MSAN 631: Design and Analysis of Experiments

Lab 4: The Lyft Experiment
Due: Thursday June 29th, 2016

Problem Description:
Ridesharing as an industry has seen remarkable growth in recent years. Much of this growth is due to an emphasis on rider and driver retention. Two notable ridesharing companies in the Bay Area are Uber and Lyft. Always eager to improve rider satisfaction and retention, Lyft decides to send its riders a promotional offer. There are many factors to consider when designing a promotional offer, for example:

- Discount Amount
- Discount Duration
- Method of Dissemination
- Ride Type

Of primary importance is an understanding of the relationship between these factors and the tendency for a user to book a ride. An understanding of this relationship can help to find a promotional offer that is optimal in the sense that it maximizes the response variable (y): booking rate.

In self-selected groups you will, through experimentation, find the levels of each factor that simultaneously maximize booking rate (y). Any experiment (or experiments) you run must be some version of a 2^k factorial (fractional or full) design. The experimental factors and their levels (in natural units) are provided in the table below.

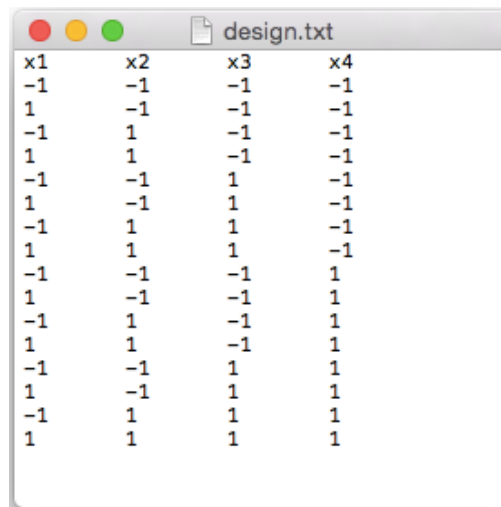| Factor	Description	Low	High
x_1	Discount Amount (%)	10	50
x_2	Discount Duration (days)	1	5
x_3	Method of Dissemination	In-app	Email
x_4	Ride Type	Standard	Shared

Once you have designed the experiment, you must email your design matrix to me (because I have the model which generates the data), and I will, in turn, provide you with response variable measurements. Note that I will perform 1000 replicates of each condition you consider and return to you the proportion of booked rides in each case. Thus no replication is required by you. If you wish to perform follow-up experiments, simply send me more design matrices. Please format your design matrices as follows:

- They should have no more than 4 columns (one column for each main effect you wish to estimate – I will generate the intercept and interaction columns myself as necessary).
- Each column contains ±1's in accordance with the typical factorial structure – what we referred to as the "natural ordering" in class.

- The matrix should be stored in a .txt file called "design.txt". Example R code that generates and saves an appropriately formatted and named .txt file is provided below. An image of an appropriately formatted and named .txt file is also shown below. Note that this corresponds to a single replicate of a full 2^4 factorial design.

```
> library(DoE.base)
> X <- fac.design(factor.names=list(x1=c(-1,1), x2=c(-1,1),
                                     x3=c(-1,1), x4=c(-1,1)),
                                     randomize = FALSE,
                                     replications = 1)
> X <- data.frame(X)
> write.table(x = X[,1:4], file = "design.txt", sep="\t",
              quote=FALSE, dec=".", row.names=FALSE)
```

x1	x2	x3	x4
-1	-1	-1	-1
1	-1	-1	-1
-1	1	-1	-1
1	1	-1	-1
-1	-1	1	-1
1	-1	1	-1
-1	1	1	-1
1	1	1	-1
-1	-1	-1	1
1	-1	-1	1
-1	1	-1	1
1	1	-1	1
-1	-1	1	1
1	-1	1	1
-1	1	1	1
1	1	1	1

Use a regression analysis to identify and quantify significant main and interaction effects. Use your fitted model to decide which promotional offer appears to maximize the booking rate. Also, be sure to calculate a point estimate of this maximum and a 95% prediction interval for this response at the optimal factor settings.

Deliverable:
Please prepare a short report that describes the experimental design strategy you have implemented and the corresponding analyses. In particular, describe the analysis method, justifying the approach you have taken, and any assumptions that have been made. Describe your analysis and any notable results that were found. Finally, indicate the promotional offer that maximizes booking rate and report a point and interval estimate of the booking rate that can be expected for this promotion. This report should summarize your findings in a clear and coherent manner.

Grading:
The members of each group will receive a common grade out of 20. Groups will be graded on the adequacy and completeness of the analysis and the reports will be graded on the quality of the written exposition.