Statistics 506 Project Proposal

Modeling Spatial and Temporal Patterns in Road Crashes: Evidence from York, UK

Brynn Woolley

1   Research Question

How do roadway design, environmental conditions, and time-of-year influence crash frequency in York, UK, after accounting for overdispersion and seasonal variation? Specifically, which roadway or environmental features are associated with higher monthly crash rates across space?

2   Data Description

I will use the publicly available *UK Department for Transport* stats19 crash dataset (accessed via the stats19 R package). It contains all police-reported road accidents in Great Britain with geocoded locations (approximately 10 m accuracy), time of occurrence, and contextual information on lighting, weather, surface condition, road type, and crash severity. This project focuses on the **York 2023 subset** (about 4,200 crashes), using only fields available within the standard casualties_wy or accidents data objects.

3   Analysis Plan

Crash counts will be aggregated by **regular hexagonal grid cells** and by **month**. The response variable will be the number of crashes per cell per month (y_{it}) with a log(days-in-month) offset to normalize month lengths. I plan to fit a **negative binomial regression model** with a log link to account for overdispersion:

$$y_{it} \sim \mathrm{NB}(\mu_{it}, \kappa), \qquad \log \mu_{it} = \log d_t + \beta_0 + \mathbf{X}_{it}^\top \boldsymbol{\beta} + \delta_t,$$

where $\mathbf{X}_{it}$ includes road class, speed limit, junction type, lighting, weather, and surface condition, and δ_t are month fixed effects. If diagnostics show excess zeros or spatial dependence, I will test a hurdle or mixed-effects variant as a sensitivity check.

All analysis will be conducted in **R** using sf, tidyverse, and MASS. Code and replication materials will be shared in a GitHub repository.