Statistical Modelling the Inventory Allocation of Bay Area Bike Share

Xiaoxiao Zhao, Junrong Wu, Tatjana Miljkovic
Department of Statistics
Miami University

Motivation and Objective

With the development of transportation and living standard in modern society, pollution from vehicles and traffic jam has become a significant issue in our daily life. As a result, Bike sharing system gains more attention in recent years. (**Figure 1**) Public bike sharing program allows citizen to rent and return bikes in any station points distributed within the city.

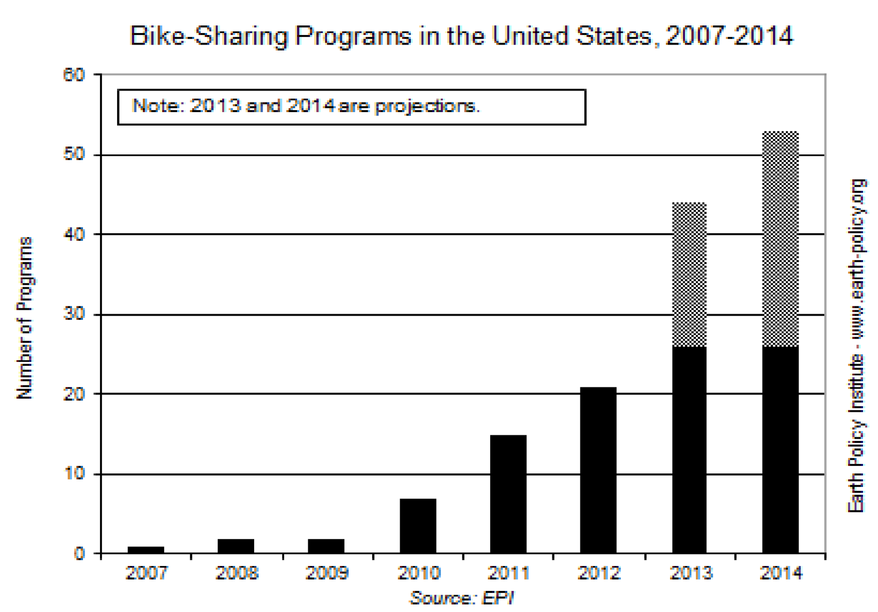

Figure 1: Bike-Sharing Programs in U.S., 2007-2014

Bay Area Bike Share(BABS) is a bike sharing system based in San Francisco bay area operated by Motivate. It provides a green, cost-saving, and more efficient way of transportation for its customers. As a healthy life-style, bike sharing system has been spread to masses of cities and it will be expanded more in the future. (**Figure 2**)

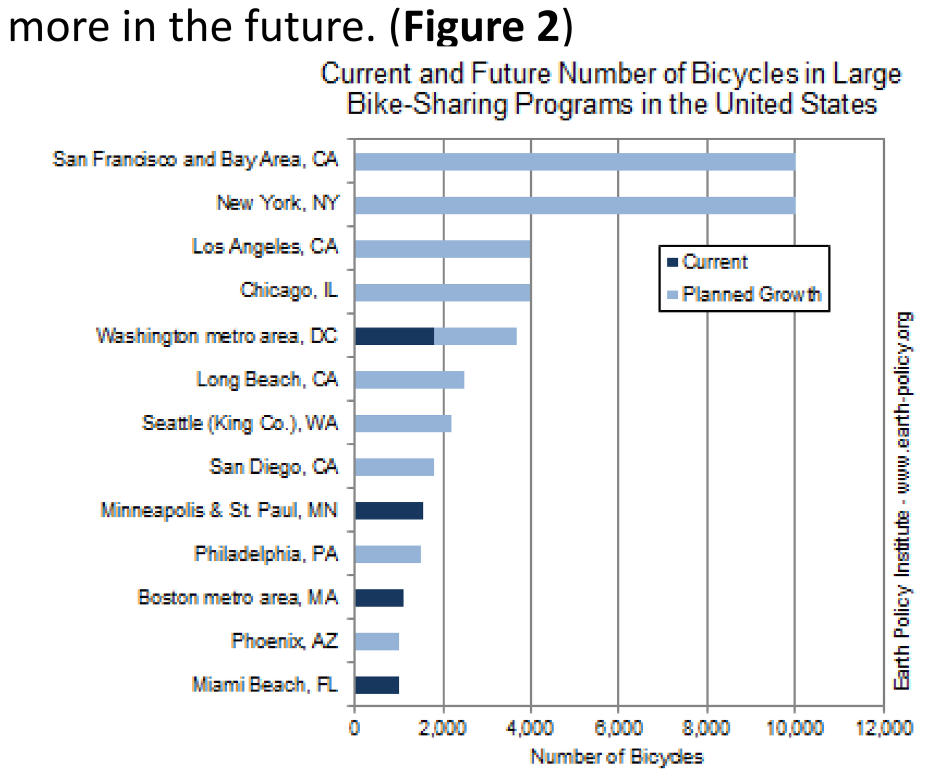

Figure 2: Current and Future Number of Bicycles in Large Bike-Sharing Programs in the U.S.

Biking is an outdoor activity, weather conditions will be considered as important factors that may affect people's decision. San Francisco, located in the west coast of the United States and surrounded by ocean, has changeable weather because of its special geographic conditions. Therefore, the research topic is to investigate **how weather conditions impact the rental bikes.** By discovering that, the firm will better allocate its inventory and achieve higher profits which will ultimately help them to expand its business and promote this green way of transportation.

Introduction

This project is going to perform generalized linear model with Poisson distribution by using six independent weather variables to predict the count of rental bikes under BABS in San Francisco.

The data is collected from 29th August, 2013 to 31st August, 2015, and there are 733 observations in total.

Methodology

The method here is to use Poisson Regression. Since the response variable in our research is the count of rental bikes per day under BABS in San Francisco, we consider a generalized linear model with link log：

$$\log(\mu_i) = x_i'\beta,$$

where μ_i is the mean of response variable, x_i is the vector of independent variables and β represents the vector of regression coefficients.
In this model, increasing the predictor x_j by one unit, is associated with an increase of β_j, the regression coefficient, in the log of the mean.

Based on the formula of Poisson distribution, we obtain our Log-likelihood function for the parameter estimation as

$$\log L(\beta) = \sum \{y_i \log(\mu_i) - \mu_i\}.$$

Model

The response variable in our research has been distributed into two groups (**Figure 3**). We discovered that the average of number of rental bikes has a significant difference between weekdays and weekends (**Figure 4**). Therefore, we obtained two log-linear regression models for each group:

$$\log(\mu_{Weekdays}) = x'_{Weekdays}\beta_{Weekdays}$$
$$\log(\mu_{Weekends}) = x'_{Weekends}\beta_{Weekends}.$$

In these two models, $\mu_{Weekdays}$ and $\mu_{Weekends}$ are the mean of number of rental bikes in weekdays and weekends, $x_{Weekdays}$ and $x_{Weekends}$ are vectors of predictors containing temperature, humidity, wind speed, visibility, cloud cover and events, where events defined as all clear day, foggy day, rainy day, fog and rain and thunderstorm. $\beta_{Weekdays}$ and $\beta_{Weekends}$ represent vectors of regression coefficients for weekdays and weekends respectively.

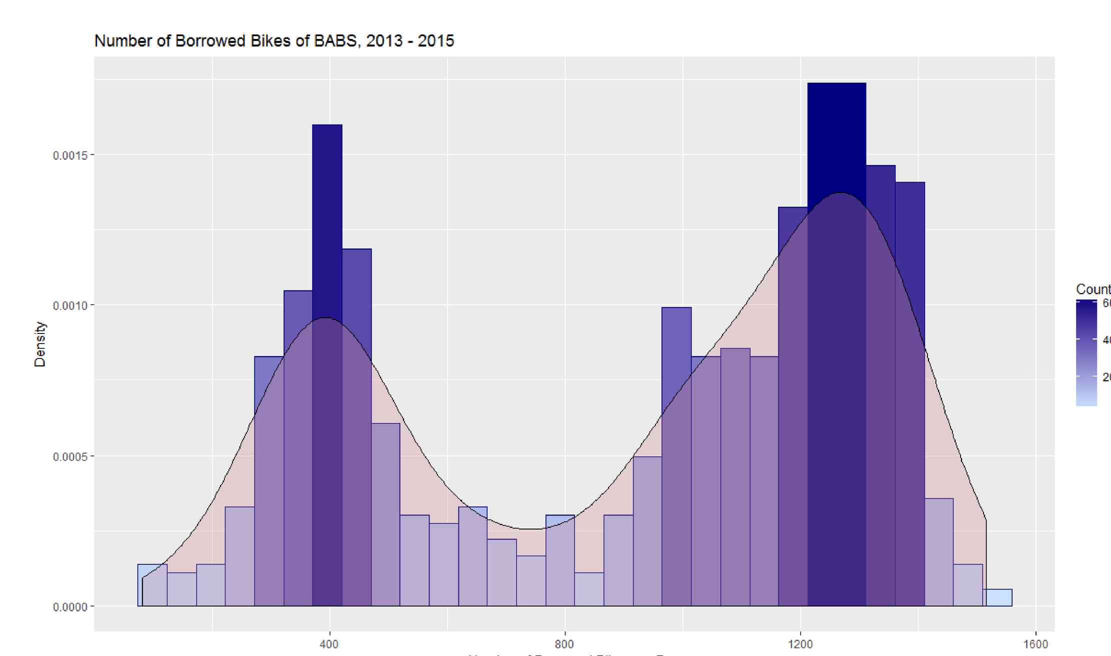

Figure 3: Number of Borrowed Bikes in BABS, 2013-2015

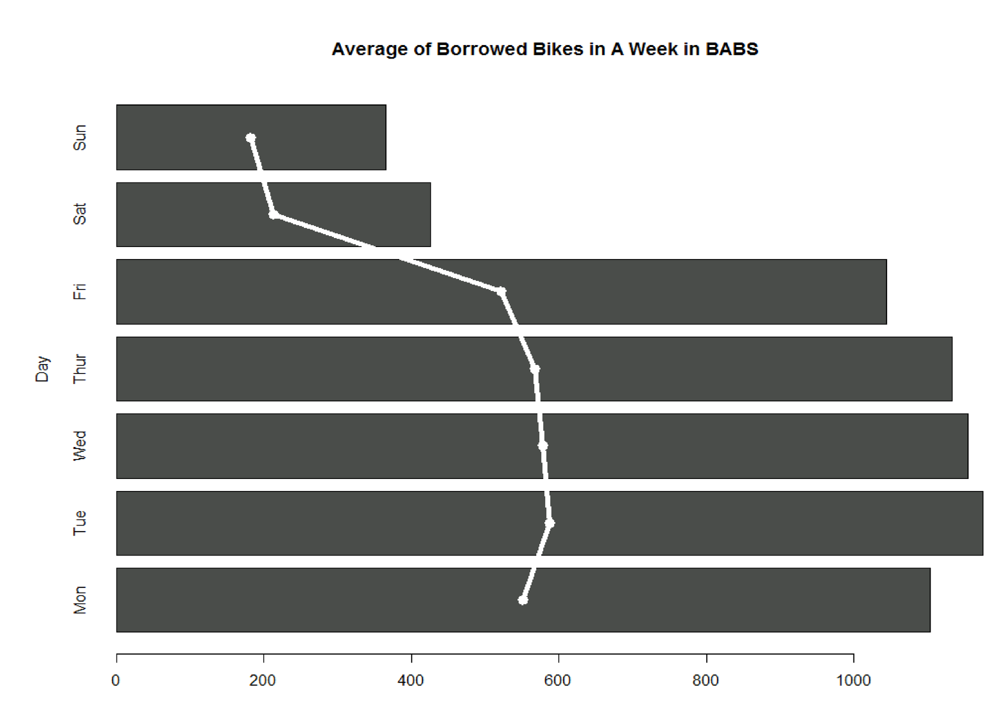

Figure 4: Average of Number of Borrowed Bikes in Week

Results

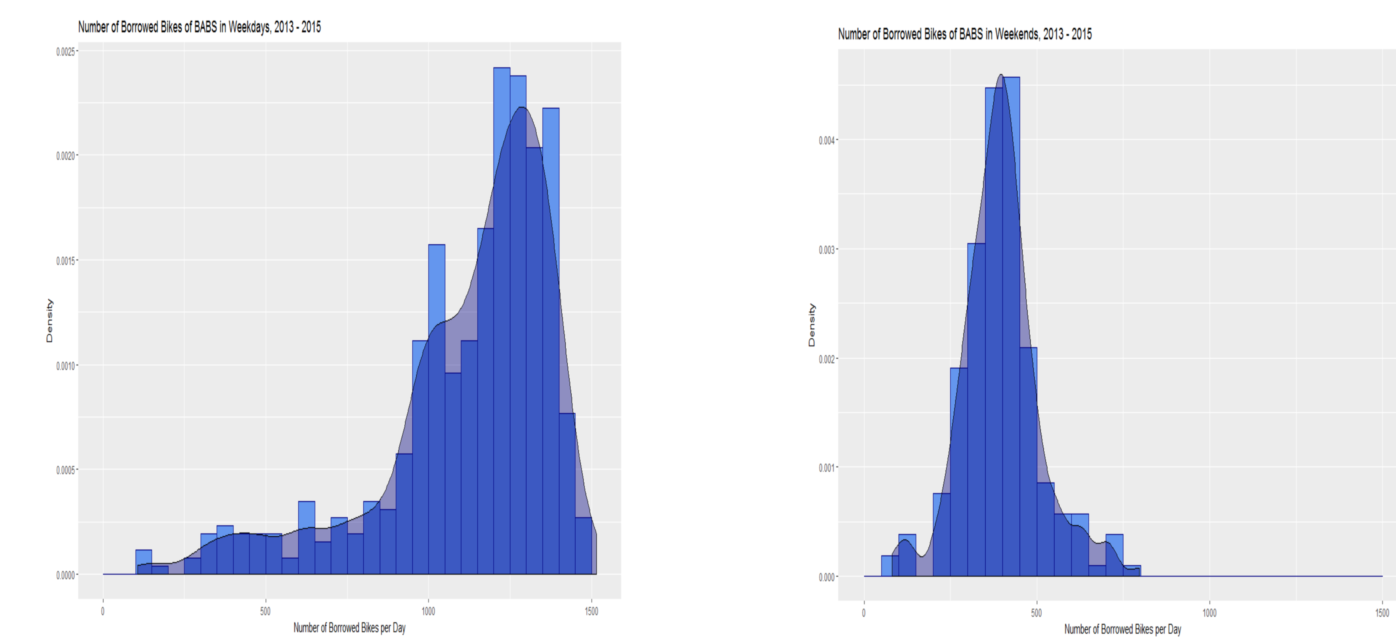

Figure 5: Number of Borrowed Bikes in Weekdays and in Weekends, 2013-2015

Model 1: glm for borrowed bikes in weekdays			
	Estimate	p-value	Standard Error
Temperature	0.017	<0.05	0.00023
Humidity	0.002	<0.05	0.00019
Wind Speed	-0.009	<0.05	0.00041
Visibility	0.028	<0.05	0.00173
Cloud Cover	0.008	<0.05	0.00081
Foggy Days	0.027	<0.05	0.00622
Rainy Days	-0.156	<0.05	0.00463
Fog and Rain	-0.213	<0.05	0.01330
Thunderstorm	0.012	0.672	0.02785

Table 1: Model Summary for Borrowed Bikes in Weekdays

Model 2: glm for borrowed bikes in weekends			
	Estimate	p-value	Standard Error
Temperature	0.019	<0.05	0.00064
Humidity	0.019	<0.05	0.00052
Wind Speed	0.002	<0.05	0.00123
Visibility	0.014	<0.05	0.00476
Cloud Cover	-0.046	<0.05	0.00225
Foggy Days	0.150	<0.05	0.01396
Rainy Days	-0.184	<0.05	0.01458
Fog and Rain	0.042	0.296	0.04067
Thunderstorm	-0.027	0.633	0.05557

Table 2: Model Summary for Borrowed Bikes in Weekends

Main Findings

- The number of borrowed bikes of BABS in San Francisco distributed into two groups. One is for weekdays and the other is for weekend.
- There is statistically significant difference of average number of borrowed bikes between weekdays and weekends. In weekdays, the average of borrowed bikes is around 1122, and for weekends, the average of borrowed bikes is around 396.
- Temperature, humidity, wind speed, visibility and cloud cover are significantly influenced the number of borrowed bikes in both weekdays and weekends.
- The majority of borrower for BABS is the resident in San Francisco.
- BABS may adjust their marketing strategy. For example: the rental bikes in weekends will be impacted more by weather conditions, the company could offer different rental policy for weekend and weekdays.

Reference

"Data." *Plan B Updates - 113: Dozens of U.S. Cities Board the Bike-Sharing Bandwagon | EPI*. N.p., n.d. Web. 15 Apr. 2017.

"Ordinary Least-Squares Regression." *The Multivariate Social Scientist* (n.d.): 56-113. Web.

"Probability Models for Count Data." *Econometric Analysis of Count Data* (n.d.): 7-62. Web.