A Concise Survey on Statistical Analysis Methods 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 plays 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 been made. Through an examination of existing studies, we explore the current trends in statistical methods to show whether certain types of infrastructure have a significant effect on cycling outcomes.

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

1 Introduction

Given that bikes have been around for just over 200 years, it is unsurprising that they have evolved considerably in their utility to us as a society in that time. Originally just for hobbyists, and inhibitively expensive, they are now one of the most accessible modes of transport [1]. Biking is an exceptional form of exercise and is a great way to achieve the World Health Organizations recommendation of a minimum of 150 minutes of moderate aerobic activity per week [25]. This is difficult to achieve in our sedentary society, but doable with cycling since it is a 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 [15]. 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.

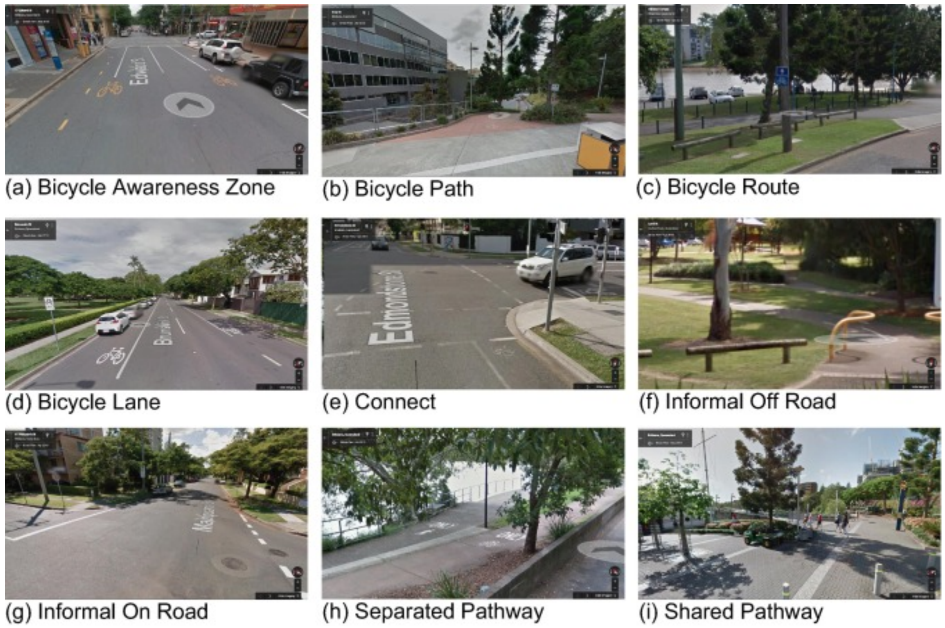
2 Background

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

2.1 Benefits of Cycling

According to Götschi et al. [9], 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 [10].

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 [10]. In recent years, there has been even more emphasis on sustainable and environmentally conscious transportation options in urban areas. Cycling fits these criteria and has risen to prominence as a



(a) Bicycle Awareness Zone (b) Bicycle Path (c) Bicycle Route
(d) Bicycle Lane (e) Connect (f) Informal Off Road
(g) Informal On Road (h) Separated Pathway (i) Shared Pathway

Figure 1: Bike Lane Comparison [14]

sustainable mode of transportation. It is a form of active transportation, which is term 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. Especially different forms of bike lanes, as this is commonly the only cycling infrastructure in the U.S. Statistical analysis comes into play here and can 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 (Informal On Road), Wide Curb-lanes, Bicycle Lanes and Multi-use Paths. The visuals in Figure 1, show an even wider array of potential infrastructures. In the U.S., commonly seen types are Bicycle Awareness Zones, Informal On Road and Bicycle Lanes. The supposition of much of the research analyzed here is that the more advanced the infrastructure is, the more likely people are to cycle [14].

Shared roadways are where bicycle and automobile traffic 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 [7].

Wide curb lanes are still require 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 [7].

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 [7].

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 [7].

The bulk of 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 the academic databases Science Direct, ProQuest, MDPI, and PubMed. These searches were conducted through the University of Connecticut library of Databases to ensure that the full content of the papers could be viewed. 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 the 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 discussed in Section 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. [14], 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 Table 4.1, where shared 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 [13].

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

These were used when the data being analyzed was user-ship of Boston's bike share program before and after the addition of bike lanes.

While a fairly simple technique, basic regression and correlation analyses are a fundamental statistical tool offering multiple advantages in data exploration and analysis. These techniques allow for the understanding and quantification of relationships between variables. Regression analysis not only reveals the nature and strength of connections between dependent and independent variables, but also aids in prediction and model interpretation. It serves as a basis for decision-making by allowing insights into influential data points, model comparison, and assumption checking. Similarly, correlation analysis succinctly measures the degree and direction of association between two variables, which gives a simple but effective way to summarizing their relationship [2]. All in all, these basic analyses provide crucial foundations for further statistical modeling, making them indispensable in telling what variable affecting cycling will provide the greatest increase in bike share usage. Even though it is a simple technique, we are able to tell in Mateo-Babiano et al. [14] that there is a significant relationship between shared pathways, bicycle paths and separated pathways on the amount of bike share station usage. Which provides a solid foundation of evidence on how to improve cycling infrastructure in urban areas.

4.2 Binary Regression Models

In the paper by Park and Akar [17], 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 [11].

The binary logistic regression model is also used in the field of statistical research on bike lanes to predict the likelihood of a successful event occurring, like a high number of cycling routes close to green areas [5]. Binary logistic regression serves as an indispensable tool in predictive modeling due to its versatile applications and tailored approach to analyzing binary dependent variables. Unlike linear regression, which is suitable for predicting continuous values, binary logistic regression is specifically designed for situations where the outcome is dichotomous, such as survival vs. death, presence vs. absence, or default vs. non-default in loan scenarios. In the case of Campos-Sánchez et al. [5], multiple variables are coded 1-0 in a binary fashion and can be seen in Figure 2.

One of the prime reasons for favoring binary logistic regression over linear regression for binary outcomes is the inherent nature of the dependent variable. In logistic regression, the model predicts the probability that the dependent variable will assume a value of 1. This probability is modeled using a logit function, ensuring the estimation of the relationship between independent variables and the probability of the event occurring.

The logistic regression model itself relies on statistical estimation methods, primarily the maximum likelihood estimation. This method aids in determining the coefficients of the model, such as 'B0' to 'Bk', which correspond to the influence of each independent variable on the log odds of the binary outcome. These coefficients are estimated using iterative algorithms like Fisher scoring or

Factors	Variables (Unit)	References	Measurements
Cyclist's behaviour	1. Flow (n)	[11,12,22]	nº of cycling routes (GPS) passing through the green area plus a 100 m buffer zone surrounding it.
Green area	2. Size (ha)	[2,23,24,25,26]	Surface area of the green area.
Environment	3. Residents (n)		nº of inhabitants within a 1 km buffer around thegreen area.
Environment	4. Commerce (m2)	[21,26,27]	Built area for commercial use within a 1 km buffer around the green area.
Network	5. Global intermediation (n)	[27,28]	Intermediation (i.e., the potential for intersections) of the access street to each green area at the city level.
Network	6. Global integration (n)	[27,28]	Integration (i.e., potential for a destination) of the access street to each green area at city level.
Green area	7. Altitude (n)	[5,21,29]	Variability in altitude in the green areas.
Cyclist's behaviour	8. Length (km)	[2,12,25,30]	Average length of the routes passing through each green area +100 m buffer.
Cyclist's behaviour	9. Time (h)	[2,12,25,30]	Average time for the routes passing through each green area +100 m buffer.
Cyclist's behaviour	10. Speed (km/h)	[2,12,25,30]	Average speed of the routes passing through each green area +100 m buffer.
Traffic	11. Crossroads (n)	[11,31,32]	nº of crossroads ≥4 roads within a 1 km buffer around the green area.
Traffic	12. Parking (n)	[1,12,21,33,34]	nº of public car parks within a 1 km buffer around the green area.
Cycling infrastructure	13. Bicycle racks (n)	[21,35,36]	Bicycle racks in each green area +100 m buffer.
Cycling infrastructure	14. Bicycle lanes (m)	[2,3,23,36]	Distance from the edge of the park to the nearest bicycle lane (or bus-taxi lane for use by cyclists).
Environment	15. University centres (n)	[21,26,27]	nº of university buildings (faculties, centres, rectorate, etc.) within a 1 km buffer around each green area.

Figure 2: Variable Table for Binary Logistic Regression Model

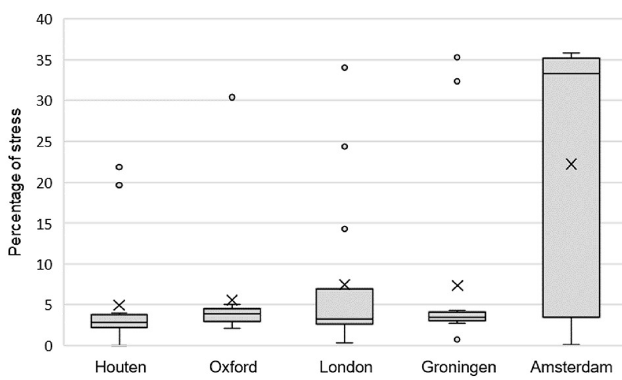


Figure 3: Distribution of Stress by City from Teixeira Study

the Newton-Raphson method, allowing for practical implementation using programming languages like R or Python [19].

4.3 Multiple Level Regression Analysis

Multiple Level Logistic Regression is used in the paper from Teixeira et al. [24] to analyze the impact of cycling infrastructure in different cities on stress markers. This analysis resulted in the data shown in Figure 3, of stress levels measured in different cities. The results of this research found that riding on a protected bike lane significantly reduces cyclists stress levels compared to cycling on a general use street.

Two types of models were used in Buehler and Pucher [4], 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 from each city being the dependent variable. This paper analyzed the role of cycling infrastructure on commuter-ship by bike in 90 of the largest 100 U.S. cities. The OLS model is used for estimating coefficients of linear regression that describe the relationship between one or more quantitative independent variables and a dependent variable [16]. With p explanatory variables, the OLS regression model is given by equation 1:

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

In the group of papers addressed, the multinomial logit (MNL) model was most commonly used to estimate the travel choice of bicycle for commuting [26]. Multilevel models, unlike traditional regression approaches, offer several crucial advantages. Firstly, they accurately handle hierarchical data structures, avoiding underestimation of standard errors and overstating statistical significance by recognizing and accounting for group-level variability. Secondly, these models address substantive inquiries related to group effects, crucial in scenarios like transportation mode to work choices or identifying 'outlying' groups, providing insights into value-added effects at different levels. Thirdly, they enable simultaneous estimation of group effects and group-level predictors without confounding their impacts, which can't be separated in fixed effects models. Lastly, multilevel models permit inference to a population of groups, treating sampled groups as a random sample from a larger population, unlike fixed effects models, which restrict inferences solely to the sampled groups. These benefits collectively enhance the depth of the analyses, particularly in understanding complex hierarchical data structures and their effects [23].

4.4 Zero-inflated Poisson Regression Model

Another approach used in the research into bike lanes is a zero-inflated Poisson regression model to estimate the change in cycle-motor vehicle collisions [3]. This model was 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 [8]. The model uses the Poisson distribution's p.m.f which is given by equation 2:

$$P(Y = y) = \exp(-\lambda)\lambda^y/y! \tag{2}$$

The Poisson model is also used in Cantisani et al. [6] 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. Zero-inflated models, by assigning a non-zero probability of observing one or more occurrences (like crashes), avoid assuming inherent safety or non-safety in intersections. Due to these reasons, zero-inflated models are useful for analyzing crash count data. It is important to note that zero-inflated models are not always superior for crash counts, but they are a viable option to consider for bike crash research [21].

4.5 Generalized Linear Model

The final method identified from the surveyed papers was the generalized linear model (GLM) being used to predict the vehicle delays caused by bicycle traffic versus other variables. The GLM has the benefit of being able to accommodate various types of response variables, including continuous, binary, count, and categorical data. It allows for different types of probability distributions such as Gaussian, Poisson, Binomial, etc., making it applicable in a wide range of scenarios. Additionally, GLMs are not restricted to normality assumptions, which makes them suitable for data that is distributed non-normally [20].

Alongside this method, the cumulative curve method as used to extract traffic flow data from videos [22]. Equation 3 gives the basic model of the generalized linear model:

$$Y_{nx1} = X_{nxp}\beta_{px1} + \epsilon_{nx1} \tag{3}$$

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 help us understand how each factor contributed to the delay of cyclists on urban streets. In Figure 4, the results are shown of the significance of multiple variables on vehicle

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 4: Results table from Research on Bicycle Traffic from Pu et al. [22]

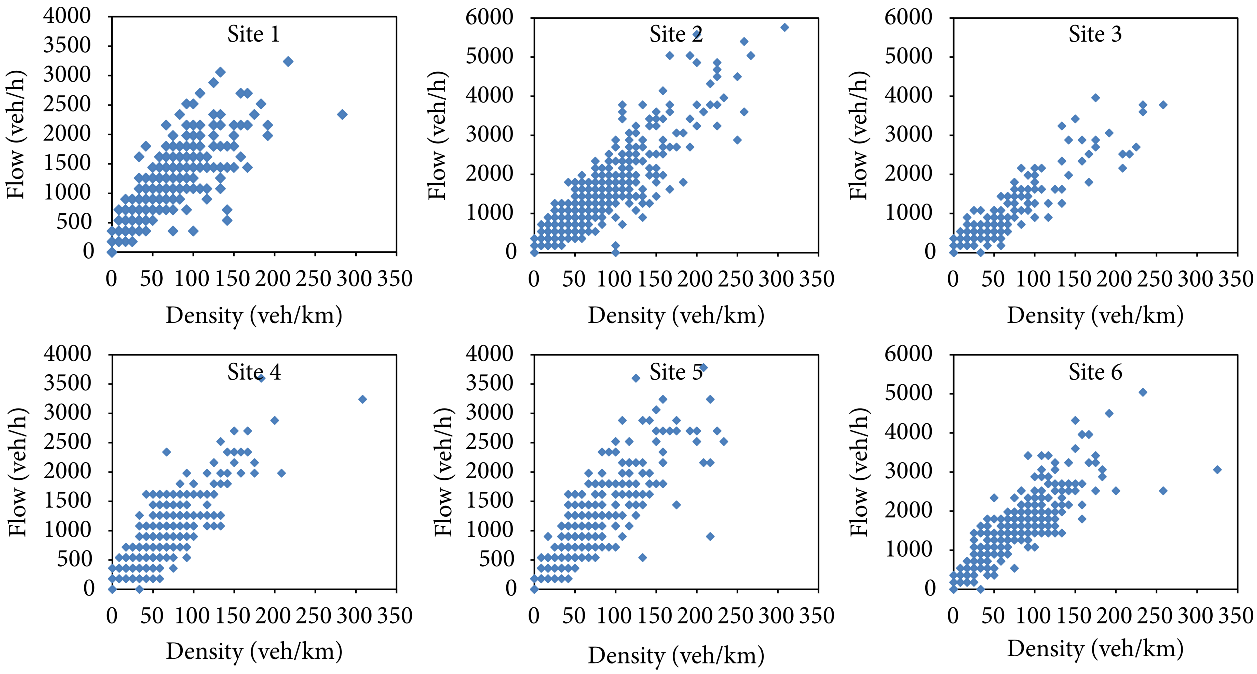


Figure 5: Relationship between bicycle flow and density from Pu et al. [22]

delay. It can be seen that with this method, bicycle and vehicle flow as well as the number of vehicle lanes, have a significant effect on vehicle delay. The results of flow vs. density are similarly shown in Figure 5, and reveals that the average bicycle speed decreases slightly as the density increases while the flow is still increasing, and when there is high density the variation of bicycle speed decreases a great deal. These results are also consistent across multiple locations of urban streets.

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; important themes, criticisms, and potential future pathways for similar research.

Investing in different types of cycling infrastructure, especially separated bike lanes, is shown in this body of research to have a significant impact on increasing cycling ridership. Whether it is an increase in bike-shares or commutership, this relationship is shown in the research. 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. [14] 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 important ideas for urban planners when it comes to designing cities. Predictive models were also used, like binary regression models in Park and Akar [17] 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 improvements this field could make include focusing on more than just high socioeconomic class areas. This type of research has only really been done in affluent countries, specifically in affluent areas, likely because much of the research has been based off of bike-share systems. Which leads into another point of criticism, where there could be better ways of collecting data on cyclists, as they are often excluding people who own and ride 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 well used bike lanes, there would then be less congestion and damage to automobile infrastructure. Focusing on this angle within bike lane research would be beneficial to persuading policy makers towards creating more bike lanes and encouraging cycling.

Further improvements for the future of the field of research into cycling infrastructure include using longitudinal studies to track changes and trends, rather that just cross-sectional research at one point in time. In addition, there is a lack of spatial analysis models in the current body of research. These models, particularly spatial regression, Geographically Weighted Regression (GWR), and spatial auto correlation analysis, would be valuable for understanding spatial patterns. For instance, clustering of bike lane usage, or identifying hot-spots of accidents and injuries along bike lanes [12]. It could also be beneficial to include causal inference techniques when studying cycling infrastructure. These tools, like directed acyclic graphs and instrumental variable analysis, could be appropriate for investigating the causal impact of interventions or policy changes related to bike lane infrastructure on outcomes like safety or usage of the infrastructure [18].

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. When working in this field of study, the choice of the best statistical model should consider the specific research objectives, the characteristics of the available data, and the assumptions of the model. Often, a combination of different models or methodologies may be more informative to comprehensively understand various aspects of bike lane-related phenomena. Consulting with experts in specific techniques and conducting exploratory data analysis can also guide the selection of the most appropriate statistical approach for a given bike lane research study.

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. As shown in this summary paper, such analysis is critical to support 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] V. Bewick, L. Cheek, and J. Ball. Statistics review 7: Correlation and regression. *Critical Care*, 7(6):451, 2003. doi: 10.1186/cc2401.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

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

[9] 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.

[10] 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.

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

[12] V. Kanade. What is spatial analysis? definition and examples, 2022. URL https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-spatial-analysis/.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] N. Pearce and D. A. Lawlor. Causal inference—so much more than statistics. *International Journal of Epidemiology*, 45(6):1895–1903, 2016. doi: 10.1093/ije/dyw328.

[19] P. Penman. Binary logistic regression – an introduction, 2022. URL https://www.datascienceinstitute.net/blog/binary-logistic-regression-an-introduction#:~:text=Binary%20logistic%20regression%20models%20the,or%20presence%20and%20so%20on.

[20] PennState. 6.1 - introduction to glms: Stat 504, 2023. URL https://online.stat.psu.edu/stat504/lesson/6/6.1.

[21] T. Pew, R. L. Warr, G. G. Schultz, and M. Heaton. Justification for considering zero-inflated models in crash frequency analysis. *Transportation Research Interdisciplinary Perspectives*, 8: 100249, 2020. doi: 10.1016/j.trip.2020.100249.

[22] 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.

[23] J. Rasbash, Aug 2023. URL https://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html.

[24] 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.

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

[26] 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.