

# STA304 Paper 1: relationship between crime and mental health

Max LI

2024-09-20

## Motivation

### Purpose

This dataset is created to explore the relationship between major crime indicators (MCI) and mental health apprehensions (MHA) across various districts in Toronto. The goal is to use crime data and MHA reports to predict crime trends using a linear regression model and identify any potential correlations between these factors.

### Context

The dataset combines two public datasets from Toronto's Open Data portal: 1. [Major Crime Indicators](#) 2. [Mental Health Apprehensions](#)

The analysis intends to assess whether MHA reports can explain variations in crime numbers and to identify residual patterns that may suggest areas for further investigation.

## Composition

### Variables

The dataset contains the following variables:

- **DIVISION**: The police division (or district) in Toronto, represented by codes such as D11, D12, etc.
- **num\_crimes**: The total number of crimes recorded in a given division.
- **num\_mha\_reports**: The total number of MHA reports in the corresponding division.

- **predicted\_crimes:** The predicted number of crimes in the division, generated from the linear model using MHA reports as a predictor.
- **residuals:** The difference between actual crime numbers and predicted values (actual - predicted), indicating how well the model performed for each division.

## Observations

The dataset includes 18 observations (one for each police division in Toronto).

## File Sources

- **raw\_data\_mha.csv:** Contains raw MHA reports by division.
- **raw\_data\_mci.csv:** Contains raw major crime reports by division.
- **clean\_data.csv:** The combined and cleaned dataset with crime and MHA information.
- **clean\_data\_with\_predictions.csv:** The final dataset that includes crime data, MHA reports, predictions, and residuals from the model.

## Data Sources

- **MHA Data:** [Mental Health Apprehensions](#), provided by Toronto Police.
- **Crime Data:** [Major Crime Indicators](#), provided by Toronto Police.
- **Model Outputs:** The final analysis dataset includes predictions from a linear regression model built in R.

## Collection Methods

1. **Download (01-download\_data.R):** The data was retrieved from Toronto's open data portal and saved locally.
2. **Cleaning (02-clean\_data.R):** The MHA and MCI datasets were cleaned and merged by division. Each division's crime and MHA counts were aggregated to provide an overall view for the analysis.
3. **Testing (03-test\_data.R):** Initial checks and summaries were conducted to ensure that the cleaned data had valid and interpretable values.
4. **Modeling (04-model.R):** A linear regression model was fit to predict `num_crimes` using `num_mha_reports` as the independent variable. The model output included predictions and residuals for each division.

5. **Replications (05-replications.R):** The predictions from the model were compared to the actual crime data, and residuals were calculated. These were saved in a final CSV file.

## Limitations

- **Sample Size:** The dataset is limited to 18 divisions, which may not provide enough variability for a robust model.
- **Linear Model:** The assumption of a linear relationship between crime and MHA reports may not fully capture the complexities of these variables.
- **Residuals:** The residuals suggest that other variables (not captured in this dataset) may explain the difference between predicted and actual crime numbers.

## References

- Toronto Open Data Portal: Major Crime Indicators ([Link](#))
- Toronto Open Data Portal: Mental Health Apprehensions ([Link](#))