Lab 05: Logistic Regression

**Due date:** Thursday, March 21, 2025 submitted as Word document to Canvas ***Lab05*** link

This lab counts 9 % toward your total grade.

**Objectives:** In this lab, you will practice your skills in

1. Explore logistic regression
2. All effects plot
3. Conditional effects plot

**Format of answer:** Submit your answers as a **Word document** with graphs and verbal descriptions, properly labeled in the task sequence, with answers in red text and only relevant content included

**Accident\_data.shp** show the spatial distribution of traffic accidents in the city of Dallas. The accident is represented as 1 (Yes, a traffic accident occurred at the location) and 0 (No, no traffic accident occurred at location).

A map with red and blue dots

AI-generated content may be incorrect.

# Task 1: Load data (2 pts)

We will use the **Accident\_data.shp** to practice the skills in logistic regression

1. Load the **Accident\_data.shp** data using **sf::st\_read()** (1 pt)
2. Check the geometry column and explain what information are stored in geometry column using **data$geometry**. (0.5 pt)
3. Convert column ‘**accident**’ to factor. (0.5 pt)

The dependent variable in the data set is whether there is a traffic accident at the location.

|  |  |
| --- | --- |
| **Dependent variable** | **Description** |
| **accident** | a factor with levels: ‘0’; ‘1’ |

The independent variable describes the geographical environments around the location.

|  |  |
| --- | --- |
| **Independent variable** | **Description** |
| **Hour** | Hour of the day, from 0 - 23 |
| **F\_SYSTEM** | Road type at different levels: 0 -7 |
| **intersect** | Whether the accident occurred at a major intersection |
| **NUM\_LANES** | The number of lanes |
| **lane\_width** | Lane width |
| **BelowFreez** | Whether the temperature is below freezing, a factor with levels: ‘FALSE’; ‘TRUE’ |
| **Fog** | Whether foggy conditions are present : ‘FALSE’; ‘TRUE’ |
| **Thunder** | Whether thunderstorms are present: ‘FALSE’; ‘TRUE’ |
| **FrozenPrec** | Whether there is frozen precipitation: ‘FALSE’; ‘TRUE’ |

# Task 2: Build logistic regression model (3 pts)

1. Build a logistic regression for the probability of accident with independent variable list above and provide the 95% confidence intervals around the logistic regression parameters.
2. Discuss the model output from logistic regression (significant of the variable) and analyze whether the results from confidence intervals align with the model results.
3. Using all effects plot to interpret the calibrated logistic regression model in terms of probabilities.

# Task 3: Perform one likelihood ratio test (2 pts)

1. Refine the model from Task 2 by dropping all variables that are not significant important for dependent variable. Test whether the removed variables had a significant influence on the dependent variable using ANOVA-based LRT or Manual LRT (logLik()). (2 pts)

Bonus (+1 point): Try both approaches and compare results. Explain if they have the same result or not.

# Task 4: Conditional effects plots (2 pts)

1. Generate conditional effects for the NUM\_LANES variable based on the refined model in Task 2.a to estimate the probability of traffic accident. (1.5 pts)

Assume two scenarios with the following values for the additional independent variable in the logistic regression model:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Low probability** | **High probability** |
| **hour** | 3 | 19 |
| **F\_SYSTEM** | 6 | 2 |
| **Intersect** | ‘no’ | ‘yes’ |

1. Discuss the conditional effects plots for the two scenarios. (0.5 pt)