Predicting Duration of Insurance Complaints

Anthony J. Mercure and Dr. Khyam Paneru*

Data Science & Mathematics with Computer Science Programs, Department of Mathematics

Abstract

Insurance disputes are a notable area of concern, affecting both consumers and the regulatory bodies tasked with enforcing varying insurance industries' compliance. This investigation presents a regression analysis of complaint duration in insurance sectors, leveraging complaint data from the Texas Department of Insurance (TDI) compiled over the last 12 years. Our study includes records of complaints lodged against insurance entities, including licensed individuals and companies. Our analysis is framed around understanding factors influencing complaint resolution times, aiming to explore potential predictors, including insurance coverage types, respondent types, and previous complaint histories. We employ multiple linear regression to examine how these factors influence the duration of complaint resolution. To enhance model reliability, diagnostic tests were conducted, and transformations on some variables were applied, with some outliers and influential observations removed to ensure more robust predictive capabilities. This analysis highlights the complex dynamics within insurance complaint resolution processes, suggesting that targeted regulatory measures could improve response efficiency. This examination contributes to the field of public policy and consumer protection, offering data-driven recommendations to refine complaint management strategies and enhance consumer protection and satisfaction within the insurance industry.

Dataset Information

Scope: The dataset includes **258,229 complaints** filed against **5,286 unique entities**, comprising licensed individuals and companies. Of these, **202,833 complaints** were not found in error, while **38,673** were confirmed as errors.

Key Variables:

Variable	Description
Complaint Duration	Dependent variable, calculated from received and closed dates.
Coverage Type	Categories include Accident & Health, Life & Annuity, Automobile, and others.
Respondent Type	Organization vs. Individual.
Confirmed Complaint	"Yes" indicates TDI confirmed the licensed person or organization was in error.
Time Between Complaints	Interval between successive complaints for the same entity.
Number of Previous Complaints	Cumulative count of prior complaints.

Data Preprocessing:

- **Exclusion of Zero-Duration Complaints:** Complaints with zero duration were excluded as invalid (e.g., immediately terminated complaints).
- **Entity Filtering:** Entities with fewer than two complaints were removed to focus on those with significant complaint histories.
- **Coverage Type Selection:** Retained the most common types (**Life & Annuity**, **Accident & Health**, and **Automobile**) for analysis.

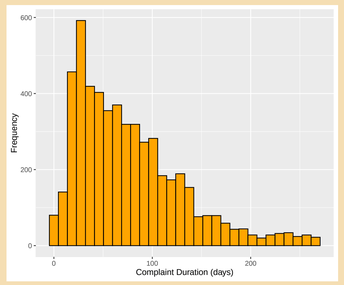

Figure 1. Histogram of Complaint Duration

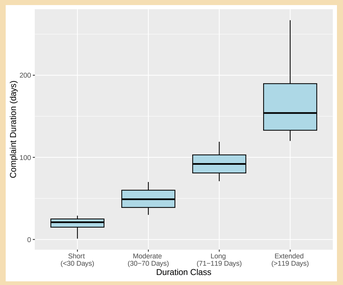

Figure 2. Boxplots of Complaint Duration

Data Preparation

- The categorical variables *Respondent Type*, *Coverage Type*, and *Confirmed Complaint* were appropriately factorized to enhance interpretability.
- **Sampling:** A **4% random sample** (**7,940 observations**) was used to ensure computational efficiency while maintaining representativeness.
- **Outlier Removal:** Using the IQR method, **2,742 outliers** were removed across key quantitative variables to improve model assumptions.
- **Influential Points Removal:** Cook's Distance and leverage values were used to identify **914 influential points**. Observations with high leverage or significant Cook's Distance values were excluded to prevent distortion of model coefficients. Removing these points reduced model bias and improved the overall goodness-of-fit.
- A **Box-Cox transformation** was applied, identifying an optimal $\lambda = 0.1414$.
  - The variables **Complaint Duration** and **Time Between Complaints** were log-transformed to stabilize constant variance and improve linearity issues found in the initial diagnostics.

Regression Results

Coefficient	Estimate	p-value
Intercept	4.333	$<0.001^{***}$
ln(Time Between Complaints)	0.001	0.881
Number of Previous Complaints	-0.00002	0.008^{**}
Respondent Type (Org.)	-0.139	0.651
Confirmed Complaint (No)	-0.225	$<0.001^{***}$
Coverage Type (Accident & Health)	0.006	0.812

Table 1. Regression Model Summary

- **Time Between Complaints**: Not statistically significant; No evidence of an association between the time between complaints and the duration.
- **Number of Previous Complaints**: Statistically significant negative effect; More previous complaints are associated with slightly shorter complaint durations.
- **Respondent Type (Individual)**: Set to zero as it represents a baseline comparison against organizations.
- **Respondent Type (Organization)**: Not statistically significant; No difference in complaint resolution times between organizations and individuals.
- **Complaint Confirmation (Yes)**: Set to zero to assess the impact of confirmed complaints with entities not in error.
- **Confirmed Complaint (No)**: Highly significant negative effect; Complaints that are not confirmed are resolved significantly faster than confirmed complaints.
- **Coverage Type (Life & Annuity)**: Set to zero as the baseline to compare other coverage categories.
- **Coverage Type (Accident & Health)**: Not statistically significant; No difference in complaint durations between Accident & Health and Life & Annuity coverage types.
- **Coverage (Automobile)**: excluded due to high multicollinearity with *CvAH*, potentially distorting the model's coefficient estimates and interpretability.

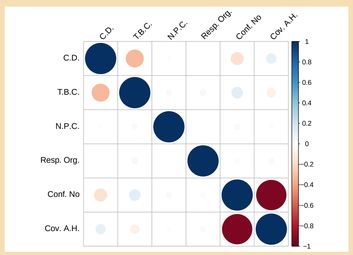

Figure 3. Correlation Matrix

Variable	Lower 5%	Upper 5%
Intercept	3.728	4.938
T.b.C.	-0.006	0.007
N.P.C.	0.000005	0.00004
Resp. Org.	-0.742	0.463
Conf. No	-0.290	-0.153
Cov. A.H.	-0.045	0.058

Figure 4. 95% Confidence Intervals for Regression Coefficients

Diagnostic Analysis

Diagnostic tests were conducted to ensure that the regression model satisfies fundamental assumptions, which are critical for accurate predictions and valid inferences. These tests evaluated residual normality, variance constancy, multicollinearity, and influential observations.

- **Residuals vs. Fitted Values:** Highlighted potential non-linear patterns.
- **Q-Q Plot:** Assessed the normality of residuals and identified deviations.
- **Scale-Location Plot:** Examined the spread of residuals across fitted values, indicating variance constancy.
- **Residuals vs. Leverage Plot:** Identified influential points and their impact on the model.

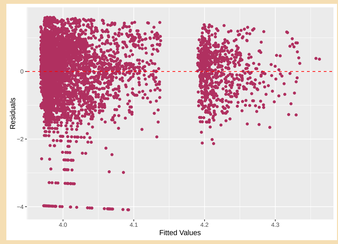

Figure 5. Residuals vs. Fitted Plot

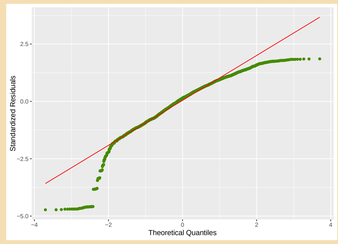

Figure 6. Normal Q-Q Plot

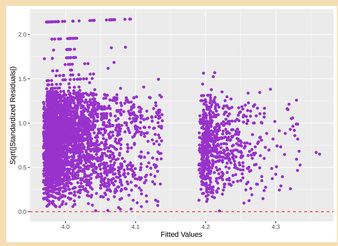

Figure 7. Scale Location Plot

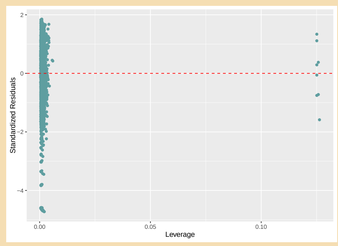

Figure 8. Residuals vs. Leverage Plot

Conclusion

Overall, the regression analysis of insurance complaints in Texas has provided significant insights into the factors influencing complaint resolution durations. For instance, the suggestion that entities with more prior complaints resolve subsequent complaints slightly faster, potentially due to improved procedural efficiencies. Similarly, when complaints were not confirmed as the entity's error, resolution times were significantly shorter, reflecting the streamlined handling of such cases. Despite the improvements made through transformations and diagnostic refinements, some issues related to the linearity of predictors and the overall model fit remain evident in the regression diagnostics. These challenges suggest that future work should focus on exploring advanced modeling approaches, such as machine learning algorithms, which may provide a more robust fit for complex patterns within the data. Addressing these issues is essential for further improving the model's predictive accuracy and enhancing its applicability for informing regulatory and operational strategies in the insurance sector. The results emphasize the necessity of tailored policy adjustments to enhance operational efficiency and consumer satisfaction within the insurance sector, while highlighting the potential for continuous improvement in the analytical framework supporting these insights.

References

Texas Department of Insurance. Insurance complaints: All data, 2024. URL `https://catalog.data.gov/dataset/insurance-complaints-all-data`. Accessed: October 21, 2024.