[Injung Kim, Konstantinos Pelechrinis, and Adam J. Lee. 2020. The Anatomy of the Daily Usage of Bike Sharing Systems: Elevation, Distance and Seasonality. In ACM SIGKDD urbcomp 2020, August 24, 2020, San Diego, CA. ACM, New York, NY, USA.](https://people.cs.pitt.edu/~injungkim/data/Urbcomp20_20.pdf)

In the introduction to their study, the authors explore the vital role of bike-sharing systems in urban settings, particularly for reducing traffic and pollution. They aim to uncover how factors like elevation, the distance of routes, and seasonal shifts impact the usage of these systems. This investigation addresses a notable gap in existing research, emphasizing the need for a deeper understanding of these influences to optimize bike-sharing infrastructure. The study focuses on Pittsburgh's Healthy Ride, employing a sophisticated Poisson regression model to analyze usage patterns. By examining variables such as elevation differences and seasonal variations, the research seeks to offer insights that could significantly improve the efficiency and appeal of bike-sharing systems. This study is about making bike-sharing a more accessible and attractive transportation option, contributing to more sustainable and livable cities.

This paper addresses several key questions related to the dynamics of bike-sharing system usage, focusing on how various factors influence user behavior. The study digs into questions such as: How do elevation differences between bike-sharing stations affect the likelihood of users choosing to bike? What impact does the distance between stations have on the frequency of bike-sharing trips? And how do seasonal changes, with their associated weather conditions, alter the usage patterns of bike-sharing systems? Through a Poisson regression model, the paper reveals insights into these crucial aspects, aiming to enhance the efficiency of bike-sharing services in urban environments.

The authors use a Poisson regression model where the dependent variable 𝑌𝑖 𝑗 is the number of daily trips from station i to station j. Poisson is suitable as 𝑌𝑖 𝑗 is a non-negative integer count.

The model estimates the average trip rate as: 𝜆𝑌 =𝑒𝛼+(b·X) , where X are the independent variables and α, β are parameters estimated via maximum likelihood estimation. Given the predicted 𝜆𝑌, the probability of observing k trips is:

𝑝(𝑌𝑖𝑗 =𝑘|X,b,𝛼) = (𝑒𝑘·(𝛼+(b·X)) /k!). 𝑒-eα+(b.X) .

Authors explored 10 different feature sets (FS1-FS10) for feature selection on the training set (7:2:1 train-validation-test split). Features include day of week dummies, season, precipitation, high/low temp, station capacity, and crucially altitude difference and distance. Altitude difference uses either continuous values or binned ranges (-300 to 300 ft in 100ft increments). Distance is also binned (0-2mi, 2-4mi, etc). Interaction terms between distance and altitude are explored. FS9 performs best with binned altitude difference, training separate summer/winter models, and including distance\*altitude range interaction. Model results show consistent signs but differing sizes of effects across seasons. Precipitation is negative. Altitude difference has largest positive effect when close to 0, reducing for downhill trips and negative for uphill. Distance effect is negative, moderated by altitude difference. On test set, RMSE is 0.16 overall but lower 0.0881 in winter due to lower variability. Key advantage is getting full Poisson trip distribution from mean 𝜆𝑌 (predicted trips).

Elevation differences have a significant negative impact on bike-sharing usage, with users less inclined to use the service for routes that involve steep inclines. Longer distances between stations tend to decrease the number of bike-sharing trips, indicating a preference for shorter, more convenient rides. Seasonal variations strongly affect usage patterns, with noticeable declines in colder or more inclement weather conditions.

**Pros:**

1) First study to explicitly model the relative altitude difference between origin-destination stations when predicting bike trips, an important factor.

2) Develops predictive models rather than just descriptive analysis, enabling planning and "what-if" analyses.

3) Uses Poisson regression which is well-suited for count data like trip numbers and provides the full trip distribution probabilities.

4) Thorough feature selection process exploring different variable representations and interactions.

5) They systematically tested different variable forms (continuous vs binned) and interactions like altitude\*distance to improve predictions.

**Cons:**

1) Analysis limited to only a single city (Pittsburgh). Generalizability to other cities needs to be examined.

2) Model feature set could be expanded further, e.g. job density, population, public transit access.

3) Only uses route distance estimates, not actual slopes/altitudes along the bike paths taken.

4) No direct comparison against other model types like negative binomial regression.

5) Using consecutive months for the train-validation-test split could introduce temporal effects related to seasonality between the sets.