Socio-Economic Indicators and Crime: A Regression Analysis of DV and Violent Crime (Queensland)

**Author:** Davi Santos Meloni  
**Date:** April 2025

**Note**: This analysis was conducted as part of a self-guided learning project. While the methods and interpretations are grounded in established statistical practices, this is the author's first attempt at modeling real-world datasets. As such, results should be viewed as exploratory and subject to the typical limitations of an individual learning exercise.

# 1. Introduction

The purpose of this analysis is to understand how socio-economic factors, as measured by SEIFA indexes, relate to violent and domestic violence (DV) crime rates in Queensland. This analysis specifically focuses on violent crime and DV-related crime (Breach Domestic Violence Protection Orders), which were chosen due to their severe impact on communities and their strong links to socio-economic disadvantage.

By narrowing the scope to these crime categories, the analysis eliminates potential noise from other crimes like property and drug-related offences, which tend to be more closely linked to factors like economic gain and substance abuse. This allows a clearer understanding of the relationship between socio-economic disadvantage and violent interpersonal crimes, which have more direct implications for community wellbeing.

# 2. Data Collection and Preparation

## Crime Data:

* Sourced from the most recent available reporting period, covering offences from 2018 onward.
* Raw offence types were consolidated into broader categories, focusing on:
  + Violent Crime
  + Domestic Violence (DV)-Related Crime (breaches of Domestic Violence Protection Orders)
* Original dataset included 335 Queensland police divisions.

## Socio-Economic Data:

* Obtained from the 2021 SEIFA indexes published by the Australian Bureau of Statistics.
* Included the following indexes:
  + IRSDA: Index of Relative Socio-Economic Advantage and Disadvantage
  + IRSD: Index of Relative Socio-Economic Disadvantage
  + IER: Index of Education and Employment
  + IEO: Index of Economic Opportunity

**Data Matching and Cleaning:**

* Matched crime data from 335 police divisions to corresponding suburb-level SEIFA indexes.
* Excluded divisions that:
  + Could not be reliably matched to SEIFA suburbs.
  + Had minimal or insufficient population sizes for meaningful analysis.
* Final dataset retained 226 matched divisions for analysis.

**3. Methodology**

**Statistical Approach:**

* The analysis employed several methods to explore the relationships between socio-economic factors and crime rates:
  + **Pearson Correlation:** To examine the linear relationships between each SEIFA index and violent/DV crime rates.
  + **Simple Linear Regression:** To quantify the strength and significance of these relationships.
  + **Multicollinearity Check:** Correlation analysis and Variance Inflation Factor (VIF) were used to identify and address multicollinearity among the SEIFA indexes.
  + **Lasso and Ridge Regression:** These models were employed to account for multicollinearity and provide more reliable estimates of the relationships.

**Model Performance Metrics:**

* **R² (R-squared):** Used to determine how well the models explained the variance in violent and DV crime rates.
* **MSE (Mean Squared Error):** Used to evaluate the predictive accuracy of the models.

**4. Exploratory Data Analysis**

**Pearson Correlation Results:**  
The Pearson correlation test revealed the following key relationships:

* **IER Score:** Moderately strong negative correlation (r = -0.610, p < 0.001). Suburbs with higher economic resources tend to have lower violent/DV crime rates.
* **IRSD Score:** Moderate negative correlation (r = -0.510, p < 0.001). Areas with greater socio-economic disadvantage have higher crime rates.
* **IRSDA Score:** Weaker negative correlation (r = -0.327, p < 0.001), consistent with the overall trend but weaker in magnitude.
* **IEO Score:** No significant correlation (r = -0.105, p = 0.114). The lack of statistical significance suggests that education and occupation factors alone may not explain violent or DV crime rates effectively.

**5. Simple Linear Regression Analysis**

**Regression Results:**  
The R² values for each index in the simple linear regression models were as follows:

* **IER Score:** R² ≈ 26%, explaining approximately 26% of the variance in violent/DV crime rates.
* **IRSD Score:** R² ≈ 20%, explaining around 20% of the variance.
* **IRSDA Score:** R² ≈ 9%, indicating a weaker explanatory power.
* **IEO Score:** R² < 2%, indicating a very weak relationship.

**Model Comparison:**  
The Pearson correlation and R² values from the linear regression provided complementary insights. For instance, IER had the strongest negative correlation (r = -0.610), and this was reflected in the high R² (~26%), indicating that IER is a significant predictor of violent/DV crime rates.

**Visualisation**  
*(Placeholder for charts/plots as per original document.)*

**6. Addressing Multicollinearity**

**VIF and Correlation Check:**

* A correlation analysis and VIF calculation revealed significant multicollinearity among all the SEIFA indexes, even when subtracting different indexes from the analysis.
* To account for multicollinearity, Lasso and Ridge regression models were employed, which are better suited to handle correlated predictors.

**7. Lasso and Ridge Regression Results**

**Lasso and Ridge Model Coefficients:**  
The coefficients from the Lasso and Ridge regression models confirmed the findings from the Pearson correlation and linear regression:

* **Lasso:**
  + IRSDA (0.0)
  + IRSD (-0.256)
  + IER (-0.337)
  + IEO (0.259)
* **Ridge:**
  + IRSDA (0.042)
  + IRSD (-0.288)
  + IER (-0.325)
  + IEO (0.237)

Both models show IER as a key predictor, with a negative relationship to crime rates, confirming that higher economic resources are associated with lower crime rates.

**Model Performance Comparison:**  
Both models provided similar performance metrics:

* **R² (Lasso):** 0.307
* **R² (Ridge):** 0.307
* **MSE (Lasso):** 0.532
* **MSE (Ridge):** 0.532

The close agreement between the models indicates reliable and consistent results, suggesting that IER plays a somewhat important role in predicting violent and DV crimes.

**Visualisation**  
*(Placeholder for coefficient plots or error metrics.)*

**8. Discussion and Interpretation**

**Overall Impact**  
Across all SEIFA indexes, the models indicate that they collectively explain approximately 30% of the variance in violent and DV crime rates. This is consistent with the Lasso and Ridge regression models, which provided slightly better predictive power than the simple linear regression, suggesting that these socio-economic factors play a notable role in shaping crime rates, particularly in the context of violent and DV crimes.

While the models explain part of the variance, it’s clear that there are other factors influencing violent and DV crime rates that weren’t captured in the analysis.

**Key Insights**

* The **IER (Economic Resources)** index was the most significant predictor of violent and DV crime rates, showing a strong negative relationship. Suburbs with lower economic resources tend to experience higher crime rates, which was evident both in the Pearson correlation (r = -0.610) and in the Lasso and Ridge regression models, where the relationship remained moderately strong and negative. This supports the idea that improving economic resources may be one of the most effective ways to reduce violent and DV crimes.
* On the other hand, the **IEO (Education/Occupation)** index showed no meaningful relationship with violent and DV crime rates. While the Pearson correlation indicated a weak negative relationship (r = -0.105), it was statistically insignificant (p = 0.114). More surprisingly, the Lasso and Ridge regression models produced positive coefficients for IEO, suggesting a very weak, albeit unexpected, relationship. This implies that education and occupation alone may not significantly impact violent and DV crime rates, potentially due to other, more dominant socio-economic factors influencing crime in these areas.

**Data Sources**

* **SEIFA Indexes (2021)**  
  Australian Bureau of Statistics  
  <https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/latest-release#data-downloads>
* **Crime Data by Police Division (2018–Latest)**  
  Queensland Government Open Data Portal  
  <https://www.data.qld.gov.au/dataset/offence-numbers-police-divisions-monthly-from-july-2001>
* **Code Repository**  
  GitHub – SEIFA & Crime Rate Analysis (Queensland)  
  <https://github.com/davimeloni/qld-seifa-crime-analysis.git>