Global Higher Education Competitiveness: Analysis of Academic Reputation and Its Influencing Factors

Hexiang Li, Fuxian Xiao, Yuchen Lin, Tommy Cheng, Yang Sun

Contributions:

Yuchen Lin: Responsible for data cleaning, constructing the MLR model, and formatting the final report.

Hexiang Li: Conducted the residual analysis and summarized the key results.

Tommy Cheng: responsible for comparing the model to the literature.

Yang Sun: Writes the data description and draws the tables, histograms and scatterplots.

Fuxian Xiao: Responsible for the introduction, literature review, hypothesis formulation, and team agreement drafting.

Introductions(364 words):

Academic reputation has emerged as a critical criterion in the assessment of global universities. It influences institutional prestige, faculty recruitment, student enrollment, research initiatives, and global visibility (Hazelkorn, 2015). QS World University Rankings 2025 assigns each university an Academic Reputation Score (ARS) based on extensive academic surveys. Both the public and academic communities remain unclear about the specific institutional factors that affect academic reputation. Institutions have to grasp these dynamics in a competitive global education scene if they are to strategically place themselves. This project explores which institutional features most strongly relate to academic reputation, using data from QS's 2025 rankings, published on Kaggle (Monfared, 2025). This dataset covers more than 10 variables describing research output, internationalization, teaching resources, employment outcomes, and sustainability performance, together with over 1,500 universities. For linear models, the academic reputation score is a continuous and ratio-scaled variable from 1 to 100. Citations per Faculty, Employer Reputation, Faculty-Student Ratio, Sustainability, Employment Outcomes, and Institution Size Classification include both continuous and categorical predictors. Before regression modelling, missing values and standardization of variables were addressed to ensure model suitability.

The study uses multiple linear regression to analyze the relationship between institutional features and academic reputation. By controlling for other variables, MLR allows us to estimate the marginal effect of each predictor, ensuring interpretability. The robustness of the model will be assessed with assumption checks and residual diagnostics. The objective is to not only fit a model but also to identify actionable levers for academic institutions by interpreting the structure of the regression coefficients.

This research extends prior studies on global rankings. Research visibility, international collaboration, and global engagement are significant factors influencing academic reputation, as indicated by previous studies (Shin & Toutkoushian, 2011). Hazelkorn posits that universities have progressively adjusted their practices to conform to ranking systems, notably by enhancing criteria associated with research output and employability(Hazelkorn, 2015). We thus hypothesize that Academic Reputation Scores are statistically correlated with several institutional factors, including research output, internationalization, faculty resources, and employability, among other institutional indicators.

Unlike QS's subjective assessment system, our framework is evidence-based and replicable. Interpreting the regression coefficients may offer strategic advice for institutions seeking global competitiveness and academic status.

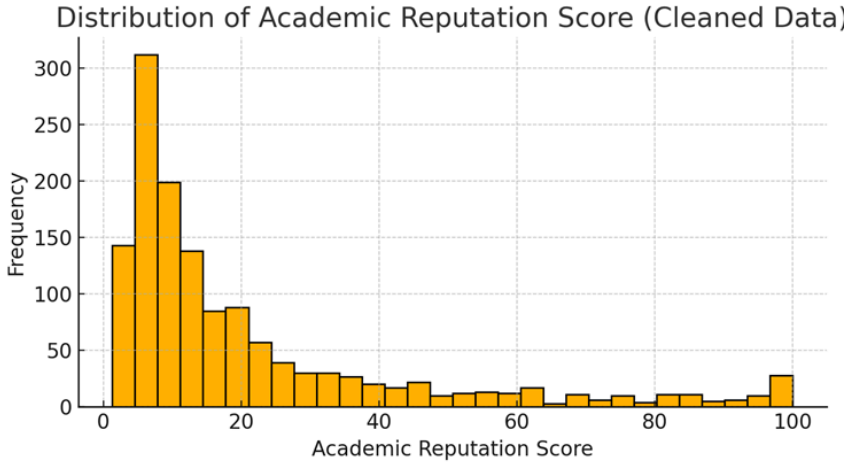
Data Descriptions（311 words):

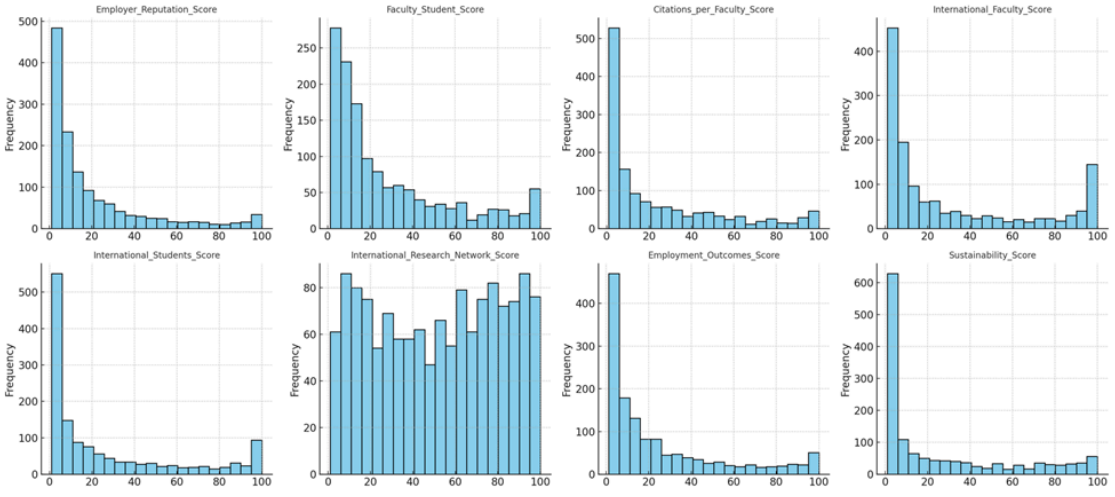
We found this dataset on Kaggle and QS.com (Monfared, 2025). Furthermore, the QS World University Rankings 2025 were compiled through a data collection process that combined institutional submissions, large-scale global surveys.

According to QS Quacquarelli Symonds (2025a), universities submitted detailed student, staff, and institutional data through QS platforms like MovelN. QS also surveyed over 151,000 institutions to assess academic reputation through extensive academic and employer opinion surveys. (QS Quacquarelli Symonds, 2025b). The usage of our original dataset is to help students, universities, and policymakers compare global institutions, which is closely related to our research direction.

Our response variable is Academic_Reputation_Score. This variable can be treated as a continuous and ratio scale variable, ranging from 0 to 100. It measures an institution's academic standing in the international higher education landscape. Because this variable is continuous and ratio-scaled, Academic_Reputation_Score is suitable as a response variable in a linear model.

The histograms show most variables are right-skewed, indicating many universities have lower scores in several areas. But the histogram of International_Research_Network is unique and it is an even spread.





Below are the tables for both 8 predictors and the only categorical variable "SIZE", indicating some relevant statistics such as mean and standard deviation.

Numerical Variable Statistics

Variable	Mean	SD	Min	25th Quantile	75th Quantile	Max
Employer_Reputation_Score	20.892	24.423	1.100	4.500	27.700	100.00
Faculty_Student_Score	28.359	27.579	1.200	7.475	41.150	100.00
Citations_per_Faculty_Score	24.600	28.236	1.000	3.000	39.375	100.00
International_Faculty_Score	31.046	34.520	1.000	4.275	53.100	100.00
International_Students_Score	26.391	31.488	1.000	3.100	40.900	100.00
International_Research_Network_Score	51.641	29.605	1.000	24.500	78.200	100.00
Employment_Outcomes_Score	24.694	27.732	1.200	4.200	35.125	100.00
Sustainability_Score	25.913	31.646	1.000	1.500	44.600	100.00

SIZE Variable Frequency

Category	Count
L	639
M	337
XL	324
S	76

Back to our research question: “which factor is most influential to the response variable Academic_Reputation_Score?” After analyzing the histogram and statistical summary, International_Research_Network_Score is a strong candidate in our analysis. It measures a university’s level of international collaboration. Additionally, the mean of this variable, which is 51.641, is significantly higher than that of other variables, and its large variance suggests it has strong potential to explain differences in Academic_Reputation_Score.

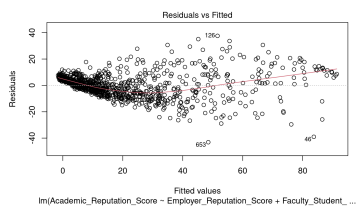
Below are scatterplots for the eight predictors. Many, especially reputation scores, show linearity with Academic_Reputation_Score, supporting their selection.

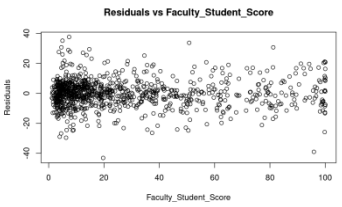
In our cleaning process to the dataset, we removed 127 rows and 17 columns with missing values, and set 1,316 missing entries to 0.

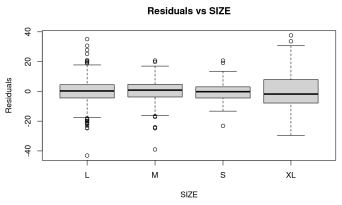
Preliminary Results(309 words):

We fitted a multiple linear regression with nine predictors which includes eight numerical predictors and one categorical predictor (SIZE). The model achieved adjusted R-squared of 0.8362, a multiple R-squared of 0.8381 and a residual standard error of 9.248 on 951 degrees of freedom. These values suggest a good fit in terms of explained variance and prediction accuracy. However, further diagnostics are required to check the assumptions of linear regressions.

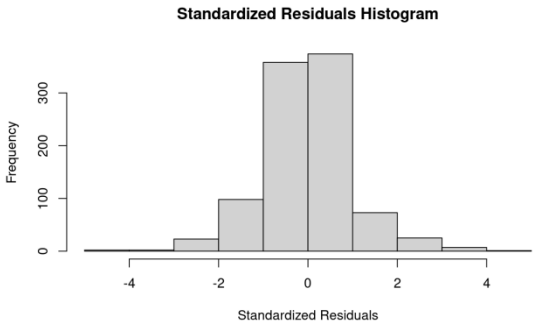
To evaluate the linear regression assumptions, we performed a rigorous residual analysis. The **Residuals vs Fitted values** plot shows a curvature in the red line and a cone-shaped spread, suggesting violations of linearity and homoscedasticity. To identify suspicious contributing predictors, we examined **Residuals vs Faculty_Student_Score** plot, which shows randomness and roughly constant spread, indicating it meets two of the assumptions. In contrast, **Residuals vs SIZE** plot shows larger variability and more outliers in group “XL”, indicating potential violation of homoscedasticity.

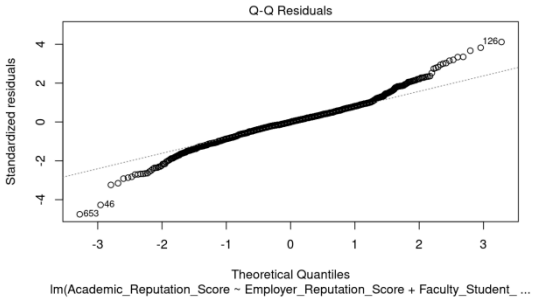






For the normality assumption, both histogram and QQ-plot of standardized residuals shows a rougly symmetric shape with a heavier right tail, suggesting slight non-normality. For the independence of the residuals, we calculated the residual covariance matrix which was not zero. Therefore, there might be some correlations between some of the residuals, potentially due to omitted variables.





The main conclusions in the literature about the factors influencing academic reputation are generally supported by the results of our early model. Visibility, international collaboration, and global engagement are significant factors influencing academic reputation. The idea is supported by these predictors' significant statistical significance and high adjusted R-square value of 0.8362. Furthermore, Faculty_Student_Score has little explanatory power for Academic Reputation Score, although meeting the requirements of linear regression by displaying linearity and constant variance in residual analysis. This implies that indicators pertaining to teaching are mainly undetectable in evaluations of global reputation, despite being internally significant for the student experience.

Bibliography:

Monfared, M. (2025). *QS World University Rankings 2025* [Data set]. Kaggle. https://www.kaggle.com/datasets/melissamonfared/qs-world-university-rankings-2025

QS Quacquarelli Symonds. (2025a). *Data collection process for QS World University Rankings 2025* [Institutional Data Submissions]. Retrieved May 16, 2025, from https://www.qs.com/data-collection-qs-rankings

QS Quacquarelli Symonds. (2025b). *Data collection process for QS World University Rankings 2025* [Academic and Employer Surveys]. Retrieved May 16, 2025, from https://support.qs.com/hc/en-gb/articles/4405955370898-QS-World-University-Rankings

Dill, D. D., & Soo, M. (2005). Academic Quality, league tables, and public policy: A cross-national analysis of University Ranking Systems. *Higher Education*. https://doi.org/10.1007/s10734-004-1746-8

Hazelkorn, E. (2015). *Rankings and the Reshaping of Higher Education*. https://doi.org/10.1057/9781137446671

Shin, J. C., & Toutkoushian, R. K. (2011b). The past, present, and future of University Rankings. *University Rankings*, 1–16. https://doi.org/10.1007/978-94-007-1116-7_1