Summary on a Page

Summary of scenario

In our global industrial company, over 10,000 employees work across multiple factories worldwide. We have operations in Asia, Europe, and South America, each regulated by its own industrial standards and safety practices. A recent incident in South America has raised concerns about the effectiveness of our company's workplace practices.

Research questions

It is necessary to perform an analysis of workplace injury data to answer the following questions:

1. Of the various safety regimes in place across the company, which one would be recommended to become the international standard for our company, based solely on injury prevention performance?
2. It has been suggested by senior management that industry experience is more important than the safety regime when it comes to preventing injuries. The idea is that a policy should be developed that is directly related to lowering employee turnover will reduce injury rates. Do the available data support this assertion?
3. If there is any relationship between:

- Injuries and the annual bonuses a proportion of employees receive.
- Injuries and whether staff have received any formal external qualifications, e.g., external safety training or a university degree.

Summary of available data

The analysis uses data from 'injury.csv', which includes counts of injuries and hours worked, aggregated by workers' experience levels and the workplace safety regime at their factory. It also provides the proportions of groups who received an annual bonus last year, completed external safety training, or have at least one university degree. The data covers the last 12 months of operation.

Modelling approach and justification

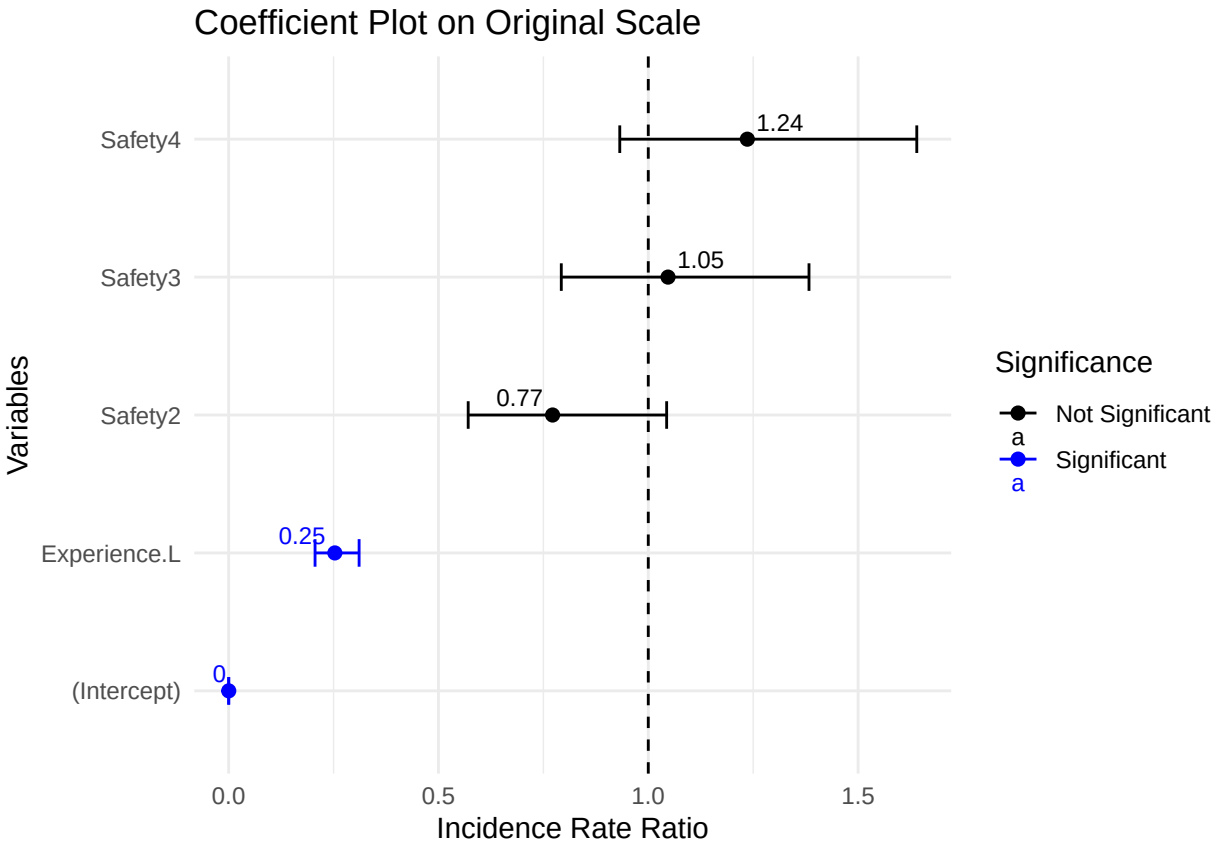
To address the questions, two models were considered: Poisson GLM and Negative Binomial GLM. Initially, both models used 'Injuries' as the target variable, with 'safety,' 'experience,' 'bonus,' 'training,' and 'university' as covariates, and 'Hours' as an offset variable. Stepwise AIC-based selection was applied to choose the variables to include. Model selection was based on comparing AIC, examining residuals, and assessing overdispersion.

Validity of model and modelling results

As a result, the Negative Binomial GLM showed a better fit to the data. Based on the final model, the number of injuries is influenced only by the experience level of the group and the safety regime (see plot). In this model, Safety Regime 1 is chosen as the baseline. Bonus, training, and having a university degree did not enter the model as they do not have a significant effect.

1. The experience level of the group has a significant effect on the number of injuries with 95% confidence. The exponentiated coefficient is 0.25, indicating a 75% decrease in the number of injuries for each increase in experience level (from 1 to 4).

2. Safety Regime 2 is marginally significant, with about 90.4% confidence. The exponentiated coefficient is 0.77, meaning there is a 23% decrease in the number of injuries compared to the baseline regime (Regime 1).Other safety regimes are not statistically significant.



Answers to research questions and recommendations

1. The analysis indicates that Safety Regime 2 results in fewer injuries than other regimes, but this effect is only marginally significant (about 90.4% confidence). Therefore, it cannot be definitively recommended as the international standard for our company based on the current data. However, given the borderline significance, further data collection and analysis are recommended.
2. The analysis strongly supports that industry experience is more important than the safety regime in preventing injuries. Experience has a significant and greater impact on reducing injuries with 95% confidence, while safety regimes do not show significant effects. Thus, the data suggest that reducing employee turnover could effectively lower injury rates.
3. Variables related to bonuses, safety training, and university degrees were not included in our final model, indicating they do not significantly impact the number of injuries.

These conclusions are based on the current data and analysis. It is advisable to continue gathering data across different periods and locations or to analyze additional variables that may influence injury rates (e.g., location, age, occupation).