Determination of Key Factors Influencing Insurance Claim Activity Using Multiple Regression: Applications to Sri Lanka and Beyond

The analysis uses regression models to identify key socio-economic and environmental factors influencing insurance claim activity, providing actionable insights for improving the Sri Lankan insurance sector and beyond.

December 20, 2024

Abstract

This study investigates the factors influencing insurance claim activity (INVOLACT) across 47 Chicago zip codes using multiple regression analysis. Key predictors—racial composition, fire rates, theft rates, housing age, income levels, and geographical side—are assessed for their impact on INVOLACT.

The analysis validates significant predictors, with race and fire rates showing a strong positive effect, theft rates a negative correlation, and housing age a moderate positive impact. Income levels were statistically insignificant and excluded in the final model. Regression assumptions, including normal residuals and homoscedasticity, are met, ensuring model validity.

This study demonstrates broader applicability to Sri Lanka's insurance market, offering insights for risk profiling, premium optimization, and policy equity. Additionally, it contributes to geographical and socio-economic research methodologies.

Introduction

Content

- **Study Context**
  - This study examines the key demographic, socio-economic, and environmental factors influencing insurance claim activity (INVOLACTT) to help insurance companies manage risk and pricing strategies effectively. It explores how variables such as race, fire and theft incidents, housing age, income level, and location impact claim patterns. By identifying the most significant predictors, insurers can enhance their underwriting processes, improve risk profiling, and tailor insurance offerings to better meet policyholder needs while minimizing financial risks. Statistical analysis and regression modeling are used to inform strategic decisions within the insurance sector.

- **Research Questions**
  - How do socio-economic factors influence INVOLACT?
  - What roles do fire and theft rates play?
  - How does housing age affect claims?
  - What is the impact of income levels and geographical location?

- **Objectives**
  - Identify significant predictors of INVOLACT.
  - Develop regression models with robust validation.
  - Apply findings to the Sri Lankan insurance sector.

Methodology

Research Approach

- Quantitative analysis using multiple regression to assess INVOLACT against six independent variables.

Regression Assumptions

- Residuals are normally distributed.
- Linear relationships exist between predictors and response
- No multicollinearity among predictors.
- Residuals exhibit homoscedasticity (constant variance).

Model Selection Criteria

- Adjusted R^2: Measures explanatory power while penalizing overfitting.
- F partial test (p-values: Variables with $p > 0.05$ are excluded.)

Data Set

- Source: Julian J. Faraway's Linear Models with R.
- Sample: 47 zip codes in Chicago.

Dependent Variable	Independent Variable
INVOLACT	Race
	Fire
	Theft
	Age
	Income
	Side

Table 1: Independent and Dependent Variables

Variable	Variable Type	Measurement Units
INVOLACT	Numeric	Percentage
Race	Numeric	Percentage
Fire	Numeric	Percentage
Theft	Numeric	Theft per 1000 population
Age	Numeric	Percentage
Income	Numeric	Median family income in $1000
Side	Factor	n-: North , s-: South

Table 2: Data Dictionary Table

- **Regression Models**
  - Model 01: Includes all predictors.
  - Model 02: Excludes “Side” (p > 0.05).
  - Model 03 (Final): Excludes “Income” (p > 0.05).

Result

- **Model Comparison**

	Number of Variables	Variables	R^2
Model 01	4	RA, FI, TH, AGE	0.747191
Model 02	5	RA, FI, TH, AGE, IN	0.750822
Model 03	6	RA, FI, TH, AGE, IN, SI	0.751052

Table 3: Adjusted R^2 values of main 3 models

Independent Variables

- Race(RA)
- Fire(FI)
- Theft(TH)
- Age(AGE)
- Income(IN)
- Side Numeric(SI)

Dependent Variable

– Involact(INV)

- **Significant Predictors (Model 03)**

    - Race (β = 0.0081, p < 0.0001): Strong positive effect.
    - Fire Rates (β = 0.0366, p < 0.0001): Strong positive effect.
    - Theft Rates (β = -0.0096, p = 0.0024): Negative impact.
    - Housing Age (β = 0.0072, p = 0.0046): Moderate positive effect.

Final Model is,

$$\mathrm{INV} = \beta_0 + \beta_1 RA + \beta_2 FI + \beta_3 TH + \beta_4 AGE + \epsilon$$

Discussion and Applications

- **Discussion**
  - Race and fire rates are dominant factors in INVOLACT.
  - Theft rates have a negative effect, possibly reflecting policyholder behavior.
  - Older housing correlates with increased claims due to higher risk exposure.

- **Applications to Sri Lanka**

  Predictive modeling for disaster-prone or high-density areas enables insurers to assess risks more accurately and allocate resources effectively. This supports risk-adjusted premium pricing strategies, ensuring premiums reflect exposure while maintaining financial sustainability. Additionally, addressing demographic disparities in insurance access fosters equitable solutions, providing underserved populations with tailored coverage to meet their unique needs.

Conclusion

- **Summary of Findings**

  Race and fire rates are significant factors influencing insurance claims, with theft rates showing a negative correlation and housing age playing a moderate role. Interestingly, income levels are statistically insignificant in predicting INVOLACT, highlighting that demographic and environmental factors outweigh economic status in this context.

- **Implications**

  - For Insurers: Enhanced risk profiling and premium optimization.
  - For Policymakers: Data-driven strategies for equitable policy access.
  - For Researchers: Extend regression modeling to diverse fields.

END