Multivariate Regression Model Estimating Injury Counts Based on Factors

EMILY SONG[1]

[1]Duke University

Duke BASS CONNECTIONS

Global Health

Introduction

Predicting the number of injuries resulting from attacks is essential for resource allocation and emergency response in the context of gauging reconstructive surgery need. We developed a **multivariate regression** that estimates daily injury counts based on attack characteristics, such as attack type, population density, and other relevant independent variables. Given that multiple factors contribute to injury severity and frequency, a multivariate approach was chosen to capture the complex interactions between multiple contributing factors.

All statistical analyses and model implementations were conducted using Python. The models were trained on one year of injury data (October 2023 – October 2024) and subsequently used to predict injury counts through January 2025. The models were implemented using the Statsmodels library (v. 3.12.2).

Selection of Independent Variables

Independent Variable	Source	Justification
Attack Type (Airdrone attacks, Shelling/artillery attacks, Ground attacks, Civil unrest, and Other)	Armed Conflict Location & Event Data Project (ACLED) Database	• Different attacks (e.g., airstrikes vs. ground combat) cause distinct injury severities and scales. • Airstrikes and explosives result in mass casualties, while ground combat is more localized. • Standardized ACLED classifications to address inconsistent reporting.
Population Density: (High density (> 20,000 people/km²), Medium density (5,000–20,000 people/km²), Low density (< 5,000 people/km²)	Situational Reports from various sources	• Injury patterns vary by target (e.g., residential vs. medical vs. industrial). • Attacks on medical facilities worsen outcomes by limiting care access. • Urban areas see higher casualties; rural areas face delayed medical response.
Infrastructure Type (urban, suburban, rural, tent, rubble, or camp)	Situational Reports from various sources, satellite imagery	• Higher density increases civilian exposure and casualties. • Dense areas overwhelm emergency response and delay medical care. • Used satellite and geospatial data to estimate real-time exposure.

Generalized Negative Binomial Model Results

Descriptive Statistics	Pseudo R-squared		Pearson chi-squared
	0.8623		337
Independent Variable	Coef.	Std. error	P-value
const	6.2393	0.052	0.000
Civil_unrest_count	-0.1348	0.042	0.001
Mean_population_density	3.9109	0.397	0.000
Rubble_count	0.0438	0.013	0.001
Camp_count	-0.0855	0.021	0.000
Suburban_count	0.3088	0.059	0.000
Airdrone_x_rubble int. factor	0.0091	0.001	0.000
Airdrone_x_tent int. factor	0.0759	0.022	0.001
Ground_x_camp int. factor	-0.0221	0.005	0.000
Predicted # of injuries (95% CI)	81971.41 (-276346.31, 440289.14)		
Predicted # of surgeries (95% Ci)	19263.28 (-64941.38, 103467.95)		

Discussion

Our findings demonstrate that a multivariate approach is essential for accurately predicting injury counts resulting from attacks, as multiple interacting factors contribute to the variability in injury outcomes.

The model demonstrated a strong overall fit (pseudo R^2 = 0.8623), indicating that it explains a substantial proportion of variance in injury outcomes. We included only statistically significant variables ($p < 0.05$) that likely contributed meaningfully to predicting injury count. **Population density** had the strongest positive association, suggesting that regions with denser populations will be more greatly affected during wartime. **Rubble** and **suburban** count also showed positive relationships with injuries, reinforcing the idea that structural destruction and population clustering contribute to higher injury rates.

Acknowledgments

Slide 5

FH0 more specific what is the multivariate reg model about? What is this slide doing? Be more precise
Frances Hasso, 2025-03-31T20:08:46.204

ES1 Add more detail to title
Emily Song, 2025-04-03T23:20:30.892

ES2 Add more detail to subheaders
Emily Song, 2025-04-03T23:20:50.997