A Concise Survey on Statistical Analysis in Bike Lane Research

Zoe Macris
University of Connecticut

December 2023

Abstract

This survey paper examines the evolving field of statistical analysis of bicycle lanes. Cycling has an important role in society, and has taken a place as a sustainable and accessible mode of transportation and exercise. In order for cycling to be accessible for the most number of people, investments into bike lanes and cycling infrastructure have to be made. Through an examination of existing studies, we explore the current trends in method to show whether bike lanes have a statistically significant effect on cycling outcomes.

Keywords— Bike Lanes, Regression Analysis, Binary Regression Models, Multiple Level Regression Analysis, Zero-Inflated Poisson, Generalized Linear Model

1 Introduction

Given that they have been around for just over 200 years, the bicycle has evolved considerably in it's utility to us as a society. Originally just for hobbyists, and inhibitively expensive, they are now one of the most accessible modes of transport [1]. The World Health Organizations recommendation is a minimum of 150 minutes of moderate aerobic activity per week [18]. This is difficult to achieve in our sedentary society, but cycling is an excellent form of exercise that can also double as the mode of transport for commuting and completing daily tasks.

Statistical research and analysis on bike lanes provides crucial insights into the usage patterns, attitudes, commute rates, bicycle traffic and safety. Other survey papers on research into bike lanes have neglected to focus on the statistical analyses used in the research [13]. This survey paper aims to provide an comprehensive examination of the existing body of statistical research into bike lines. With this, we hope to show trends in the research, give an overview of statistical findings and identify areas that can be further developed.

The rest of the paper will be organized in the following manner. Section 2 will present a background about bike lanes and why they are important. Section 3 will present how the survey was conducted. Section 5 will discuss the findings and recommendations. Then 6 will conclude the paper.

2 Background

In order to understand the research into bike lanes, these topics should be understood.

2.1 Benefits of Cycling

According to Götschi et al. [8], getting the appropriate amount of exercise has an estimated risk reduction of 30% for all-cause mortality, including from cardiovascular disease, coronary heart disease, stroke, diabetes and cancer . It also can help people's mental health as well as physical, some studies have even demonstrated a decrease in depression symptoms from cycling for transportation [9].

Beyond these individual benefits, cycling has a benefit for the environment and society as a whole. Transportation is a large source of greenhouse gas emissions for many countries [9]. In recent years, there has been even more emphasis on sustainable and environmentally conscious transportation options in urban areas. Cycling fits the these criteria and has risen to prominence as a viable sustainable mode of transportation. It is a form of active transportation, which is defined as human powered, and thus does not release greenhouse gas emissions. Bikes also take up considerable less space and require significantly less infrastructure than cars as an added benefit.

In order to promote cycling and make use of the benefits of alleviating traffic congestion, reduced emissions, and enhancing public health, it is necessary to plan and evaluate cycling infrastructure, and bike lanes more specifically. Statistical analysis can come into play here and help prove whether bike lanes make a significant impact on the propensity of people cycling and their safety while doing it.

2.2 Types of Infrastructure

The Connecticut State government describes the common facility types used to accommodate cyclists as Shared Roadways, Wide Curb-lanes, Bicycle Lanes and Multi-use Paths.

Shared roadways are where bicycle and automobile travel are allowed, all streets and roads other than controlled highways, are in this category. They do not provide any added protection to cyclists, but a street's designation as shared roadway indicates to cyclists that there are advantages to using the route over alternatives [6].

Wide curb lanes are still for sharing between bicycle and motorized traffic, but they provide more distance between bicyclists and vehicles and protection from high traffic speeds. They are created by widening roadways or narrowing traffic lanes, or both [6].

The category of bicycle lanes is defined as painted and signed lanes on the shoulder of streets, to improve conditions for bicyclists and to protect them from traffic volume and speed. Being designated as a bike lane street, means that more "pavements surface improvements, stronger sweeping programs, special signal facilities, etc" occur on them to increase safety and comfort for cyclists [6].

The last level of cycling infrastructure is multi-use paths, on which cyclists have exclusive rights-of-way and don't cross The flow of automobiles more than necessary. These can be both recreational opportunities or high=speed commuter routes for cyclists [6].

The research in this survey focuses on the infrastructure of bicycle lanes, and whether they are sufficient to increase the propensities of people to cycle, and their safety while they are cycling.

3 Method and Data

To conduct this comprehensive survey of statistical analyses on bike lanes, a search was conducted of academic databases such as Science Direct, ProQuest, MDPI, and PubMed. These searches were conducted through the University of Connecticut library of Databases to ensure they were papers that were a able to be accessed. Specific terms were used including "bike lane", "statistical analysis", and "cycling infrastructure" in order to identify relevant papers. The date range for literature was defined as 2012 to the present, upon evaluation of typical date range of the available literature.

3.1 Criteria of Inclusion and Exclusion

Literature was included in the survey if it met these criteria:

1. Whether the paper held sufficient statistical analysis on the topic of bike lanes.
2. Being publicized in a peer-reviewed, reputable journal.
3. Availability of paper for access using University of Connecticut student status.

3.2 Data

The data in this survey paper is the types of statistical analyses conducted by the identified literature from method given above. A selection of eleven papers resulted from this, and the approach used to identify relevant data from the papers to be discussed in the survey paper was done as follows. The specific type of statistical analysis conducted in the paper was identified. Alongside the method and variables used by the researchers in the study. The results from the data collection are displayed in 4.

4 Results

4.1 Correlation and Regression Analysis

The correlation and regression analysis approach of statistical research on bicycle lanes is used to establish significant relationships amongst variables. In Mateo-Babiano et al. [12], the approach using Spearman's correlation coefficient rho was used, which was suitable because the variables deviated from a normal distribution. The variables used were Public bicycle-sharing programs (PBSP) station usage frequency and length of bike-ways, which do not follow a normal distribution. The results of the Spearman correlation of station usage and infrastructure type can be seen in 4.1, where shred pathways, bicycle paths, connections, and separated pathways were shown to have a significant impact on PBSP usage.

Other analyses used include Differences-in-Differences, Pearson's R correlation coefficient, Panel regressions, and Counterfactual based on control [11]. These were used when the data being analyzed was usership of Boston's bike share program before and after the addition of bike lanes.

Table 1: Length of bikeway by bikeway type in Brisbane and CityCycle areas. Spearman correlation of station usage and infrastructure type: from Mateo-Babiano paper

Infrastructure type	Length Brisbane Area (km)	Lenth in CityCycle area (km)	Correlation coefficent	P-value
Shared Pathway	327.9	18.0	0.42	¡0.01
Bicycle Lane	186.7	21.3	0.03	0.69
Bicycle Awareness Zone (BAZ)	303.7	40.3	-0.07	0.39
Bicycle Path	23.9	2.6	0.27	¡0.01
Bicycle Route (BAZ)	76.9	8.2	0.13	0.11
Connect	19.5	5.0	0.16	0.049
Informal Off Road	65.8	4.6	0.00	0.98
Informal On Road	18.2	4.1	-0.13	0.12
Separated Pathway	1.8	1.3	0.29	¡0.0

4.2 Binary Regression Models

In the paper by Park and Akar [15], binary probit regression analysis was used to analyze the choice of cycling for commuting, since the dependent variable to choice to bicycle to commute has two outcomes:

- 0, the individual does not commute by cycling.
- 1, the individual commuted by cycling at least once per week.

For the binary model, a latent or unobserved variable y* is assumed that ranges from ∞ to $-\infty$, which is related to the independent variable by the equation $y_i^* = x_i\beta + \epsilon_i$. In the context of this survey paper, this latent variable are what underlies the decision making process of bicycle commuting. Where X_i is a matrix of independent variables and β is a vector of coefficients to be estimated and ϵ is random error [10].

The binary logistic regression model is also used in this field of statistical research to predict the likelihood of a successful event occurring, like a high number of cycling routes close to green areas [4]. In order to conduct this model, all the variables had to be recoded to dichotomous categorical variables (1-0). For example, bicycle racks where 1=bicycle racks in buffer area, 0= no bicycle racks.

4.3 Multiple Level Regression Analysis

The paper from Teixeira et al. [17] uses Multiple Level Logistic Regression to analyze the impact of cycling in various cities with different cycling infrastructure on stress markers. This analysis resulted in the data shown in Table 1.

Two types of models were used in Buehler and Pucher [3], a log-log Ordinary Least Square (OLS) model "with the dependent variable being the bike commuters per 10,000 population. As well as a Binary Logit Proportional model "with the share of bike commuters in each city as the dependent variable". This paper analyzed the role of cycling infrastructure on commutership by bike in 90 of the largest 100 U.S. cities. The OLS model is used for estimating coefficients of linear

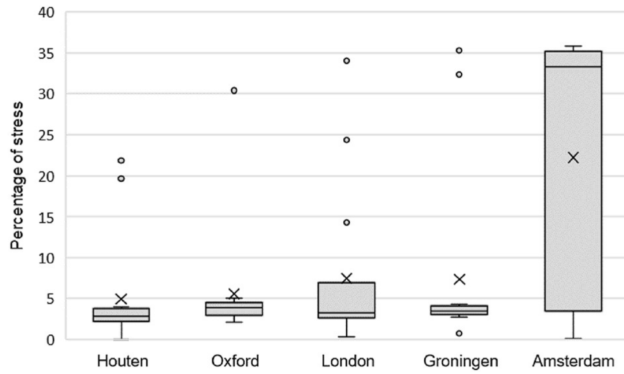

Figure 1: Distribution of Stress by City from Teixeira Study

regression that describe the relationship between one or more quantitative independent variables and a dependent variable [14]. With p explanatory variables, the OLS regression model is:

$$Y = \beta_0 + \sum_{j=1,\ldots,p} \beta_j X_j + \epsilon$$

In the group of papers selected, the multinomial logit (MNL) model was used to estimate the travel choice of bicycle for commuting [19].

4.4 Zero-inflated Poisson Regression Mode

A different approach that was used in the research into bike lanes was to use a zero-inflated Poisson regression model to estimate the change in cycle-motor vehicle collisions [2]. This model is used since the design is looking at collision rates pre to post instillation of bike lanes, and since the model allows for an over-abundance of zero counts in the data [7]. The Poisson distribution's p.m.f is:

$$P(Y = y) = \exp(-\lambda)\lambda^y / y!$$

The Poisson model is also used in Cantisani et al. [5] to "address the randomness due to the perception and decision of road users at the intersections", as their research is specifically on the crash likelihood of cyclists at roundabouts with different infrastructure types.

4.5 Generalized Linear Model

The final method identified from the surveyed papers was to use the generalized linear model to predict the vehicle delays caused by bicycle traffic versus other variables. Alongside this method, the cumulative curve method as used to extract traffic flow data from videos [16]. The basic model of the generalized linear model is:

$$Y_{nx1} = X_{nxp}\beta_{px1} + \epsilon_{nx1}$$

Where the matrix of observed values of dependent variables is Y_{nx1}, X_{nxp} is the matrix of observed explanatory variables, β_{px1} = is the matrix of coefficients for explanatory variables, and ϵ_{nx1} is the error. Using this model is reported to help to understand how each factor contributed to the delay of cyclists on urban streets.

Table 2: Modeling results of vehicle delay.

Variable	Coef.	Std. err.	z	$P > \|z\|$	[95% conf. interval]
Bicycle flow (thousand)	32.67	7.67	4.29	<0.001	[17.33, 48.00]
Vehicle flow (thousand)	38.00	8.67	4.35	<0.001	[20.67, 55.33]
Number of vehicle lanes	−30.87	3.03	−10.12	<0.001	[−36.93, −24.80]
Width of bike lane	−5.83	3.10	−1.88	0.061	[−12.03, 0.367]
Constant	80.43	5.07	15.85	<0.001	[70.30, 90.57]

Statistics. log likelihood = −1077.782; Prob = <0.001; AIC = 3.063.

Figure 2: Results table from Research on Bicycle Traffic from Pu et al.

5 Discussion

The findings from this survey paper show the multifaceted approaches being taken in the field of research into cycling infrastructure, and how important they are to encourage sustainable urban transportation. In this discussion section will include; key themes, criticisms, and potential future pathways for similar research.

The different types of cycling infrastructure, especially bike lanes, are shown in this body of research to have a significant impact on increasing cycling ridership. Whether it's an increase in bike-shares or commutership, this relationship is shown in the papers. This is essential when it comes to appealing to policy makers and decision makers in communities to convince them to add cycling infrastructure to their area. Statistical analyses are influential in providing actionable insights, including the correlation between infrastructure types existing in an area and increased cycling.

Studies like Mateo-Babiano et al. [12] show how effective simple methods like Spearman's correlation coefficient can be at revealing the impact of bike lanes on PBSP usage. These sorts of results can lead to insights for urban planners when it comes to designing cities.

Predictive models were also used, like binary regression models in Park and Akar [15] to provide insight into all the factors influencing an individuals' choice to cycle for commuting. These predictive models offer ways for policy makers to encourage cycling as a successful commuting option in their cities.

Some future directions this field could go include focusing on more that just high socioeconomic areas. This type of research has only really been done in affluent countries, likely because a lot of the research has been based off of bike-share systems. Which leads into another point of improvements, where there could be better ways of collecting data on cyclists, as they are often excluding people who own their own bikes from the data. Additionally, there is room for more cost-benefit analyses on the economic impact of investing in cycling infrastructure. When you create such infrastructure, there could be less congestion and damage to the automobile infrastructure. Which would be beneficial to persuading policy makers towards creating more bike lanes and encouraging cycling.

6 Conclusion

In conclusion, this survey of statistical research on bike lanes offers a current snapshot of frequently used methods in the analysis of bike lanes. The significance of bike lanes in increasing cycling highlights the need for continued research and innovative statistical analysis to understand how to overcome the challenges of congestion, safety, and sustainability in transportation. This supports the efforts of enhancing cycling infrastructure for the benefits of individuals experiences while cycling and the betterment of entire communities.

7 Acknowledgements

I would like to express my appreciation to Professor Jun Yan for his guidance in creating this paper.

References

[1] F. J. Berto. bicycle. *Encyclopædia Britannica*, 2023.

[2] D. Bhatia, S. A. Richmond, C. J. Loo, L. Rothman, C. Macarthur, and A. Howard. Examining the impact of cycle lanes on cyclist-motor vehicle collisions in the city of toronto. *Journal of Transport &amp Health*, 3(4):523–528, Dec. 2016. doi: 10.1016/j.jth.2016.04.002.

[3] R. Buehler and J. Pucher. Cycling to work in 90 large american cities: new evidence on the role of bike paths and lanes. *Transportation*, 39(2):409–432, July 2011. doi: 10.1007/s11116-011-9355-8. URL https://doi.org/10.1007/s11116-011-9355-8.

[4] F. S. Campos-Sánchez, L. M. Valenzuela-Montes, and F. J. Abarca-Álvarez. Evidence of green areas, cycle infrastructure and attractive destinations working together in development on urban cycling. *Sustainability*, 11(17):4730, Aug. 2019. doi: 10.3390/su11174730. URL https://doi.org/10.3390/su11174730.

[5] G. Cantisani, C. Durastanti, and L. Moretti. Cyclists at roundabouts: Risk analysis and rational criteria for choosing safer layouts. *Infrastructures*, 6(3):34, Mar. 2021. doi: 10.3390/infrastructures6030034.

[6] CTDOT. *Connecticut Statewide Bicycle and Pedestrian Transportation Plan*, chapter 4. CTDOT, 2023.

[7] D. Giles. Notes on the zero-inflated poisson regression model. Department of Economics, University of Victoria, 2010.

[8] T. Götschi, J. Garrard, and B. Giles-Corti. Cycling as a part of daily life: A review of health perspectives. *Transport Reviews*, 36(1):45–71, June 2015. doi: 10.1080/01441647.2015.1057877. URL https://doi.org/10.1080/01441647.2015.1057877.

[9] S. Green, P. Sakuls, and S. Levitt. Cycling for health. *Canadian Family Physician*, 67(10):739–742, Oct. 2021. doi: 10.46747/cfp.6710739. URL https://doi.org/10.46747/cfp.6710739.

[10] J. F. J. Scott Long. *Regression Models for Categorical Dependent Variabls Using STATA*. Stata Press, 2001.

[11] E. Karpinski. Estimating the effect of protected bike lanes on bike-share ridership in boston: A case study on commonwealth avenue. *Case Studies on Transport Policy*, 9(3):1313–1323, Sept. 2021. doi: 10.1016/j.cstp.2021.06.015. URL https://doi.org/10.1016/j.cstp.2021.06.015.

[12] I. Mateo-Babiano, R. Bean, J. Corcoran, and D. Pojani. How does our natural and built environment affect the use of bicycle sharing? *Transportation Research Part A: Policy and Practice*, 94:295–307, Dec. 2016. doi: 10.1016/j.tra.2016.09.015. URL https://doi.org/10.1016/j.tra.2016.09.015.

[13] F. J. M. Mölenberg, J. Panter, A. Burdorf, and F. J. van Lenthe. A systematic review of the effect of infrastructural interventions to promote cycling: strengthening causal inference from observational data. *International Journal of Behavioral Nutrition and Physical Activity*, 16(1), Oct. 2019. doi: 10.1186/s12966-019-0850-1. URL https://doi.org/10.1186/s12966-019-0850-1.

[14] ORDINARY LEAST SQUARES REGRESSION (OLS). URL https://www.xlstat.com/en/solutions/features/ordinary-least-squares-regression-ols. Website from XLSTAT on OLS.

[15] Y. Park and G. Akar. Understanding the effects of individual attitudes, perceptions, and residential neighborhood types on university commuters' bicycling decisions. *Journal of Transport and Land Use*, 12(1):419–441, 2019. ISSN 19387849.

[16] Z. Pu, Z. Li, Y. Wang, M. Ye, and W. D. Fan. Evaluating the interference of bicycle traffic on vehicle operation on urban streets with bike lanes. *Journal of Advanced Transportation*, 2017:1–9, 2017. doi: 10.1155/2017/6973089. URL https://doi.org/10.1155/2017/6973089.

[17] I. P. Teixeira, A. N. R. da Silva, T. Schwanen, G. G. Manzato, L. Dörrzapf, P. Zeile, L. Dekoninck, and D. Botteldooren. Does cycling infrastructure reduce stress biomarkers in commuting cyclists? a comparison of five european cities. *Journal of Transport Geography*, 88:102830, Oct. 2020. doi: 10.1016/j.jtrangeo.2020.102830. URL https://doi.org/10.1016/j.jtrangeo.2020.102830.

[18] WHO. *WHO guidelines on physical activity and sedentary behaviour: at a glance*. World Health Organization, Genève, Switzerland, Nov. 2020.

[19] P. Zhao. The impact of the built environment on bicycle commuting: Evidence from beijing. *Urban Studies*, 51(5):1019–1037, July 2013. doi: 10.1177/0042098013494423. URL https://doi.org/10.1177/0042098013494423.