Problem Statement

The objective of this study is to develop an effective predictive model for violent crime rates (crimes per 100,000 people) using demographic and socio-economic indicators at the state level in the United States. The dataset comprises variables such as state ID, percent of the population in metropolitan areas, percent white, education level, poverty, and single-parent household prevalence. The goal is to identify significant predictors and suggest actionable insights to help reduce crime rates.

Crime Rate Prediction Using SAS – Report

1. Data Overview and Preprocessing

The dataset includes the following variables:

- sid, state: Identifiers
- crime: Target variable (Violent crimes per 100,000 people)
- Predictors: murder, pctmetro, pctwhite, pcths, poverty, single

Steps Taken:

- Removed identifier columns (sid, state) for modeling
- Checked for missing values and outliers
- Standardized continuous variables where needed
- Explored correlations to detect multicollinearity

2. Exploratory Data Analysis

- Strong positive correlations were observed between crime and variables like poverty and single.
- pcths and pctwhite showed a weak or negative correlation with crime.

3. Model Development

The dataset was divided into:

- Training Set: 70%
- Testing Set: 30%

We built multiple regression models:

- Model 1: Multiple Linear Regression with all predictors
- Model 2: Backward elimination to remove insignificant predictors
- Model 3: Stepwise Regression
- Model 4: Ridge Regression for multicollinearity handling
- Model 5: Cross-validated Linear Regression (10-fold)

Best Model:

The Stepwise Regression Model selected the best subset of predictors with high adjusted R^2 and low RMSE on the test set. The final predictors were:

- poverty
- single
- pctmetro

4. Regression Assumptions Checked

- Linearity: Satisfied through residual plots
- Normality: Histogram and QQ-plots of residuals showed near-normal distribution
- Homoscedasticity: Residuals showed constant variance

- Multicollinearity: VIF values < 5 in the final model

5. Evaluation on Test Data
- RMSE and MAE metrics were calculated
- Predicted vs Actual plots indicated good fit with minimal bias

6. Suggestions to Reduce Crime
Based on the model:
- Reduce poverty: Strongest positive predictor of crime
- Support for single-parent households: Correlated with higher crime rates
- Invest in urban safety measures: pctmetro showed mild positive correlation with crime

Policy Recommendations:
- Launch targeted education and employment programs in high-poverty areas
- Increase access to childcare and support services for single parents
- Community policing and urban planning initiatives in high-density metro areas