UNITED STATES MILITARY ACADEMY
WEST POINT

Inferential Analysis of Burglary Counts in Chicago Census Blocks

CDTs Evan Asuncion and Aimee Rohan

v2

Background & Data

- Chicago afflicted with high crime rates across the city.
- Policymakers and law enforcement need to implement well-informed policies/initiatives to confront crime issue.

Chicago News Is Nothing but Crime

Problem Statement

Our research seeks to infer the impact of both socio-economic and temporal variables on burglary rates in Chicago's Census Block groups.

Research Questions

- Do unemployment, wealth, or presence of youth influence differences in burglary counts?
- Do month and year affect the number of burglaries within Chicago?
- Are these potential explanatory variables solely covariates or is there some aspect of randomness?

Dataset

- GitHub provide raw data of 552 distinct Census block groups in Chicago tracking burglary frequency from 2010 to 2015.
- Supplementary Data: Adjacency matrix, population stats, unemployment rates, wealth distribution, and count of young males.
- Final dataset: 39,744 observations of blocks in a given month and year

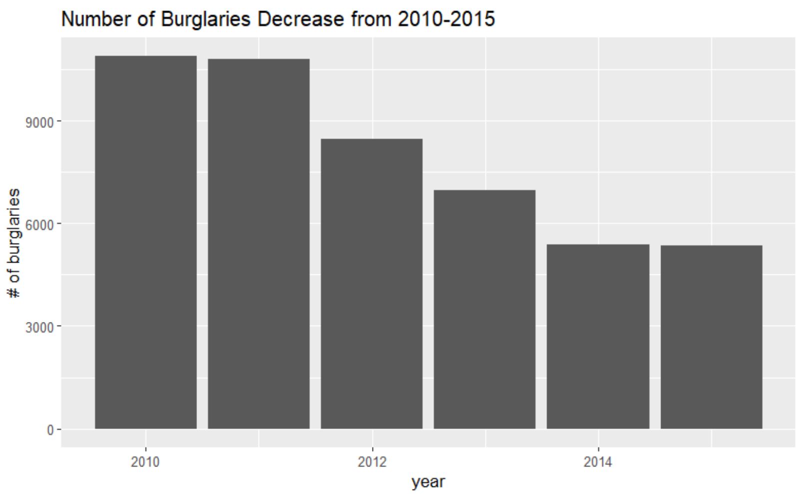
UNITED STATES MILITARY ACADEMY
WEST POINT

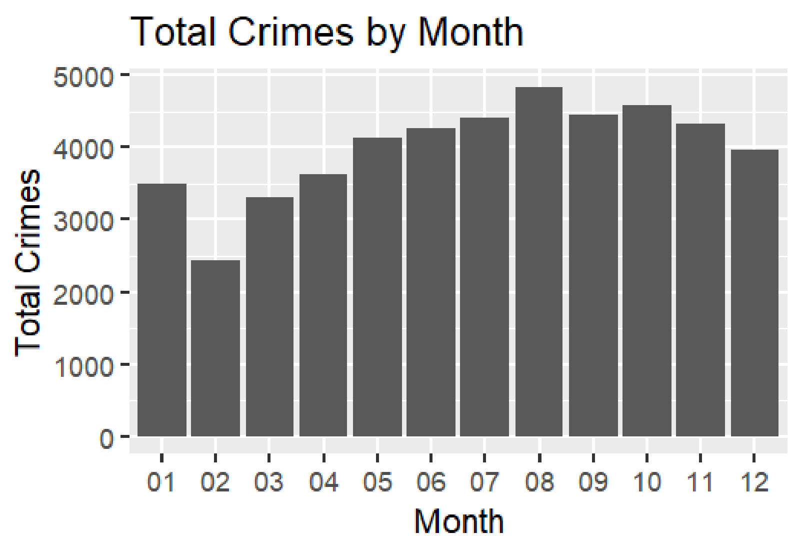
Methodology

- Data Transformation:
  - Added a column for the year and month data quantitatively.
  - Grouped the year variable into categories ranging from 0 to 5 (2010-2015), put month variable on scale from 0 to 12, standardized population variable.
- Data Exploration:
  - Conducted exploratory data analysis to understand the relationships between the covariates and the response variable.
  - Examined the distribution of variables, correlations, and potential patterns in the data.
- Model Selection:
  - Identified that count data were best modeled using Poisson, zero-inflated Poisson (ZIP), mixed-effects models, or generalized linear mixed-effects models (GLMMs).
  - Used AIC to compare models if they have the same y.
- Modeling:
  - Selected appropriate modeling techniques based on the observed relationships and data characteristics.
  - Fit all models via MLE.
  - Assessed variable significance, model fit and performance using goodness of fit/likelihood ratio tests and AIC.

Data Exploration: Temporal Trends

- Visual analysis reveals a declining trend in burglary counts over the years.
- A declining trend in burglary counts is also noted visually across the colder months.

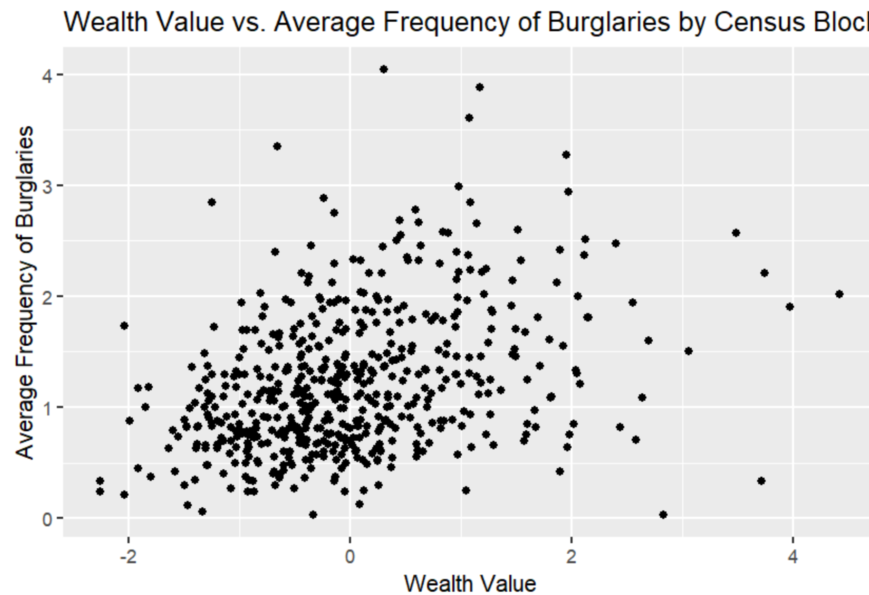


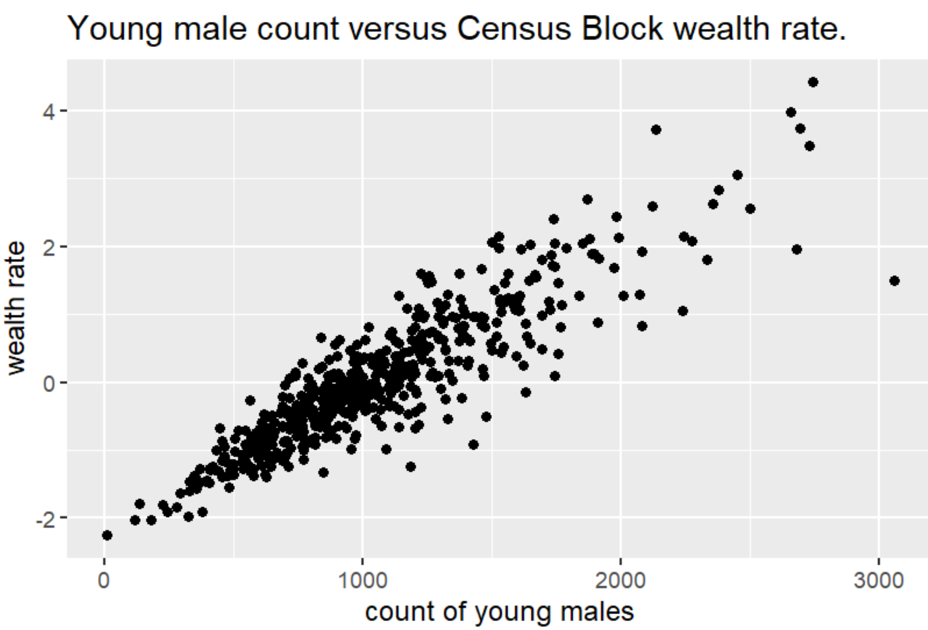


UNITED STATES MILITARY ACADEMY
WEST POINT

Analyzing Wealth

- A Positive trend observed: Higher wealth ratings in census blocks correspond to increased average burglary counts.

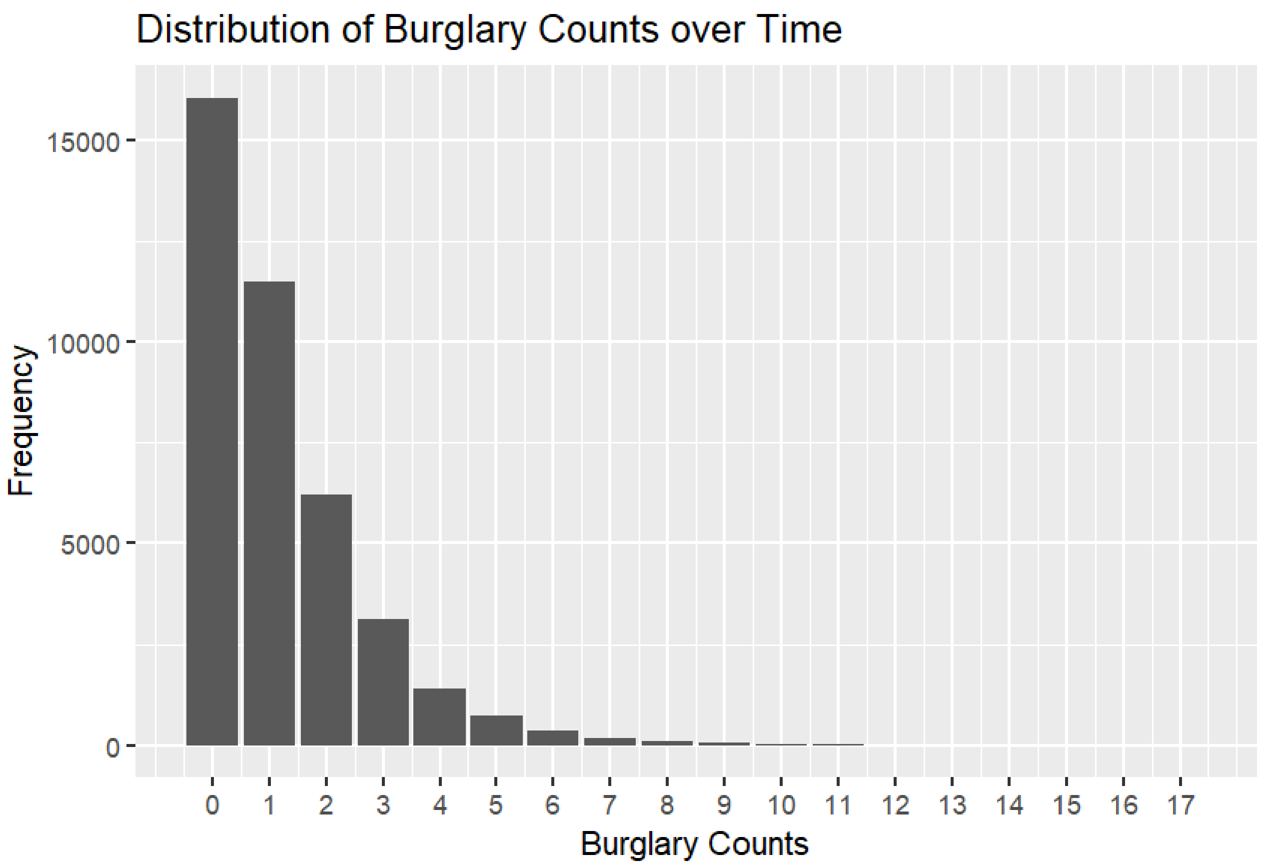




- Possible Multicollinearity due to strong visual relationship between wealth and count of young men per census block.

Distribution of Burglary counts

Graphing the reported distribution of burglary counts indicate reported count to predominantly be 0 occurrences, while also declining as counts increase.



UNITED STATES MILITARY ACADEMY
WEST POINT

Objectives

- Find a model that best fits the data to explain what are factors in increasing burglary counts.
- Within each model conduct tests (likelihood ratio, AIC, etc.) to verify factor significance.
- Explain if there was a need to adding a random effect on month or block.

Challenges

- Issues with adding year as a covariate due to there not being many years present (only 6).
- Facing under-dispersion (mixed model) and possible over dispersion issues.
- How will the high frequency count in 0's affect our models' accuracy?

Poisson GLM

$$y_{ijk} = \textit{count of burglary for census block } i \textit{ in year } j \textit{ and month } k$$

$$y_{ijk} \sim Pois(\lambda_{ijk})$$

$$\eta_{ijk} = \log(\lambda_{ijk})$$

$$\log(\lambda_{ijk}) = \beta_0 + \beta_1 Wealth_i + \beta_2 Year_j + \beta_3 Unemployment_i + \beta_4 Month_k + \log(Population_i\ )$$

- Simplest model for count data.
- Assuming there is not overdispersion in our data.
- Variable selection resulting from data exploration.
- Added offset term to account for differences in population density.

Results

- Subsequent likelihood ratio tests suggest that wealth, month, year, and unemployment are all significant fixed effects.
- AIC = 120958
- Year (est: -.165, p-val <2 e-16):For each one-unit increase in year, the expected count of crimes decreases by a factor of 0.848.
- Unemployment (est: 0.469, p-val <5.5 e-14): For each one unit increase in unemployment, the expected count of crimes increase by a factor of 1.59.
- Goodness of fit test: [0] suggests that we prefer a saturated model. This means that our model does not adequately capture the variance.

ZIP Model

$$y_{ijk} \sim \begin{cases} 0 & \text{w/p } \phi_{ij} \\ Pois(\lambda_{ij}) & \text{w/p } (1-\phi_{ij}) \end{cases}$$

$$z_i \sim Bern(\phi_{ij})$$

$$y|z_i = (\frac{e^{-\lambda_i j}\lambda^2}{y!})^{1-z_i}$$

$$\eta_{ij} = log(\lambda_{ij}) = \beta_0 + \beta_1 Wealth_i + \beta_2 Unemployment_i + \beta_3 Year_j + \beta_4 Month_k + log(Population_i)$$

- Confronts issue of large number of blocks in given year with zero reported burglaries.
- Choose zip model to explain that possibly there are many zeros due to there being a mechanism of whether one reports or not.
- OUTPUT: The count model coefficients describe the changes in the expected counts of non-zero events (burglaries). The zero inflated model coefficients describe the likelihood of observing excess zeros.

Results

- Zero inflated model output: **Intercept (-1.51680)** represents the log odds of excess zeros. By it being negative it suggests that there is a low probability of observing excess zeros.
- Count model: all the coefficients in the model have small p-values (less than 0.05) indicating that they are statistically significant predictors of burglary counts.
- AIC = 129534 (w/out month), 129112.3 (month)
- Issues with adding year as a covariate due to there not being many years present (only 5).

GLMM

y_{ijk} = burglary count for census block i in year j and month k

$$y_{ijk} \sim \text{Pois}(\lambda_{ijk})$$

$$\log(\lambda_{ijk}) = \eta_{ijk} = \beta_0 + \beta_1 \cdot \text{wealth\_d}_i + \beta_2 \cdot \text{un.emp}_i + \beta_3 Year_j + b_{0i} + b_{1k}$$

where

y_{ijk} is the observed crime count,

λ_{ijk} is the expected mean of the Poisson distribution,

$b_{0i} \sim \mathcal{MVN}(0, \sigma^2_{\text{block}})$ represents the random effect for block i,

$b_{1k} \sim \mathcal{MVN}(0, \sigma^2_{\text{month}})$ represents the random effect for month k,

- Considers random effect from different census blocks and month.
- We could not add offset population due to lack of convergence. We also did not add year as a fixed effect for this final model due to lack in performance.

Results

- AIC: 111635.6
- Adding random effects due to both block and month yields lower AIC.
- From Wald Test, wealth seems to continue to be statistically significant ($p < 0.05$), but unemployment does not ($p > 0.05$).
- Fails goodness of fit test.

AR-1 Poisson GLMM

y_{ijk} = count of burglary for block i in year j and month k

$y_{ijk} \sim \text{Pois}(\lambda_{ijk})$

$\eta_{ijk} = \log(\lambda_{ijk}) = \beta_0 + \beta_1 \text{Wealth}_i + \beta_2 \text{Unemployment}_i + \beta_3 Year_j + u_k + \alpha_i$

$u_k = \phi_{u_{k-1}} + v_k$

$v_k \sim (0, \sigma_v^2)$

$\alpha_i \sim N(0, \sigma_\alpha^2)$

- Also takes into account random effect in different Census block and month.
- Considers year, wealth, and unemployment as fixed effects.
- AR-1 structure on month assumes some months are more like each other than others (seasonal similarities).

Results

- AIC = 139324.9
- There appears to be a random effect due to block.
- Month alone is not a significant random effect
- Suggests significant fixed effects for year, wealth, and unemployment.
- Under-dispersion? $\phi = 0.1525082$

Conclusion

- No model adequately
  explains variance in Chicago burglaries.
- Cannot compare all models to each other due to response variables not matching.
- Socio-economic effect of wealth and temporal effect of year seem to have significant fixed effect each block's number of burglaries experienced.
- Each block seems to have own random effect.
- Different months bring uniqueness to crime rate.
- Effect of unemployment is not clear.
- Overall models cannot explain the intertwined effects from count of young men and wealth.

UNITED STATES MILITARY ACADEMY WEST POINT

Future work

- Open to better predictive measures if able to capture more information about the population such as gender ratio, ethnicity stats, and education ranking for each block.
- Try fitting a GEE model to capture at the population level impact for burglary frequency in Chicago overall.

Questions?