**Statistical Analysis of Traffic Incidents**

**and Weather Conditions**

A logo of a university of calgary

Description automatically generated

**Data 602 Group Project**

Anitha Joseph, Jincy Thomas, Megha Radhakrishnan Sanitha, Prichelle Lal

February 12th, 2025

**Introduction**

Transport Canada has undertaken to release a subset of its National Collision Database (NCDB) – a database containing all police-reported motor vehicle collisions on public roads in Canada. Selected variables (data elements) relating to fatal and injury collisions for the collisions from 2001 to 2021 data is collected from NCDB Online.

Highlights

1. Do weather variables, such as wind, rain, snow, speed limit, significantly influence road traffic incidents?
2. what are the specific effects of each weather parameter on traffic conditions? Additionally, to what extent are these impacts likely to be interconnected?
3. Gaining insight into the connection between weather and traffic could lead to more effective management of traffic and transit systems.

**Literature review**

<we have to some sentences here>

**Statistical Analysis**

**Summary of Tasks:**

1. **Task 1:** Monte Carlo simulation for high fatalities (threshold = 20).
2. **Task 2:** Monte Carlo simulation for at least 2 fatalities.
3. **Task 3:** Poisson regression to analyze the impact of weather conditions.
4. **Task 4:** Statistical tests for fatalities over time (Chi-Square and Linear Regression).

**Task 1: Monte Carlo Simulation for Predicting the Probability of High Fatalities in Weather Conditions**

**Objective:** Perform a Monte Carlo simulation to estimate the probability of exceeding a specific number of fatalities (e.g., 20) for each weather condition.

**Steps:**

1. **Data Preparation:**
   * Use the dataset to determine the average rate of fatalities per year for each weather condition (e.g., snowing, raining, freezing rain).
2. **Monte Carlo Simulation:**
   * Simulate the number of fatalities for each weather condition using a Poisson distribution.
   * Run 1,000 simulations for each weather condition.
   * Count how many times the simulated fatalities exceed the threshold (e.g., 20).
3. **R Script Implementation:**
   * Write an R script to perform the simulation and calculate the probability of exceeding the threshold for each weather condition.
4. **Output:**
   * Provide the probabilities of exceeding the threshold for each weather condition.
   * Compare the Monte Carlo results with theoretical Poisson probabilities using ppois().

**Task 2: Monte Carlo Simulation and Probability Estimation for At Least 2 Fatalities**

**Objective:** Estimate the probability of at least 2 fatalities occurring in a year for each weather condition using Monte Carlo simulation.

**Steps:**

1. **Data Preparation:**
   * Determine the average number of fatalities per year (λ) for each weather condition from the dataset.
2. **Monte Carlo Simulation:**
   * Simulate the number of fatalities for each weather condition using a Poisson distribution (10,000 simulations).
   * Count how many times the simulated fatalities are ≥ 2.
3. **R Script Implementation:**
   * Write an R script to calculate the probability of at least 2 fatalities for each weather condition.
4. **Comparison with Exact Poisson Probability:**
   * Compare the Monte Carlo probabilities with the exact Poisson probabilities using the formula:

P(X≥2)=1−P(X=0)−P(X=1)*P*(*X*≥2)=1−*P*(*X*=0)−*P*(*X*=1)

1. **Output:**
   * Present the results in a table and/or plot, showing both Monte Carlo and exact Poisson probabilities.

**Task 3: Poisson Regression for Analyzing the Impact of Weather Conditions on Fatalities**

**Objective:** Model the relationship between weather conditions and fatalities using Poisson regression.

**Steps:**

1. **Data Preparation:**
   * Prepare the dataset with fatalities as the dependent variable and weather conditions as independent variables.
2. **Poisson Regression:**
   * Fit a Poisson regression model in R.
   * Test the significance of weather conditions using likelihood ratio tests.
3. **Hypothesis Testing:**
   * Null Hypothesis (H₀): Weather conditions have no effect on fatalities.
   * Alternative Hypothesis (H₁): At least one weather condition has a significant effect.
4. **Output:**
   * Provide the regression coefficients and p-values for each weather condition.
   * Interpret the results at a significance level of 0.05.

**Task 4: Statistical Tests for Fatalities Over Time**

**Objective:** Analyze changes in fatalities over time using statistical tests.

**Subtasks:**

1. **Chi-Square Test for Independence:**
   * Test if the distribution of fatalities across weather conditions is independent of the year.
   * Null Hypothesis (H₀): Fatalities are independent of the year.
   * Alternative Hypothesis (H₁): Fatalities depend on the year.
   * Perform the test in R and interpret the results at a significance level of 0.05.
2. **Linear Regression for Snowing Conditions:**
   * Test if the rate of fatalities in snowing conditions has changed over time.
   * Null Hypothesis (H₀): The rate of fatalities is constant over time.
   * Alternative Hypothesis (H₁): The rate of fatalities has changed over time.
   * Perform linear regression in R with fatalities as the dependent variable and year as the independent variable.
   * Interpret the slope and p-value at a significance level of 0.05.

**Conclusion**

**Recommendations & Next Steps**

**References**