

Written Report: Mortgage Lending Fairness

Joseph Rodriguez

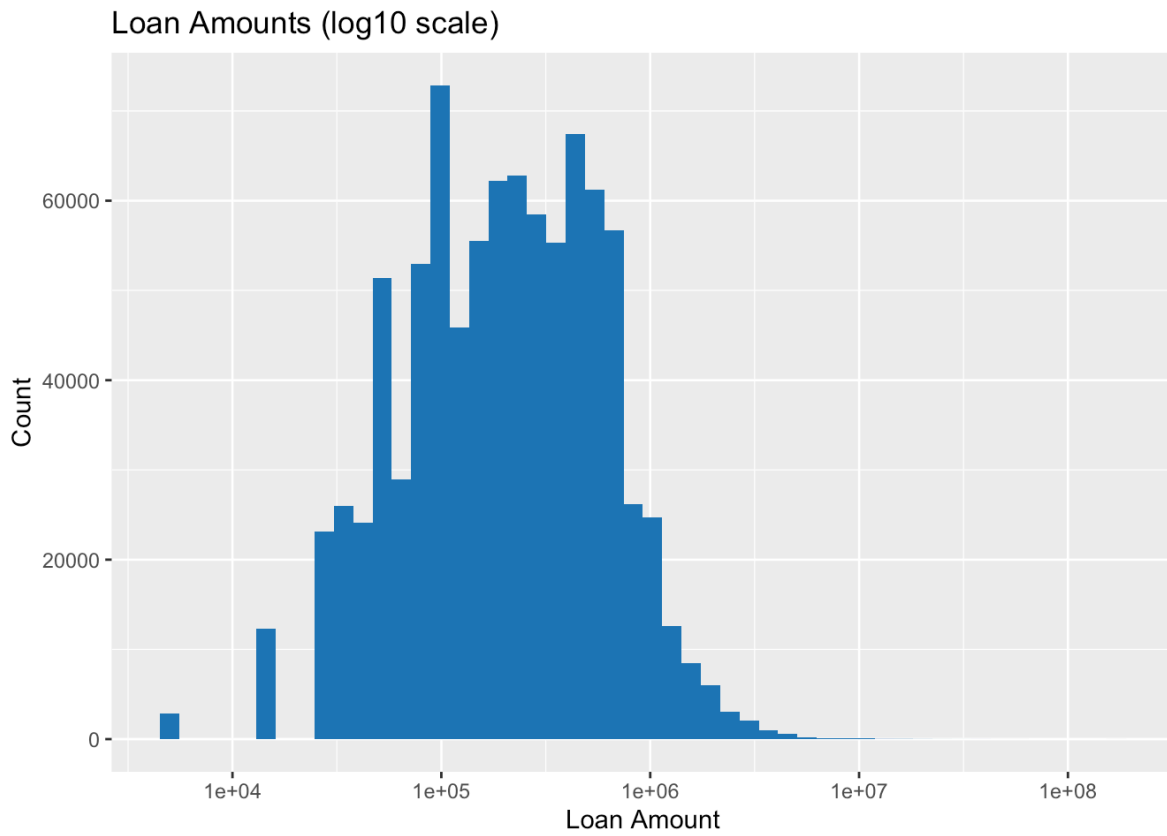
Introduction

This report presents a comprehensive exploration of potential disparities in mortgage lending across racial and regional lines in the U.S., using data from the 2023 Home Mortgage Disclosure Act (HMDA). It includes exploratory data analysis, a logistic regression model examining loan approval decisions, and a Monte Carlo simulation to validate statistical inference procedures.

Exploratory Data Analysis

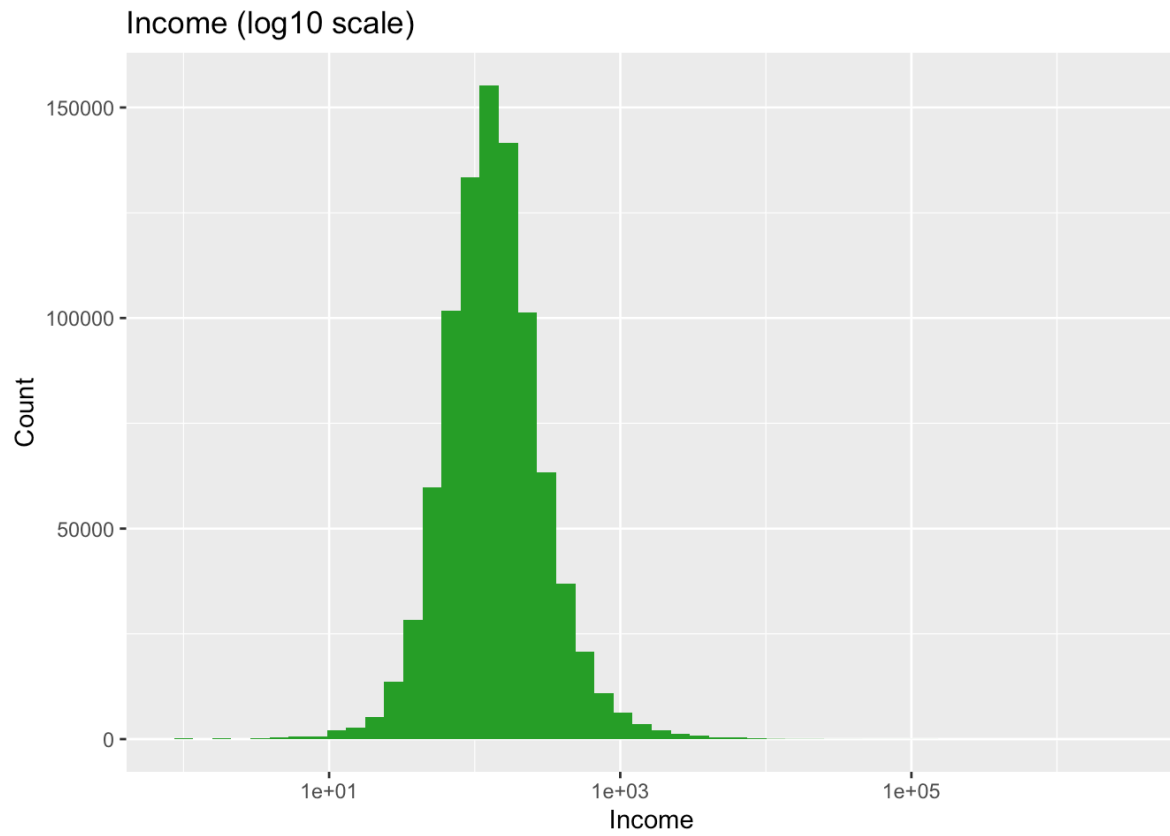
We begin by understanding the distribution of financial variables and loan approval patterns across regions.

Distribution of Loan Amounts



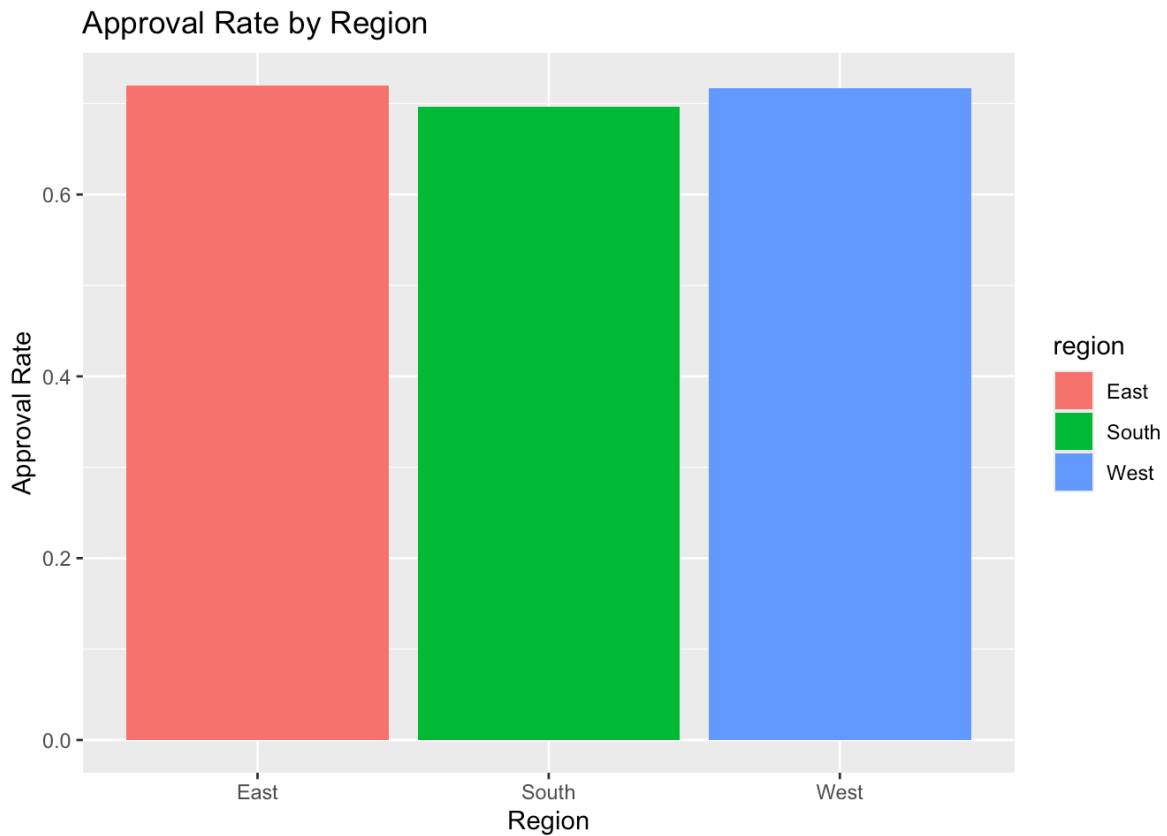
Caption: Loan amounts are right-skewed, with a wide range extending into millions. A log transformation was applied to visualize the distribution more clearly.

Distribution of Income



Caption: Incomes are also highly skewed, resembling a log-normal distribution. Most applicants report moderate incomes, with fewer very high earners.

Approval Rates by Region



Caption: Approval rates are highest in the East and West regions, slightly lower in the South. This raises questions about whether these differences remain after adjusting for other factors.

Project 3: Logistic Regression Modeling

To investigate systemic disparities, we fit a logistic regression predicting loan approval using applicant demographics, region, income, and neighborhood variables.

Model Summary Output

Caption: Model summary showing deviance, log-likelihood, and number of observations used in the logistic regression.

Interpretation

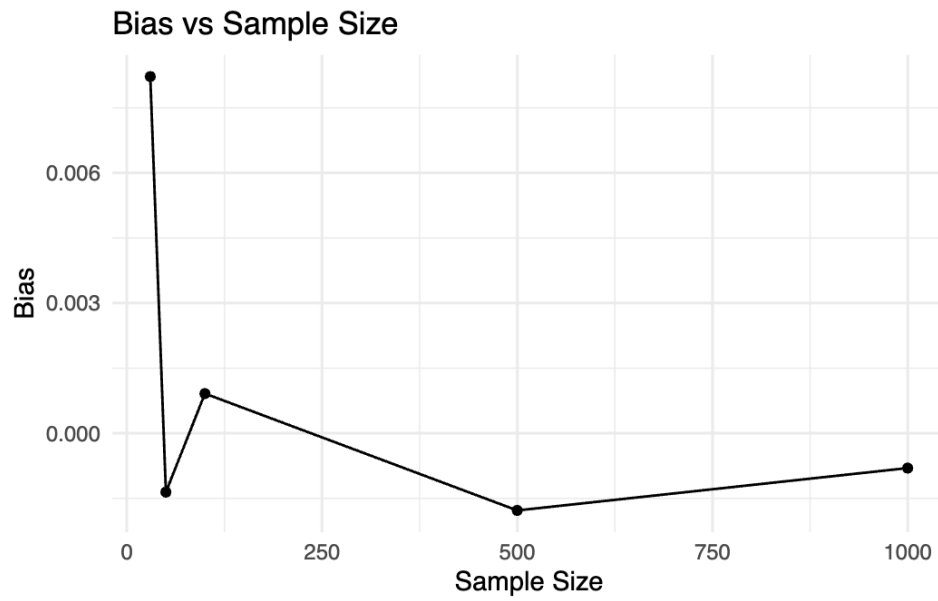
- **Financial factors** (income, loan amount) behave as expected — positively associated with approval.
- **Regional disparities:**
 - Applicants in the South had **12% higher odds** of approval than those in the East.
 - West Coast applicants had **12% lower odds** than those in the East.
- **Racial disparities:**
 - White applicants had **157% higher odds** of approval.
 - Black applicants saw **34% higher odds**, and Asian applicants **105% higher**.
 - Applicants with “Free Form Text Only” race had **~25% lower odds** of approval.
- **Neighborhood effect:**
 - A 1% increase in minority population in a tract corresponded to a **0.3% decrease** in odds of approval.

These findings suggest systemic disparities even after accounting for financial indicators.

Project 4: Monte Carlo Simulation

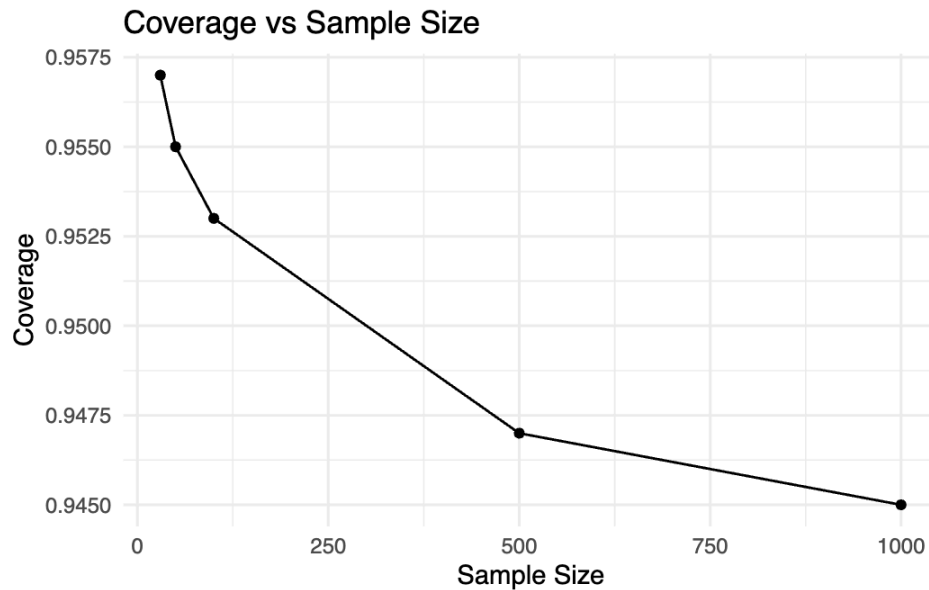
To assess the bias and confidence interval performance of our statistical approach, we conducted a Monte Carlo simulation.

Bias vs Sample Size



Caption: As sample size increases, bias approaches zero, validating the use of logistic regression in large-sample settings.

Coverage vs Sample Size



Caption: Coverage of the 95% confidence interval remains stable across sample sizes, generally close to the nominal level of 0.95.

Conclusion

This analysis reveals consistent patterns of disparity in mortgage lending approval. Even after adjusting for income and loan amounts, race and region significantly affect the likelihood of loan approval. The Monte Carlo simulation supports the validity of inference based on the logistic model. Future work should explore lender-level effects, credit history, and extend the analysis nationally.