Ministry of Higher Education
and Scientific Research

University of Tunis

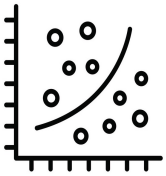
Tunis Business School

وزارة التعليم العالي و البحث العلمي

جامعة تونس

المعهد العالي للأعمال بتونس

Project Report for Econometrics Course



Factors Influencing Traffic Accidents: A Linear Regression Analysis

Elaborated By :
Fatma BEN REJEB

Supervisor :
Prof. Amira DRIDI

Academic year : 2024/2025

Contents

Dataset Source

Google Colab Link

1 Introduction

This report provides an analysis of a dataset containing key factors influencing traffic accidents across urban and rural areas. The dataset includes 8,756 observations, covering both environmental, infrastructural, and behavioral variables. It aims to contribute to improving road safety through insightful analysis that can inform policy, infrastructure planning, and predictive modeling.

2 Dataset Overview

The dataset consists of the following variables:

- **Accidents:** The number of recorded traffic accidents, ranging from minor incidents to major collisions.
- **Traffic Fine Amount:** The average traffic fines (in thousands of USD) in the observed area, reflecting enforcement efforts and driver behavior.
- **Traffic Density:** A score (0-10) indicating vehicle volume, with 0 representing low traffic and 10 indicating high traffic density.
- **Traffic Lights:** The proportion of intersections with traffic lights, serving as a proxy for control at intersections.
- **Pavement Quality:** A rating (0-5) for road conditions, with higher values indicating better infrastructure.
- **Urban or Rural Classification:** A binary indicator (1 for urban, 0 for rural).
- **Average Speed:** The typical speed of vehicles (in km/h) observed in the area.
- **Rain Intensity:** A scale (0-3) reflecting the intensity of rain, where 0 is no rain and 3 is heavy rain.
- **Vehicle Count:** Estimated number of vehicles (in thousands) present in the area.
- **Time of Day:** The time at which the accidents occurred, on a 24-hour scale.

3 Purpose of the Analysis

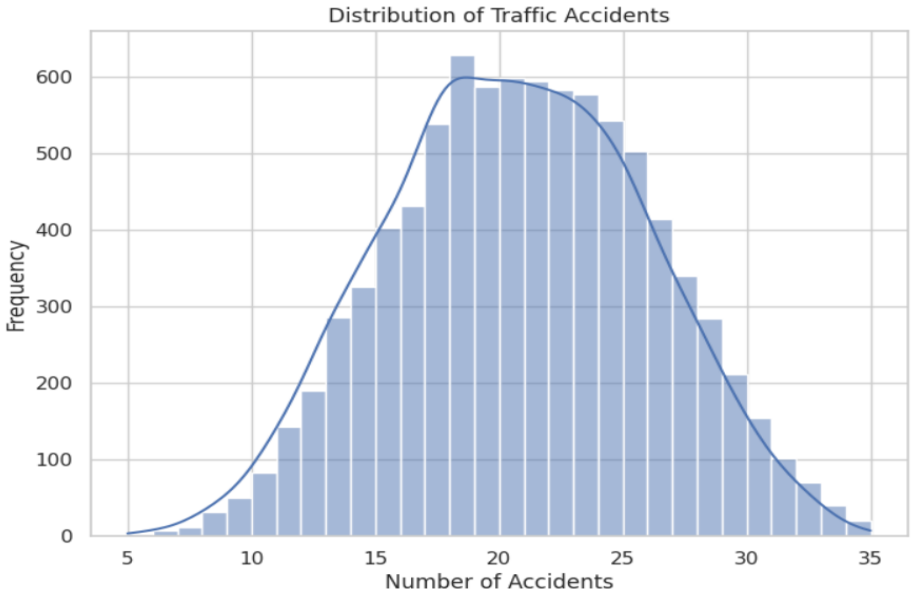
The dataset is designed to support:

- **Traffic Safety Analysis:** Understanding the factors contributing to accidents.
- **Urban Planning and Infrastructure Improvement:** Identifying areas needing better infrastructure or traffic management.
- **Predictive Modeling:** Developing models to predict accident hotspots and identify high-risk conditions.
- **Policy-making:** Informing decisions aimed at reducing traffic-related incidents and enhancing road safety.

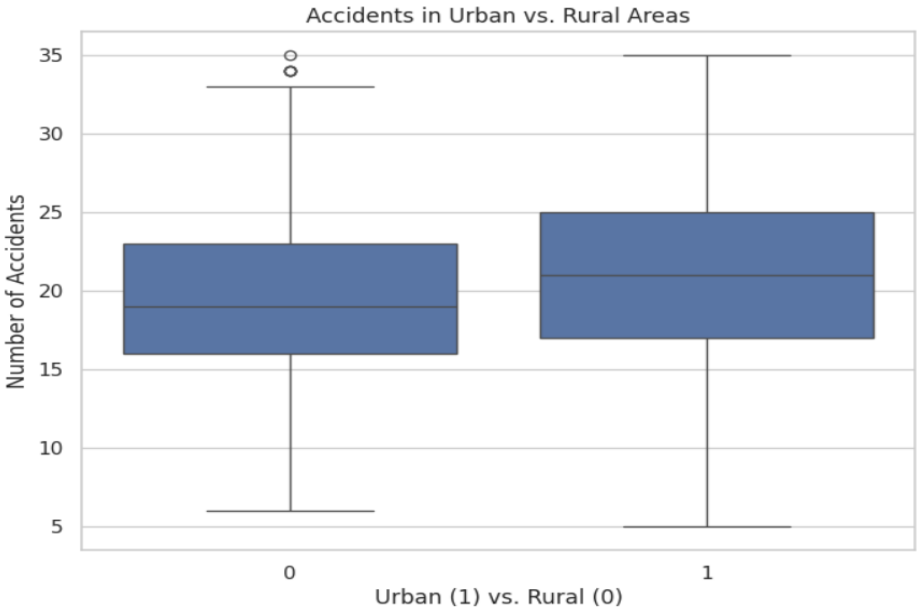
4 Methodology

4.1 Exploratory Data Analysis (EDA)

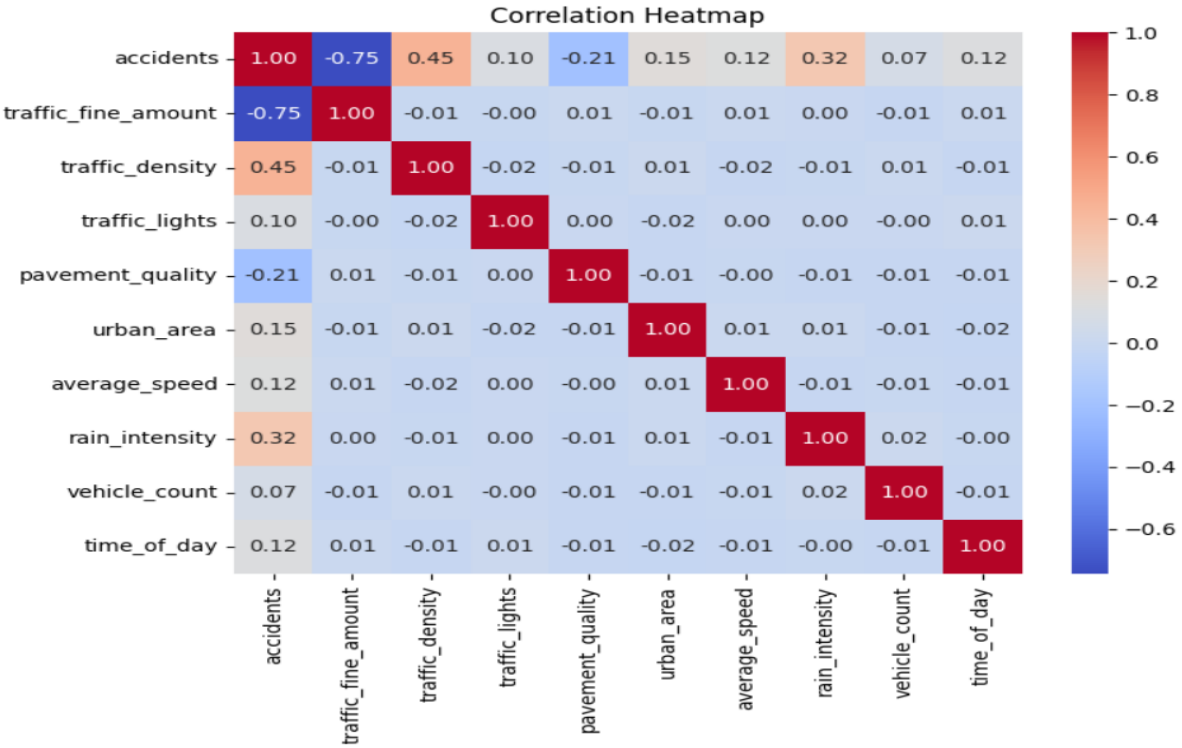
- I began by examining the dataset to gain insights into the distribution of key variables, particularly the dependent variable (accidents). A histogram was plotted to visualize the distribution of accident occurrences, revealing patterns such as skewness or potential outliers.



- Box plot analysis was conducted to compare accident occurrences in urban versus rural settings.



- A correlation heatmap was generated to identify significant relationships between variables.

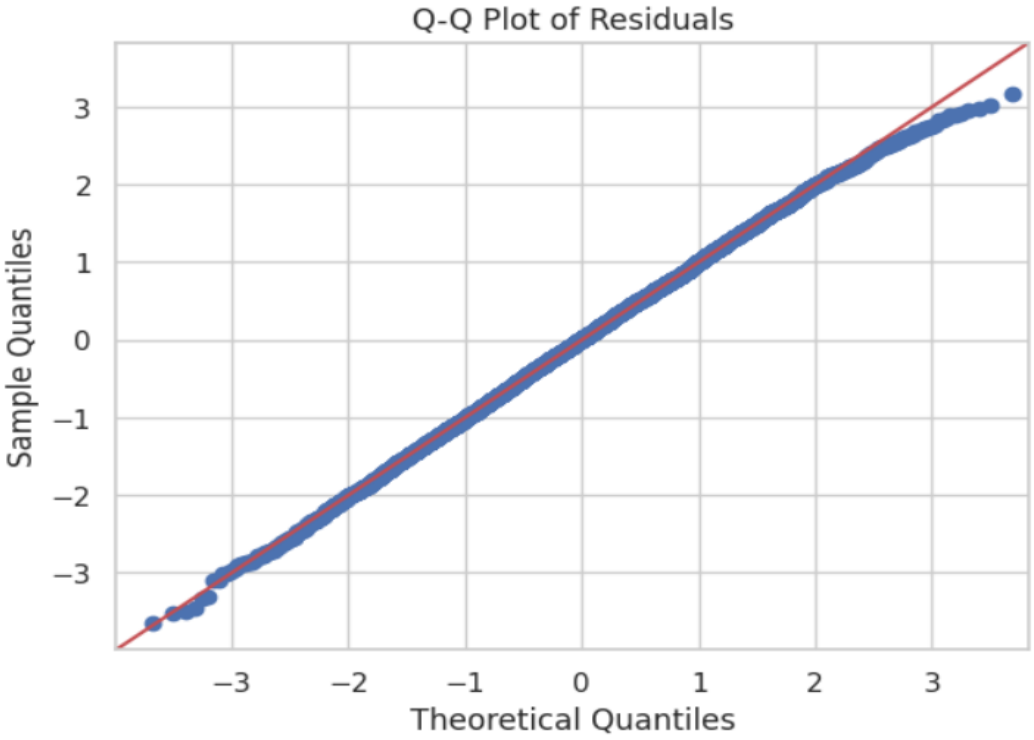


4.2 Feature Selection

I performed a correlation analysis to identify the most relevant variables. Features with a low correlation with the dependent variable were excluded. The final selected features for regression were: `'traffic_density'`, `'rain_intensity'`, and `'traffic_fine_amount'`. For the `'pavement_quality'`, although it showed a weak negative correlation with accidents, it was considered for additional testing to determine if adding it to the model would improve performance.

4.3 Statistical Tests

To ensure the model assumptions were met, I conducted the Durbin-Watson test to detect autocorrelation in the residuals and the Breusch-Pagan test to check for heteroscedasticity. I also examined the residuals through histograms and Q-Q plots for normality. Multicollinearity was assessed using the Variance Inflation Factor (VIF) to ensure the independence of the predictors.



4.4 Model Creation

A linear regression model was initially built using the selected features. Subsequently, I tested if adding `'pavement_quality'` as an additional predictor improved the model's performance.

4.5 Model Evaluation

I compared the performance of the initial model with the new model including `'pavement_quality'`. Evaluation metrics such as R^2, Adjusted R^2, MAE, and RMSE were used to assess the models.

OLS Regression Results

Dep. Variable:	accidents	R-squared:	0.894
Model:	OLS	Adj. R-squared:	0.894
Method:	Least Squares	F-statistic:	1.575e+04
Date:	Wed, 29 Jan 2025	Prob (F-statistic):	0.00
Time:	22:20:57	Log-Likelihood:	-13666.
No. Observations:	7004	AIC:	2.734e+04
Df Residuals:	6999	BIC:	2.738e+04
Df Model:	4		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
const	23.5468	0.078	303.209	0.000	23.395	23.699
traffic_fine_amount	-1.4887	0.008	-192.893	0.000	-1.504	-1.474
rain_intensity	1.9742	0.024	83.148	0.000	1.928	2.021
pavement_quality	-0.7202	0.014	-50.728	0.000	-0.748	-0.692
traffic_density	0.7985	0.007	114.058	0.000	0.785	0.812

Omnibus:	9.693	Durbin-Watson:	2.030
Prob(Omnibus):	0.008	Jarque-Bera (JB):	9.690
Skew:	-0.083	Prob(JB):	0.00787
Kurtosis:	2.927	Cond. No.	32.6

4.6 Interpretation of Coefficients

The coefficients of the independent variables were analyzed to interpret their significance and influence on traffic accidents. I focused on the impact of traffic density, rain intensity, and traffic fines, and evaluated whether pavement quality improved the model's predictive power.

Multiple Linear Regression Equation:

Accidents = 23.55 - 1.49 × traffic_fine_amount + 1.97 × rain_intensity - 0.72 × pavement_quality + 0.798 × traffic_density

5 Key Findings and Insights

- **Traffic Fine Amount:** Higher traffic fines tend to correlate with fewer accidents, highlighting the role of enforcement in promoting safer driving.
- **Traffic Density:** Higher traffic density significantly increases accident risk, particularly in urban areas.
- **Pavement Quality:** Poor road conditions (low pavement quality) are strongly linked to more accidents, stressing the need for better infrastructure.
- **Rain Intensity:** Heavy rain increases the likelihood of accidents, emphasizing the importance of weather preparedness for driving.

- **Urban vs Rural Areas:** Urban areas exhibit higher accident rates due to higher traffic density and more complex infrastructure.

6 Strategic Actions

- **Insurance Companies:** Could use rain intensity, pavement quality, and traffic density in their risk models to adjust premiums based on environmental and traffic conditions.
- **Government and Urban Planners:** Should focus on improving pavement quality and reducing traffic congestion through smart urban planning and investment in public transport infrastructure.
- **Transportation Services:** Could consider developing systems that promote safe driving behaviors, especially during poor weather conditions, by offering incentives or penalties.

7 Conclusion

This analysis provides valuable insights into the key factors affecting traffic accidents. The findings can guide future urban planning, infrastructure improvements, and traffic safety measures. Policymakers can use this information to allocate resources efficiently and prioritize areas for road safety initiatives. Additionally, predictive models based on this dataset can be further developed to predict accident hotspots, which can be crucial in proactive traffic management and accident prevention.