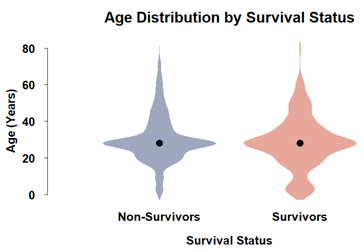
Influences on Titanic Survival

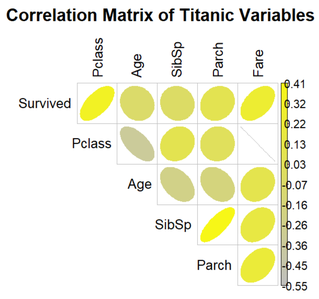
Alexis Pastas

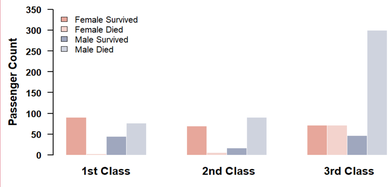
Introduction

The Titanic disaster is one of the most studied events in history. Predicting **survival** during the incident helps understand the factors that influenced passenger outcomes. By analyzing data on **class**, **age**, **gender**, and other variables, we try to identify key predictors of survival and assess how logistic regression can model these outcomes effectively.

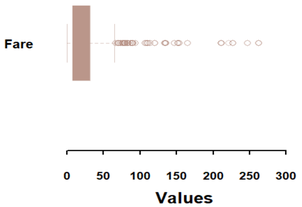
QA / EDA

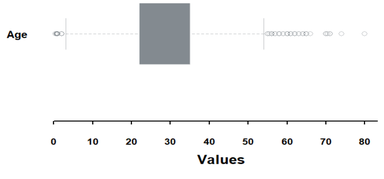


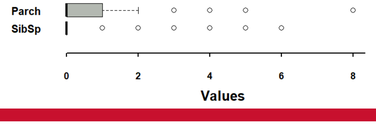




Survival rates varied significantly by gender/class



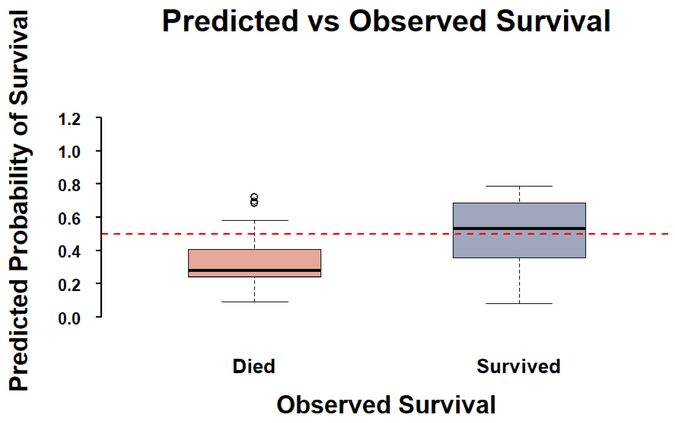




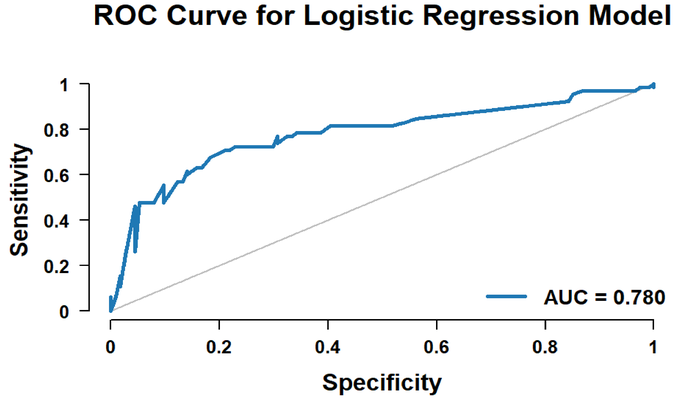
Analysis / Results

Died group (red): Most predicted survival probabilities are below 0.5

Survived group (blue): Predicted probabilities are mainly above 0.5



Variables used for Model: Age and Pclass



Discussion

The analysis showed that **Pclass**, **age**, and **fare** were key predictors of survival.

Pclass: First-class passengers had better access to lifeboats, leading to higher survival rates. Third-class passengers faced more challenges.

Age: Younger passengers survived more often, likely due to rescue priorities, while older passengers had lower survival chances.

Fare: Higher fares were linked to first-class tickets, improving survival odds.

Analysis and Results: The Predicted vs Observed Survival graph demonstrated that the model effectively separates survivors from non-survivors, though some misclassifications remain, particularly for lower probabilities in the “Died” group. Lastly, the ROC curve produced an AUC of 0.787, confirming the model performs well overall, with good sensitivity and specificity.

Conclusion: The results highlight that survival on the Titanic was heavily influenced by passenger class, age, and fare. More improvement can be done to tell the story of the Titanic.

References

https://www.kaggle.com/datasets/yasserh/titanic-dataset

File: Data-Visualization---Assignment-3.html