Alumni Donation Case Study

Introduction

The goal of this report is to explore factors influencing alumni donation rates and develop a predictive linear regression model using available data, while focusing on identifying the “best” model and addressing any violations of model assumptions to improve predictive performance and interpretability.

The dataset contains information about 49 universities, with an emphasis on significant characteristics that may influence the primary variable of interest, alumni giving rates. The dataset consists of the following variables:

* **School**: The name of the university.
* **Percent of Classes Under 20**: A numerical variable representing the percentage of classes with fewer than 20 students.
* **Student-Faculty Ratio**: A numerical variable representing the ratio of students to faculty.
* **Alumni Giving Rate** A numerical variable representing the percentage of alumni who donate to the university.
* **Private**: A binary indicator variable where 1 denotes private institutions and 0 denotes public institutions.

We ran a series of linear regression studies to look at the association between alumni donating rates and institutional features. Exploratory data analysis demonstrated a negative correlation between the student-faculty ratio and the alumni donating rate, while other predictors had weaker associations.

To reduce skewness and increase model fit, the response variable was transformed logarithmically. Stepwise selection with BIC revealed that the student-faculty ratio was the most important predictor. Diagnostic plots revealed no significant violations of regression assumptions in the final model. The resulting model has a mean squared error (MSE) of about 0.11, indicating high predicted accuracy.

Key findings indicate that universities with lower student-faculty ratios have greater alumni giving rates, although private status and small class percentages had no impact.

Data Description

The dataset used in this analysis included 49 universities and significant institutional characteristics such as student-faculty ratios, class size percentages, and private status. A summary of the variables provides the following insights:

* **Alumni Giving Rate**: The response variable ranges from 7% to 67%, with a mean of 29.27%, reflecting wide variability in alumni donation behaviors across institutions.
* **Student-Faculty Ratio**: Ranges from 3 to 23, indicating significant differences in class size experiences across schools.
* **Percent of Classes Under 20**: Varies between 29% and 77%, highlighting variation in the prevalence of smaller class sizes.
* **Private Status**: Approximately 69% of the universities in the dataset is private institutions.

Visual analysis using histograms, scatter plots, and box graphs showed essential patterns. The alumni giving rate had a right-skewed distribution, and scatter plots revealed a strong negative association between the student-faculty ratio and the alumni giving rate. Box plots revealed that private institutions have higher donation rates than public organizations.

A graph of a column

Description automatically generated with medium confidence

A graph showing a number of numbers and a number of columns

Description automatically generated with medium confidenceA graph with pink dots

Description automatically generated

Methods

To model alumni giving rates, we began with exploratory data analysis and identified student faculty ratio, percent of classes under 20, and private as potential predictors. A stepwise forward selection process, using the BIC, was performed to select the best model. We started with a base model containing only the intercept and iteratively added predictors, including interaction and quadratic terms, to find the model with the lowest BIC.

The response variable (alumni giving rate) was log-transformed to address skewness and heteroscedasticity, improving model interpretability and residual diagnostics. The final model included student faculty ratio and private as predictors. Diagnostic plots were generated to confirm the model assumptions (linearity, homoscedasticity, and normality of residuals). The model was evaluated on a test set using the mean squared error (MSE) as the performance metric.

Results

The stepwise forward selection process identified the best model as:

**log(alumni giving rate) = 3.533 + 0.41(private} - 0.052(student faculty ratio).**

Key Findings:

1. Coefficients:

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

* + The coefficient for private ( 0.407) indicates that private institutions have, on average, e^0.407 ≈ 1.50 times higher alumni giving rate than non-private institutions, holding student faculty ratio constant.
  + The coefficient for student faculty ratio (−0.055) suggests that a one-unit increase in the student-faculty ratio reduces the alumni giving rate by approximately 5.5%, on average, holding private constant.

1. Model Diagnostics:
   * R2=0.6383, adjusted R2=0.6222: The model explains approximately 63.8% of the variance in the log-transformed alumni giving rates.
   * Residual standard error: 0.328, indicating the typical deviation of log-transformed predictions from observed values.

A graph with red line and black dots

Description automatically generated

A graph of a student and faculty ratio

Description automatically generated

1. Model Evaluation:
   * The residuals showed no patterns or correlations, confirming that the model assumptions were met.
   * The test set MSE was 0.13, indicating good predictive performance.
2. Other Variables:
   * Adding percent\_of\_classes\_under\_20 or additional interaction terms did not improve the model and were excluded due to lack of significance.

Discussion

The final model highlights that student faculty ratio and private are the most significant predictors of alumni giving rates. Private schools generally have higher alumni giving rates, likely due to stronger alumni networks or more resources dedicated to engagement. A lower student-faculty ratio also positively impacts giving rates, potentially reflecting stronger student-teacher connections that improve alumni loyalty.

Future Study:

* Further research could explore additional predictors, such as regional demographics or institutional financial support, to improve the model.
* Interaction effects, such as the combined influence of private and student faculty ratio, could be examined in greater depth.

Limitations:

* The dataset is limited in size, which might affect the model's generalizability.
* The log-transformation simplifies interpretability but may not capture all nuances of the alumni giving behavior.

This analysis provides a robust framework for understanding and predicting alumni giving rates, offering actionable insights for improving alumni engagement strategies.