Problem Set - Uber Case

Yuxuan Mei

Install the packages if you don't have them installed yet. Install them only once, and not when you run your knitting.

1. Read the HBS Case. What is the difference between Uber POOL and Express POOL? No more than two sentences.

Unlikely Uber POOL where riders stays at the original location to wait for drivers, Express POOL requires riders to walk a short distance and wait a short time.

2. How did Uber use surveys in designing Uber Express Pool?

Uber used surveys to measure consumers' sensitivity to different variables, then used the results to design Express POOL's pricing strategy.

3. Suppose Uber was considering a new algorithm to recommend ride destinations in the app. Which type of research strategy should they use (A/B Test, Switchback, Synthetic Control)? No more than two sentences.

They should use Switchback. Switchback can be used to evaluate the effects of the new algorithm, and its randomization is done on time units instead of user-level units to reduce dependency impacts.

4. Suppose Uber was considering a radio advertising campaign. Which type of research strategy should they use (A/B Test, Switchback, Synthetic Control)? No more than two sentences.

They should use Synthetic Control. Randomization on geographic region can reduce spillover effects, and could help attribute observed changes in aggregate outcomes to the radio advertising campaign.

5. Create two new columns in the dataset that represent the total number of trips for both pool products and the profit from these products. (10 points)

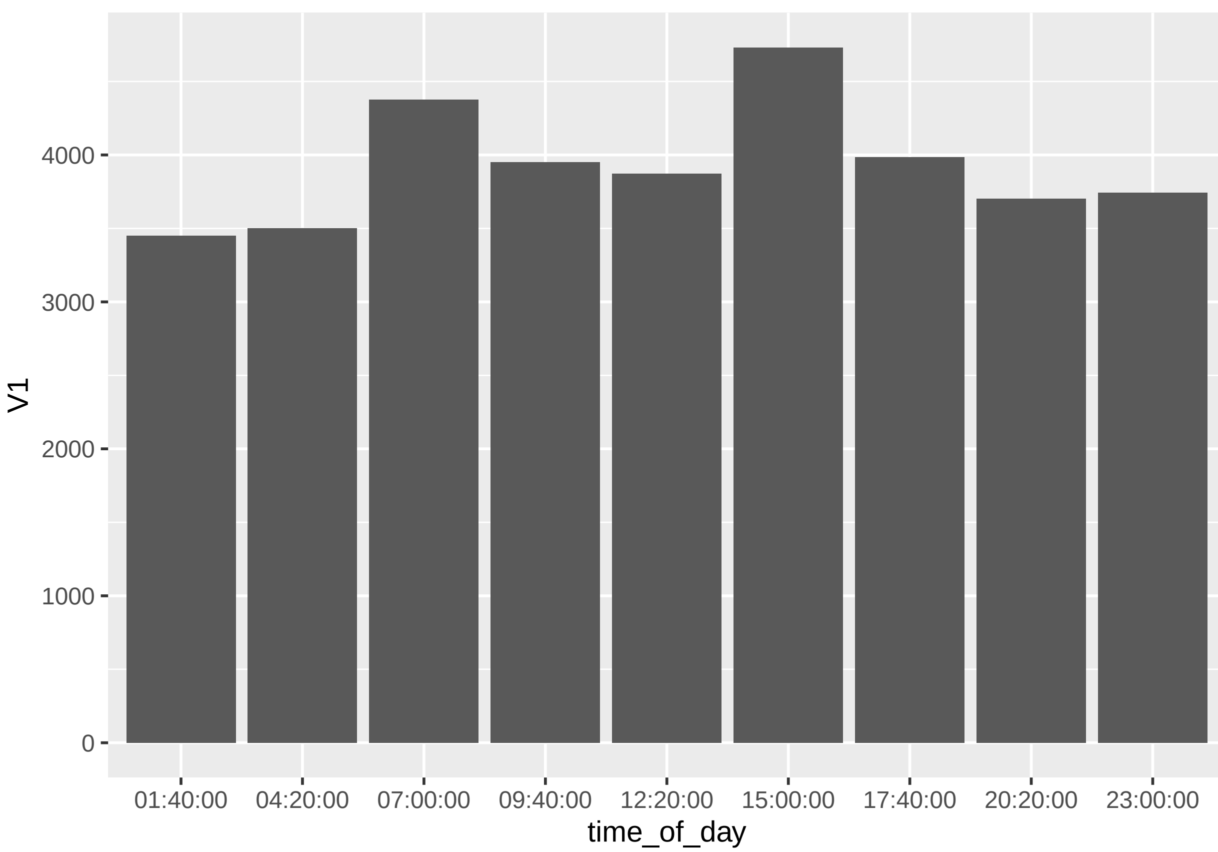
(remember you can create a new column by: data[, new_col_name := whatever you want the new column to contain])

```
data[, sum_trips := trips_pool + trips_express_pool]
data[, sum_profit := revenue - total_driver_payout_sr]
```

6. Plot the average number of trips as a function of the time of the day. Describe a reason why this pattern exists (no more than 2 sentences). (20 points)

Hint: You can use ggplot to do this. As in assignment 1, you'll first have to create a dataset with the average number of trips by time of the day.

```
ntrips_time = data[, mean(sum_trips), time_of_day]
trip_time_plot <- ggplot(data = ntrips_time, aes(x = time_of_day, y = V1)) + geom_bar(stat = 'identity')
trip_time_plot
```



Number of POOL trips increases before/around morning rush hours and before afternoon rush hours. A reason might be that people are reluctant to use POOL services during rush hours because they may perceive POOL rides to be slow and inefficient if they are in a rush.

7. Conduct a regression analysis of the experiment (considering the outcomes: revenue, total__driver__payout__sr, rider__cancellations, total__trips). Make sure to think carefully about the correct regression specification. The regression output should be easy to read, so use 'etable' or 'modelsummary'. What do you learn in words from this regression analysis (no more than 5 sentences but it can be less)?

Hint: We should control for the fact that different times of the day and different days have different demand patterns. (Please refer to p.13 of the HBS article to see why) Hint: The syntax for fixed effects is: feols(outcome ~ treatment_name | fixed_effect_name1 + fixed_effect_name2, data = data, se = 'hetero') Hint: You can output multiple regressions in this way: etable(reg1, reg2)

```
reg1 <- feols(sum_profit ~ treat | time_of_day + date, data = data, se = 'hetero')
reg2 <- feols(rider_cancellations ~ treat | time_of_day + date, data = data, se = 'hetero')
reg3 <- feols(sum_trips ~ treat | time_of_day + date, data = data, se = 'hetero')
reg4 <- feols(revenue ~ treat | time_of_day + date, data = data, se = 'hetero')
reg5 <- feols(total_driver_payout_sr ~ treat | time_of_day + date, data = data, se = 'hetero')
etable(reg1, reg2, reg3, reg4, reg5)
```

```
##                              reg1                reg2           reg3
## Dependent Var.:        sum_profit rider_cancellations      sum_trips
##
## treatTRUE       1,829.8*** (496.8)    26.63*** (7.058) -80.86 (81.19)
## Fixed-Effects:  ------------------ ------------------- --------------
## time_of_day                    Yes                 Yes            Yes
## date                           Yes                 Yes            Yes
## _______________ __________________ ___________________ ______________
## S.E. type       Heteroskedas.-rob. Heteroskedast.-rob. Heterosk.-rob.
## Observations                   126                 126            126
## R2                         0.42068             0.55432        0.50728
## Within R2                  0.11961             0.12685        0.00997
##                          reg4                   reg5
## Dependent Var.:       revenue total_driver_payout_sr
##
## treatTRUE       -170.6 (930.5)      -2,000.4* (775.0)
## Fixed-Effects:  -------------- ----------------------
## time_of_day                Yes                    Yes
## date                       Yes                    Yes
## _______________ ______________ ______________________
## S.E. type       Heterosk.-rob. Heteroskedastici.-rob.
## Observations               126                    126
## R2                     0.57973                0.50234
## Within R2              0.00035                0.06308
```

After adjusting for fixed effects for time of the days and dates, we observed that the treatment (increase waiting time from 2 mins to 5 mins) led to a statistically significant increase in total profit. While the most part of increase in profit came from reduced driver payouts (statistically significant at 5%). The treatment also led to a statistically significant decrease in trip cancellation. The treatment did not have a statistically siginificant impact on total number of trips.

8. One of your data scientists suggests that the optimal wait time may differ by whether it's a commuting period. Test whether the effects of a 5 minute wait period on total trips and cancelations differ by whether it's a commuting period (the column 'commute'). Which policy works better during commute times? (10 points)

```r
reg_c <- feols(sum_trips ~ treat | time_of_day + date, data = data[commute == TRUE])
reg_c1 <-feols(rider_cancellations ~ treat | time_of_day + date, data = data[commute == TRUE])
reg_uc <- feols(sum_trips ~ treat | time_of_day + date, data = data[commute == FALSE])
reg_uc1 <- feols(rider_cancellations ~ treat | time_of_day + date, data = data[commute == FALSE])

etable(reg_c, reg_c1, reg_uc, reg_uc1)
```

```
##                              reg_c               reg_c1         reg_uc
## Dependent Var.:          sum_trips rider_cancellations      sum_trips
##
## treatTRUE       -322*** (NaNe-Inf) 56.80*** (4.13e-15) -44.06 (64.24)
## Fixed-Effects:  ------------------ ------------------- --------------
## time_of_day                    Yes                 Yes            Yes
## date                           Yes                 Yes            Yes
## _______________ __________________ ___________________ ______________
## S.E.: Clustered    by: time_of_day     by: time_of_day by: time_of_..
## Observations                    20                  20            106
## R2                         0.61897             0.71472        0.36839
## Within R2                  0.20985             0.54935        0.00570
##                             reg_uc1
## Dependent Var.: rider_cancellations
##
## treatTRUE            20.53* (7.970)
## Fixed-Effects:  -------------------
## time_of_day                     Yes
## date                            Yes
## _______________ ___________________
## S.E.: Clustered     by: time_of_day
## Observations                    106
## R2                          0.33429
## Within R2                   0.18110
```

We can see that during the commute period, the treatment led to highly statistically significant decrease in total number of trips as well as increase in the number of total cancelled trips. While the impact on total trips during non-commute periods was not statistically significant. Therefore, for commute periods, Uber should probably not implement the treatment because it would cost decreased number of trips and lower user satisfaction significantly (riders probably value time more during commute periods), which would potentially damage Uber's long term profits.