# Exploring the Multivariate Relationship between University Characteristics and Teaching Scores

December 20, 2023

This paper investigates the association between various academic and international attributes of universities and their respective Teaching Scores, using data from the 2023 QS World University Rankings. The study addresses a research gap by simultaneously considering predictors such as 'Location,' 'No. of Students per Staff', 'International Students', 'Research Score', 'Citations Score', and 'Industry Income Score'. Drawing on existing literature, the study examines the impact of these factors on teaching quality.

Transformations and model selection techniques were used to improve the predictors for statistical analysis and inference. Unexpected findings include a positive coefficient for 'Students per Staff,' challenging conventional wisdom on student-to-staff ratios. The negative coefficient for 'Research Score' aligns with the hypothesis that research-focused institutions may allocate fewer resources to teaching. Regional analyses reveal varying impacts, with 'East Asia & Pacific' outperforming other regions in teaching quality. The study acknowledges limitations and suggests future research avenues. The model achieves an adjusted $R^2$ of approximately 0.7234, affirming the feasibility of predicting teaching scores based on selected predictors.

## Background and Significance

We explore the question of whether the overall academic and international qualities of a university can be used to estimate its 'Teaching Score'. Linear regression is a useful tool to identify the strength and direction of these relationships.The dataset we are using to investigate this relationship is from the 2023 QS World University Rankings.

The basis for which we believe these predictors will have a measurable impact on 'Teaching score' stems from peer-reviewed academic research articles. Benito et al. [1] provide insight as to the significance of the location of a university on the quality of education. This motivates the categorical variable of 'Location'. Ramirez et al. [2] explore the relationship between research and teaching at academic institutions, motivating the Research and Citation numerical variables. Lastly, See et al. [3] demonstrate that teaching can be affected by a range of factors such as location and industry. This motivates the use of a linear regression model to discern the multivariate relationship involved in teaching.

Assessing the multivariate impact of this set of predictors has not been seen in the literature, and we feel that the results of this analysis can motivate a better understanding of what results in higher teaching quality in universities. The aspect of the fitted model that would give us the answer is the coefficients and the statistical significance of the relationships found.

## Methods

### Data Collection

An initial linear regression model was constructed using the variables relevant to the research question, using data taken from the "QS World University Rankings" dataset. These rankings are assembled based on expert opinion and thorough literature review. This data is reliable as it is comprehensive and approved by the International Ranking Expert Group. Numerical variables were treated as is, and categorical variables were added using one-hot-encoding.

### Variable Creation

The variables in our analysis involve 'Teaching Score' (0-100 rating of the learning environment), using the predictor variables 'Location' (country), 'No. of Students per Staff' (numerical ratio), 'International Students' (fraction of total student population), 'Research Score' (0-100 rating based on research volume/income), 'Citations Score' (0-100 rating based on research quality/influence), 'Industry Income Score' (0-100 rating based on income/patents), and 'International Outlook Score' (0-100 rating based on international collaboration).

The 'Location' score was transformed to a "Region" identifier- the locations were individual countries and this prevented us from interpreting the analysis, with ~1700 datapoints. We transformed the countries to their geographical regions, to capture the information in a less granular way.

### Analytic Methods

We test linearity, constant variance, and error correlation between predictors using pairwise scatter plots of all the variables from the model and against residuals of the model. We used Q-Q plots on the residuals to assert normality, applying approximate Boxcox transformations. We also considered various forms of problematic points such as leverage points, outlier points, and influential observations. We noted, but did not remove these points since this would diminish the predictive capacity of our model. Then, we manually used partial F-test and adjusted $R^2$ to verify the optimal set of statistically significant predictors.

## Results

The final model relates five predictors: ‘Region’ (transformed from ‘Location’ which named every country, making interpretation difficult), log-transformed ‘No. of Students per Staff’, the $-8$ power transform of ‘International Students’, log-transformed ‘Research Score’, and the 0.5 power transform of ‘Citations Score’, predicting the $-0.5$ power transform of ‘Teaching Score’. Summaries of the numerical variable distributions are below:

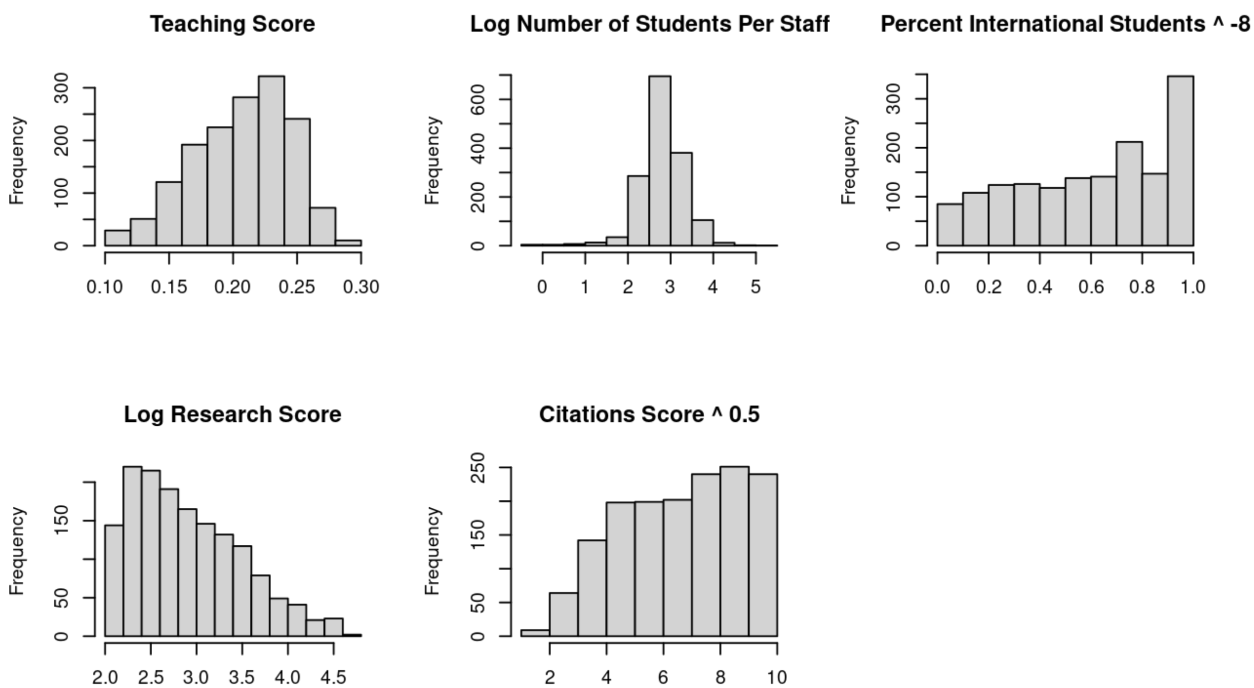


Figure 1: Predictor and Response Numerical Summaries

The transformations improve normality and ensured constant error variance- visually, they correct the distributional skews. We removed the redundant “International Outlook Score”, which had high collinearity with “International Students”. We reject the null hypothesis of the ANOVA test for overall significance given the p-value of less than $2.2e-16$, meaning the model contains a significant relationship with at least one predictor. There are a large number of problematic observations, that exceed cutoffs for leverage, outliers, Cook’s Distance, DFFITS, and DFBETAS. The coefficients of our final model are given below, all non-region coefficients are statistically significant:

Table 1: Final Model Coefficients

| Coefficient | Value |
|---|---:|
| Intercept | 29.9079219 |
| Region Europe & Central Asia | -0.1577419 |

| Coefficient | Value |
|---|---|
| Region Latin America & Caribbean | -0.4498322 |
| Region Middle East & North Africa | -1.8773786 |
| Region North America | -1.6551098 |
| Region South Asia | -2.2746432 |
| Region Sub-Saharan Africa | -0.5607930 |
| Log Number of Students Per Staff | 2.0369901 |
| Percent International Students ˆ -8 | 0.4644004 |
| Log Research Score | -5.1417150 |
| Citations Score ˆ 0.5 | 0.0639710 |

## Discussion

The model implies an association between a university's characteristics and its teaching ability, specifically its students to staff ratio, research score, citations score, region, and proportion of international students.

First, the coefficient for students per staff is positive, meaning with all other predictors constant, more students per staff improve teaching. However in [2], a lower student-to-staff ratio increases teaching scores. We hypothesize that the size of the institution (which we don't have a predictor for) is a confounding factor - larger, more established institutions tend to have more students and better teaching.

Secondly, there is a negative coefficient for 'Research Score'. This makes intuitive sense, as institutions more devoted to research would have fewer resources devoted to teaching. This contradicts [2], where there is no significant relationship between research productivity of professors and the number of classes taught. We hypothesize this contradiction may be in the difference between the number of classes taught (predictor in the study), and teaching quality (predictor in our data). Since classes taught does not directly translate to teaching quality.

Third, the seven regions ordered from best average teaching to worst are: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Sub-Saharan Africa, North America, Middle East & North Africa, South Asia. The high ranking of the "East Asia & Pacific" region agrees with See et al. [3]. The rankings of "Latin America & Caribbean" and "Sub-Saharan Africa" as higher than "North America" does not agree with that paper, which may be a model limitation due to the low number of data points ( 34).

Last, the "Industry Income" factor did not have a statistically significant influence on teaching score given the other predictors and was removed. This makes sense given that the impact of "Industry Income" can be attributed indirectly through the other predictors.

Future directions of this work could be assessing the confounding effect of a university's size on the data and a more detailed analysis of how various regions and countries impact teaching quality, as we found that our limited dataset size did not lend well to studying these effects.

In conclusion, we address our original research question of whether can we create an accurate predictive model for 'Teaching Score'. It is important that we keep in mind our model's limitations- the high number of problematic observations may have affected coefficient estimates. However, we can still tentatively answer yes to our original research question, since our model achieves an adjusted $R^2$ of $\approx 0.7234$.

## References


[1] Benito M, Gil P, Romera R. 2020. Evaluating the influence of country characteristics on the Higher Education System Rankings' progress. J Informetr. 14(3):101051. doi:10.1016/j.joi.2020.101051. http://dx.doi.org/10.1016/j.joi.2020.101051.

[2] Ramirez-Montoya MS, Ceballos HG, Martínez-Pérez S, Romero-Rodríguez LM. 2023. Impact of teaching workload on scientific productivity: Multidimensional analysis in the complexity of a Mexican private university. Publications. 11(2):27. doi:10.3390/publications11020027. [accessed 2023 Oct 3]. https://www.mdpi.com/2304-6775/11/2/27.

[3] See KF, Ng YC, Yu M-M. 2022. An alternative assessment approach to national higher education system evaluation. Eval Program Plann. 94(102124):102124. doi:10.1016/j.evalprogplan.2022.102124. https://www.sciencedirect.com/science/article/pii/S0149718922000787.

## Appendix

### QQ Plot

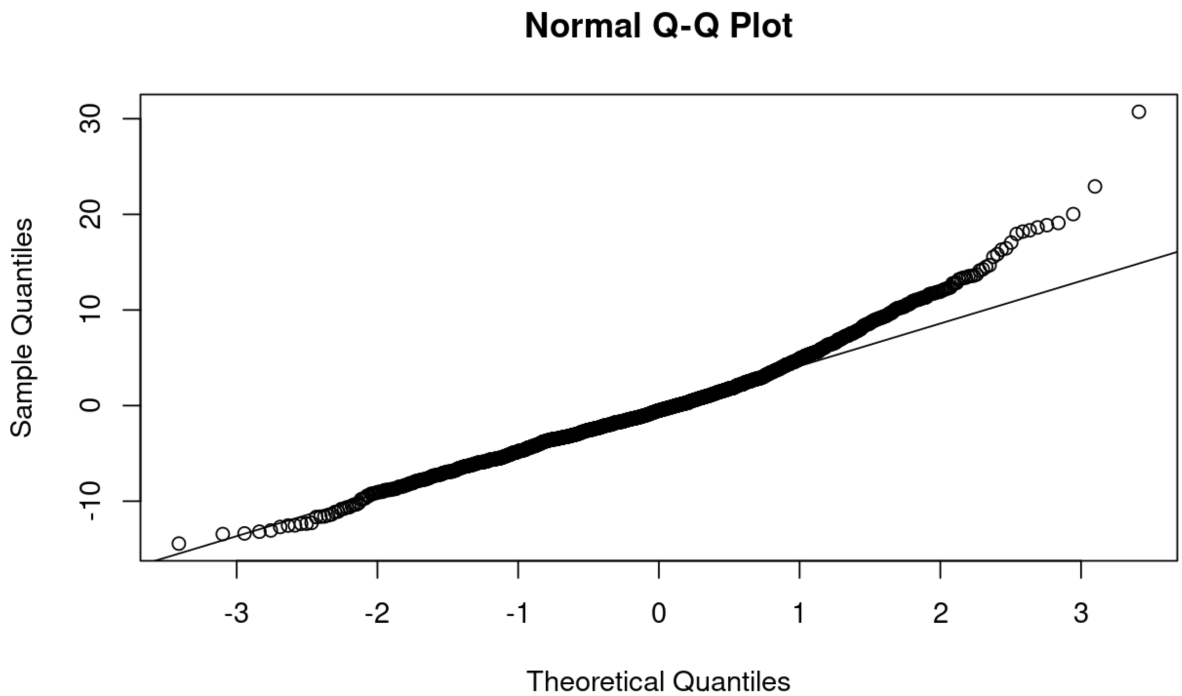


Figure 2: QQ Plot Showing Violated Normality Assumption

### Ethics

We chose to use manual selection methods. We already had a reasonnably small number of predictors to begin with, and there were also a number of predictors that qualified as predictors of interest due to their presence in the literature, that we would not have wanted to eliminate anyways, regardless of their statistical significance.

If the number of predictors had been greater, we might've chosen to do things automatically instead. Regardless of the chosen method, we still ultimately bear responsibility for the resulting model. So when deciding which of the two approaches (manual or automatic selection) is most ethical, we should consider which of the two is likely to produce the best model. Sometimes, automated methods might produce the better model in cases where there are too many predictors for a human to manually test the different potential models.

However, in a case like ours with relatively few potential predictors, the manual approach is likely better, because humans can understand models in a more detailed way than an automated selection algorithm. For a specific ethical example that applies to our model, a human can identify that using region as a predictor might be ethically problematic if we wanted to use our model to help students decide what institutions to apply to, since it might lead to more talented students leaving certain regions (sometimes known as “brain drain”), but an automated selection algorithm wouldn't take this into account at all.