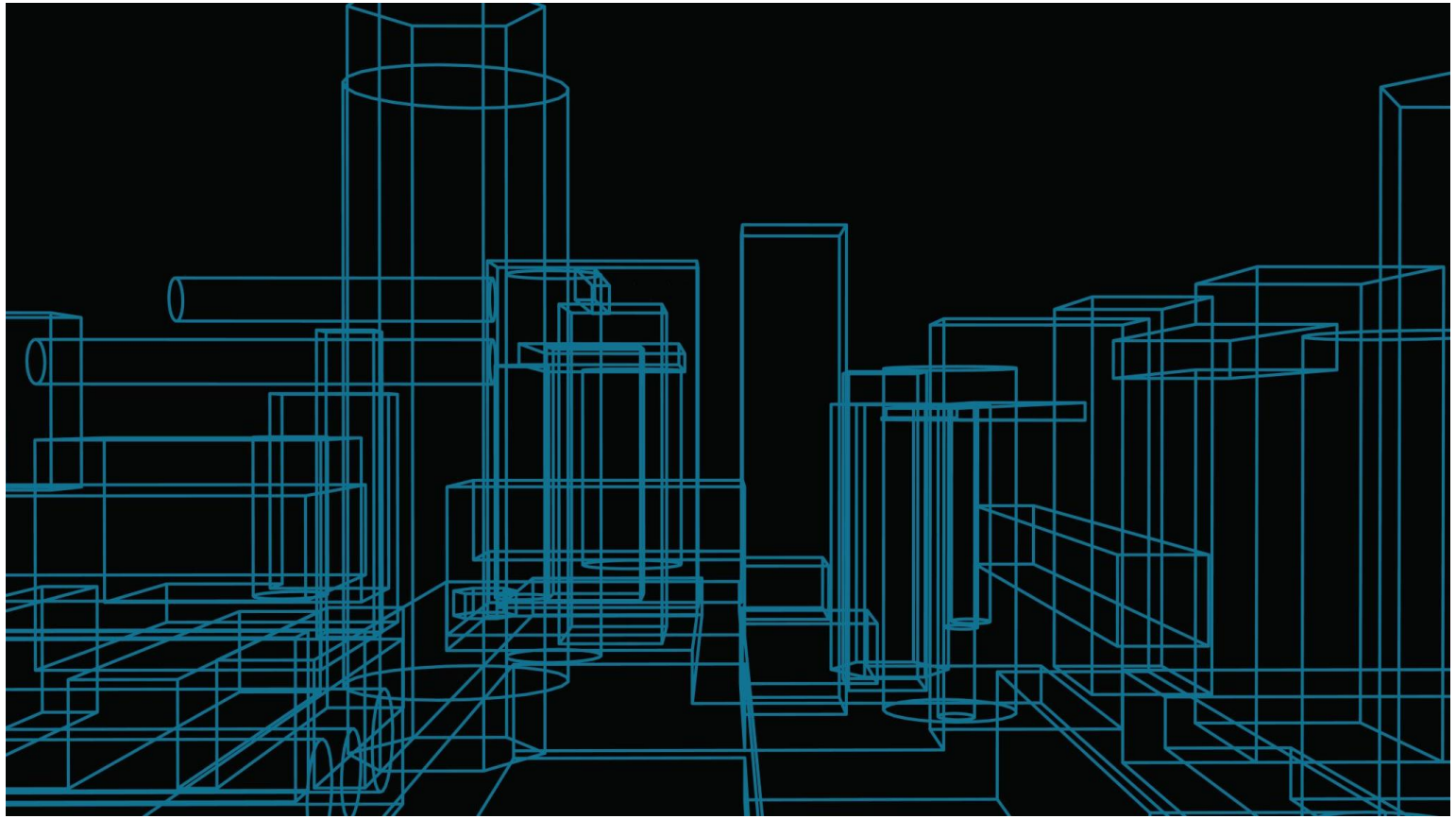
Ultimate Technologies

Part 2 Experiment Metrics and Design



Experiment Metrics and Design - Plan

- Company Goal – Have two different fleets of driver partners spend the same amount of time on two different cities
- Program - The company plans to start reimbursing all toll cost to drivers on both cities to encourage them to be available in both cities
- We need data from last 3 months prior to the start of the program.
- The following features would be helpful in a dataset for experimentation:
- stop_id, time_of_day, day_of_week, cost_per_stop , revenue_per_stop , profit_per_stop, distance
- Collect the same data for at least 3 months after the program has had a solid start and is widely adopted

Experiment Metrics and Design - Plan

- Collect the agreed upon data for each fleet of driver partners
- We will assume the fleet sizes are the same for each city. Otherwise, we can adjust calculations by using percentages

Experiment Metrics and Design - Plan

- We will assume we have all the data on each the populations (the population of stops for both fleets)
- This is a reasonable assumption since we are tracking each stop for each fleet, but each stop is tracked as a transaction by necessity
- Because of the frequency of transactions, we can consider a large n (number of samples)
- No further statistical work would have to be done to make sure the samples have enough power as the frequency of transactions is expected to be high
- A large n also allows us to use the Law of Large Numbers, and the distributions can be considered to be approximately normal

Experiment Metrics and Design - Measurements

- Since the distribution of stops can be considered to be approximately normal and the because of the previous assumptions we have discussed in the slides, like equal fleet size, we have two similar normal distributions of stops for each of the fleets.
- We will take the difference of the means between the two fleets to be our metric, with a large enough n (number of samples) the expected value of the difference of two random variables simply becomes the difference of the averages which is very straight forward to calculate and monitor (since our averages are approximately normal so is the distribution of differences)

Experiment Metrics and Design - Metric

- the ideal situation would be that after the experiment Ultimate Gotham and Ultimate Metropolis are getting the same number of stops on both cities during busy hours
- Considering our assumptions that the fleets are the same size, and stops are approximately normal, independent distributions, the difference between the average number of stops between the two different fleets in a perfect world would be zero

Experiment Metrics and Design - Metric

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- We can go further, when we feel the distributions have indeed become equal, we can test
- Our metric would be as we discussed - the difference between two independent means
- On the left is the independent sample t-test which we can use to check if the difference is statistically significant.
- More details on next two pages.

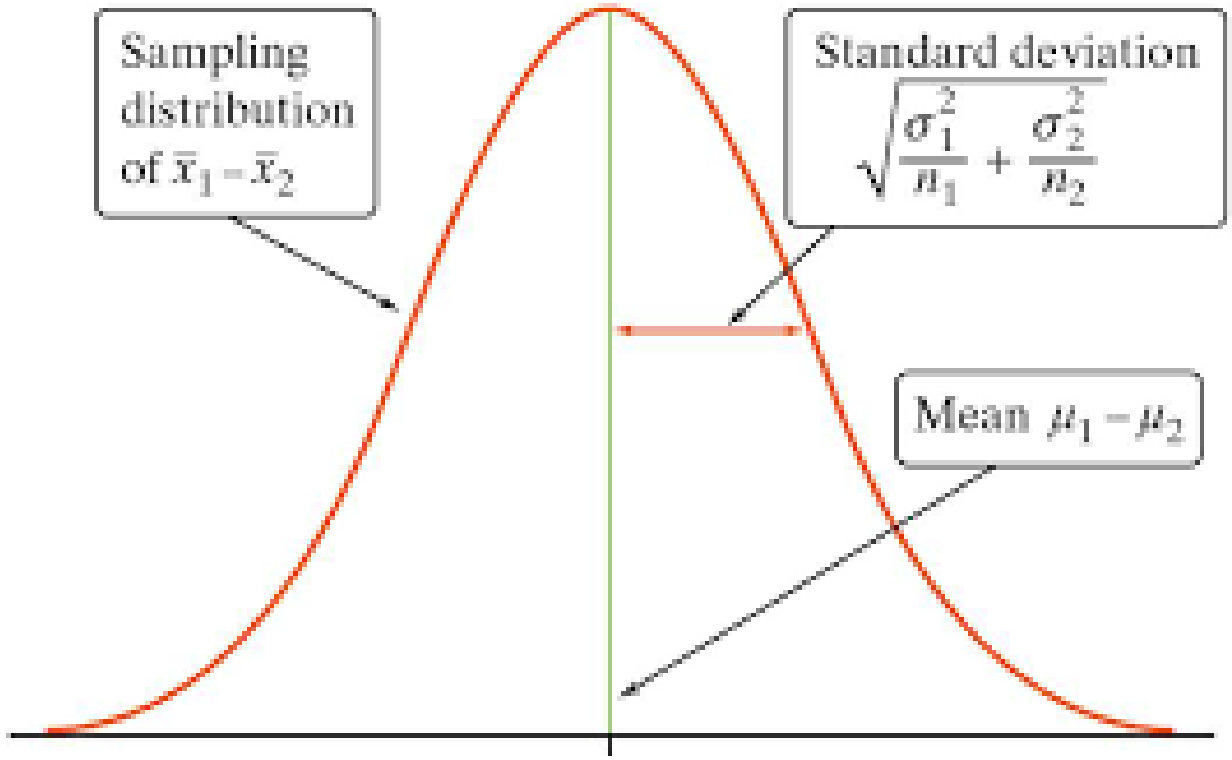
https://mat117.wisconsin.edu/hypothesis-testing-for-the-difference-of-two-independent-means/

Experimental Metrics and Design - Performance

- Based on the useful information provided by the company, we have three tests to perform
- Let's say we wanted perfect serving across both cities by both fleets. We could use hypothesis testing to verify this
- Build one distribution for the ***difference*** in average number of stops between the fleets for each of the following scenarios
- If the 95% CI of the difference contains 0, then there is no difference between groups. If it doesn't contain 0, then there is a statistically significant difference between groups.
- First, we want to test if equal time is being spent by both fleets on weekdays, using information given to us by the company, we'd want to check that there is no statistical difference between the average number of stops between Ultimate Gotham and Ultimate Metropolis fleet stops in Gotham during nighttime

https://towardsdatascience.com/why-overlapping-confidence-intervals-mean-nothing-about-statistical-significance-48360559900a?gi=a1fceb634d45

Experimental Metrics and Design - Performance



- Second scenario is that there is no statistical difference between the average number of stops of each fleet to Metropolis during the day.
- Last scenario is that there is no statistical difference between the average number of stops of each fleet to either city during the weekend.
- No statistical difference implies the perfect situation where fleets are participating equally on both cities

https://www.cusd80.com/cms/lib/AZ01001175/Centricity/Domain/8950/Comparing%20Means%20Notes.pdf

Experiment Metrics and Design - Performance

- In reality, we will be expecting improvements to be more gradual:
- After the start of the program, we should see a significant reduction in the mean difference in each of the three scenarios. We mentioned that with our assumption of a large n the expected value of the difference becomes the difference between the averages of the groups, which we can easily monitor for improvement (reduction in size)
- Ideally as time goes by, with complete participation, the 95% CI of the difference contains 0 and there's no longer a difference between the means.

Experiment metrics and design – summary and potential caveats

- The company is mainly concerned about participation from both driver partner fleets on both cities and so we focused on measuring the success of that goal as the difference in average stops between the fleets for each scenario (a simple measurement with solid statistical grounding that makes sense for this problem). Scenarios are the following:
    - Gotham at night during the weekdays
    - Metropolis at daytime during the weekdays
    - Metropolis during the entire day on the weekend
    - Gotham during the entire day on the weekend
- Feature recommendations for the dataset included items like cost_per_stop, revenue_per_stop , profit_per_stop so that this data will already be collected for further analysis in case any issues come up that affect profitability when trying to maximize participation only
- Adding features such as driver_home_location and fuel_costs may also help assess unexpected expenses
- Competition in certain areas from competitor fleets might also create unexpected changes