## Introduction

In recent years, there has been a notable decline in both college enrollment and college applications in the United States (Pavlov and Katsamakas, 2020). This decline has prompted research into various factors that influence students' motivation to apply to college. The motivation of this study is to examine students' motivation to attend college by examining the number of applications schools receive. Specifically, we will examine whether financial considerations, college staff issues, and different credentials that colleges might offer can significantly influence the number of applications.

The financial considerations examined in this study include college tuition and fees and the availability of financial aid. Numerous studies have highlighted the importance of financial factors in students' college selection process. For example, Schumacher (2015) interviewed 26 students about their college selection criteria, and the majority indicated that the tuition and loan assistance offered by the school was the primary reason for their selection. Caroline (2004) explored the impact of financial aid on the college decisions of high-achieving students by developing an economic model, highlighting that financial packages are positively correlated with the number of school applications.

Another aspect that could influence an institution's attractiveness to students is the quality of educational resources provided by its faculty. Faculty members play a critical role in a school's reputation and student experience (Womble, 2018). Therefore, it is imperative to examine how faculty members influence students' learning experiences and how their compensation affects this dynamic. One important indicator of a college's educational resources is the student-faculty ratio. McDonald (2013) found that the student-faculty ratio contributes to a school's reputation and learning outcomes. Alter and Reback (2014) further found that a school's reputation significantly affects the number of college applications. In addition, Ehrenberg and Zhang (2005) examined the economic determinants of faculty salaries and their influence on institutional reputation. Therefore, two factors related to faculty will be examined, which are student-faculty ratio and faculty salary.

The third aspect of this study is to look at whether students are more interested in applying baccalaureate degrees or certificate offering programs. This is because students may value different financial benefits that different credentials could offer. For example, Kim and Tamborini (2019) examined the long term financial benefit of higher education and found that those with baccalaureate degrees earn more than those with associate degrees. There is a similar study that compares earnings differences by credential type and gender. At the median level, male certificate holders earn more than 40% of men with associate's degrees and 24% of men with bachelor's degrees. Similarly, at the median level, female certificate holders earn more than 34 percent of women with associate's degrees and 24 percent of women with bachelor's degrees (Carnevale et al., 2012).

While there is extensive research on the individual impact of these factors, there is limited research on how these factors collectively affect college applications. Therefore, the purpose of this study is to understand how these factors - tuition, student-faculty ratio, faculty salary, and the percentage of students receiving financial aid - collectively affect college applications, as well as whether each of these factors is a significant predictor. This understanding will help colleges improve their ability to attract students.

## Sample and Method

Data description

The data used for this study is from the Integrated Postsecondary Education Data System (IPEDS) project on the National Center for Education Statistics website (*The Integrated Postsecondary Education Data System,* n.d.). The system consists of 12 connected surveys conducted annually that collect data from every college, university, and other higher education institution in the U.S. The IPEDS website provides a custom data selection tool that allows users to merge only the variables of interest. The dataset used for this study includes data collected during the fall 2022 application season from all New York State higher education institutions using this tool.

The original dataset contains some missing values. Primarily, the missing value is the number of applications reported by each school; 211 out of 422 schools did not report the number of applicants they received. We decided to eliminate these missing values, since after elimination there are still 182 observations, which is still enough for this research. The names of the institutions are eliminated, leaving only the ID for privacy reasons. The final dataset consists of 182 observations and 8 observations. The variable names are changed to shorter names for easier analysis and interpretation. Table 1 is a list of continuous variable descriptions and Table 2 is the description for the categorical variable.

**Table 1**

***Numeric variable descriptions***

| Variable | Description |
|---|---|
| unitid | University ID |
| year | The year in which the data was collected (2022) |
| sf_ratio | Student to faculty ratio |
| salary | Average faculty salary |
| fees | Total cost of tuition and other school fees |
| finanAid_percent | Percent of students receiving financial aid |
| applicants | The number of total applicants in 2022 |

**Table 2**

***Categorical variable description***

| Vairable | Description | Levels | Description |
|---|---|---|---|
| category | Category of schools classified by the level of offerings | baccalaureate | Degree-granting, primarily baccalaureate or above |
| | | non-baccalaureate | Degree-granting, not primarily baccalaureate or above |
| | | asso&certificate | Degree-granting, Associate's and certificates |

Table 3 and Table 4 show descriptive statistics of numeric and categorical variables, respectively. Figure 1 is a visualization comparison of the distribution of application numbers of three levels of school category. On average, schools that primarily offer baccalaureate or higher degrees receive the most applications, followed by schools that primarily offer nonbaccalaureate and associate's degrees and certificates.

**Table 3**

***Descriptive statistics of numeric variables***

| Statistic | sf_ratio | salary | fees | finanAid_percent | applicants |
|---|---|---|---|---|---|
| Min. | 3 | 16000 | 1008 | 12 | 0 |
| Max. | 47 | 185958 | 68365 | 100 | 100662 |
| 1st Qu. | 9 | 55346 | 10012 | 74 | 99.5 |
| Median | 13 | 73254 | 17465 | 89.5 | 2181.5 |
| Mean | 13.3 | 75119 | 26217 | 83.88 | 7283.8 |
| 3rd Qu. | 15 | 96374 | 38757 | 98 | 8297.5 |

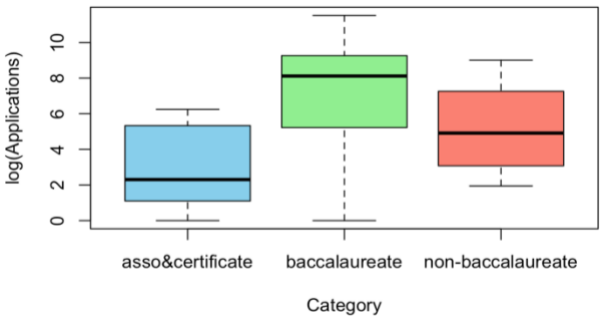
**Table 4**

***Frequency school category***

| Levels | n | Frequency |
|---|---|---|
| baccalaureate | 152 | 0.0604 |
| non-baccalaureate | 19 | 0.8351 |
| asso&certificate | 11 | 0.1043 |

**Figure 1**

***Category variable frequency***




*Note. This box plot is plotted taking the logarithm of the number of applicants because of the large differences in mean and distribution among the three categories.*

Method description

Multiple linear regression (MLR) is used in this study to make predictions about the collective influence of multiple factors on student decision making. MLR is a powerful statistical technique that allows researchers to understand the relationship between one dependent variable and multiple independent variables. In the current study, MLR provides a way to analyze how different factors - tuition, student-staff ratio, staff salary, and the percentage of students receiving financial aid - collectively affect the number of applications received by colleges. This is critical to this study because the decision-making process of students applying to colleges is influenced by a combination of factors rather than a single variable.

Interpretation of results depends on significance testing and coefficient analysis. For significance testing: Perform hypothesis tests (t-tests) on the regression coefficients to determine the significance of each independent variable. A p-value less than 0.05 is considered statistically significant. Coefficients will be interpreted in terms of magnitude and direction to understand the impact of each independent variable on the number of college applications. ANOVA tests (F tests) will be used to determine the overall significance of the model. A p-value less than 0.05 will be considered statistically significant.

**Analysis and Findings**

First, we estimated a model where the number of applicants received for each school served as the dependent variable and all other numerical continuous variables were independent variables, denoted as lmout1.

*lmout1: applicants=β0+β1sf_ratio+β2salary+β3fees+β4finanAi_percent+ϵ.*

In this model, all coefficients were found to be insignificant (p>0.05). Diagnostic checks for constant variance, multicollinearity, and normality were performed to identify potential problems. A violation of normality (right-skewed) and the presence of several outliers (absolute value of the studentized residual > 3) were identified. These issues were addressed by removing the outliers and taking the logarithm of the dependent variable (log applicants).

After the model diagnostics, a second log applicants model is run with the same dependent variables. Denoted as lmout2.
*lmout2: log(applicants)=β0+β1sf_ratio+β2salary+β3fees+β4finanAi_percent+ϵ.*

In lmout2, the student-staff ratio was found to have a non-significant contribution to the model (p>0.05). Therefore, the student-staff ratio was removed and we reran the model called lmout3.

*lmout3: log(applicants)=β0+ β1salary+β2fees+β3finanai_percent+ϵ.*
In lmout3, overall, all coefficients for average salary, annual tuition and fees, and number of students receiving financial aid are significant predictors of the log of the number of applicants (p<0.05). Table 5 is a result of lmout3.

***Table 5***

***lmout3 results***

| Predictor | Estimate | Std. Error | t value | p value |
|---|---|---|---|---|
| (Intercept) | -0.03368 | 1.009 | -0.033 | 0.9734 |
| salary | 0.0000571 | 0.00000629 | 9.081 | <0.001*** |
| fees | 0.0000206 | 0.00000976 | 2.113 | 0.036* |
| finanAid_percent | 0.02434 | 0.009644 | 2.524 | 0.0125* |

*Note. Residual standard error: 2.104 on 178 degrees of freedom. Multiple R-squared: 0.4503, Adjusted.mR-squared: 0.441. F-statistic: 48.6 on 3 and 178 DF, p-value: < 2.2e-16. Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05.*

From Table 5, we can first conclude that the coefficient for average faculty salary is 5.710e-05 and it's highly significant (p < 0.001). This means that for every unit increase in the average faculty salary, the log of the number of applicants increases by 5.710e-05, holding other variables constant. The coefficient on fees is also significant (p = 0.0360). It suggests that for each unit increase in the total cost of tuition and other school fees, the log of the number of applicants increases by 2.061e-05, holding other variables constant. The coefficient on the percentage of students receiving financial aid is 2.434e-02, which is also significant (p = 0.0125). This suggests that for every unit increase in the percentage of students receiving financial aid, the log of the number of applicants increases by 2.434e-02, holding other variables constant.

Overall, the model is statistically significant (F= 48.6, p< 2.2e-16). And the adjusted R-squared indicates that about 44.1% of the variability in the log of the number of applicants can be explained by the model.

Given this satisfactory result, we add the school category to the model, denoted as lmout4.

*lmout4: log(applicants)=β0+ β1salary+β2fees+β3finanai_percent+ β4category +ϵ.*

The result in Table 4. indicates that annual tuition and fees, the number of students who received financial aid become insignificant (p>0.05), see the result in Table 6. Institutions offering baccalaureate degrees is positively and significantly related to the log number of applications (p < 0.001). The average difference in log(applications) between institutions offering baccalaureate degrees and those offering associate or certificate degrees is 3.132, holding other variables constant. Being a non-baccalaureate institution is positively and significantly related to the log number of applicants (p < 0.01). Compared to the associate or certificate category, being a non-baccalaureate institution increases the log number of applicants by approximately 2.516. Figure 2 is a visual illustration of the categorical differences.

***Table6***

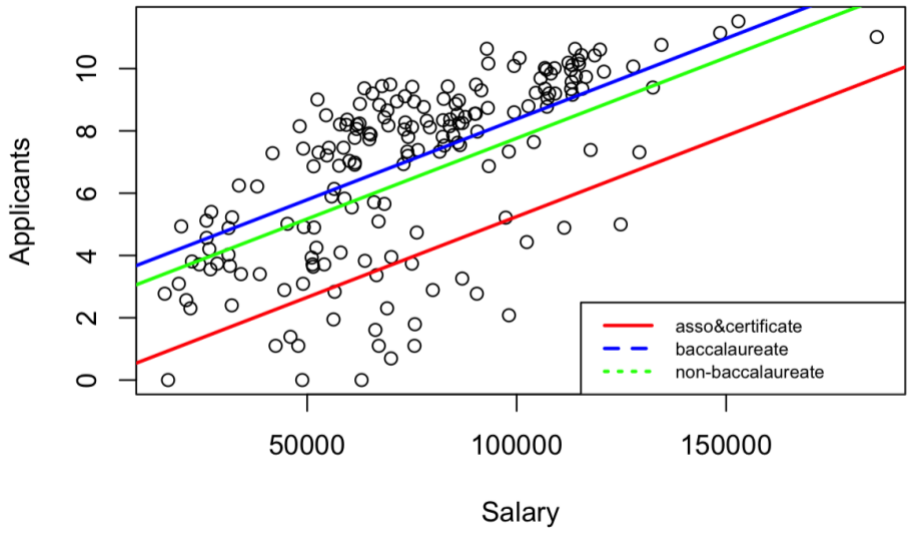
***lmout4 results***

| Predictor | Estimate | Std. Error | t value | p value |
|---|---|---|---|---|
| (Intercept) | -1.452 | 1.019 | -1.424 | 0.15616 |
| salary | 0.0000518 | 0.0000061 | 8.495 | <0.001 *** |
| fees | 0.0000182 | 0.00000925 | 1.967 | 0.05080 . |
| finanAid_percent | 0.0124 | 0.009432 | 1.316 | 0.18991 |
| categorybaccalaureate | 3.132 | 0.649 | 4.826 | <0.001 *** |
| categorynon-baccalaureate | 2.516 | 0.765 | 3.29 | 0.00121 ** |

*Note. Residual standard error: 1.987 on 176 degrees of freedom. Multiple R-squared: 0.5153, Adjusted R-squared: 0.5016. F-statistic: 37.43 on 5 and 176 DF, p-value: < 2.2e-16. Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1.*

***Figure 2***

***Regression Lines of Applicants vs. Salary by Institutional Category of lmout4***

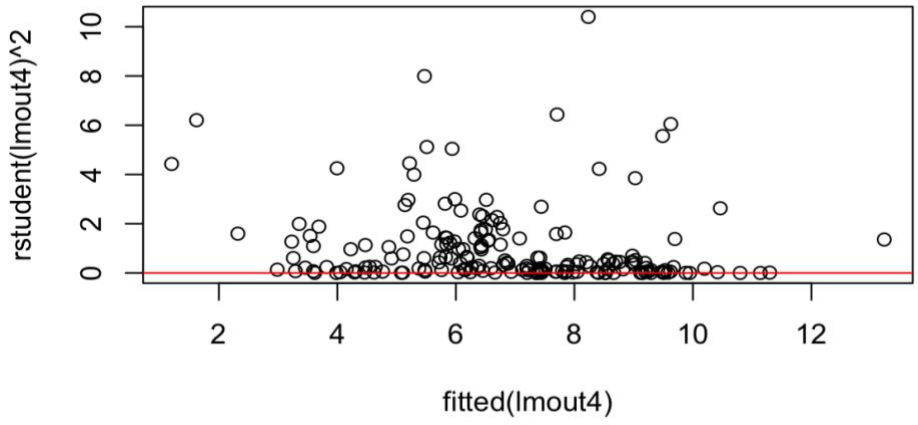



*Note. The plot shows the regression line of applicants and salary in lmout4, with three different regression lines representing each institutional category, by controlling fees and financial percent by their mean.*

Although the overall lmout 4 is significant (F=37.43, p<0.001), there are some variables in this model that do not contribute significantly to the prediction, such as fees (p=0.0508) and the finanAid_percent (p=0.189). This could be due to a violation of the constant variance in this model, shown in Figure 3, that the variables are not evenly distributed around y=0.

***Figure 3***

***Squared Studentized Residuals vs. Fitted Values of lmout4***




*Note. The plot shows the squared studentized residuals against the fitted values from the regression model (lmout4). In this plot, most points appear to cluster near the horizontal axis, but some variability increases with the fitted values, suggesting potential heteroscedasticity. The red line indicates y=0.*

## Discussion

In summary, this study found that average faculty salary, annual tuition and fees, the number of students receiving financial aid, and school category were significant predictors of the number of college applications. These findings are consistent with the existing literature, which emphasizes the importance of financial and institutional factors in shaping the decision of students to attend college. However, the student to staff ratio (SSR) was found to be an insignificant predictor of the number of applications. This inconsistency may be because previous studies have focused primarily on the impact of SSR on school reputation rather than its direct effect on application numbers. And this study assumed that SSR indirectly affects applications through its effect on reputation.

There are several limitations to this study that should be acknowledged. First, there is a challenge in the treatment of outliers. While outliers can bias the results of a regression model, excluding outliers may also lead to inaccurate models. The result could be more robust if more ways of dealing with them could be adopted. Second, the problem of inconsistent variance or heteroskedasticity in regression models was not adequately addressed in the last model lmout4. Heteroskedasticity can lead to inefficient estimation and affect the validity of hypothesis testing. Future research should use methods such as weighted least squares (WLS) or robust standard errors to correct for this problem.

In addition, this study was limited to data from New York State and may not be representative of the state of education in the United States as a whole. Factors affecting college applications may vary significantly by region, and a more comprehensive dataset would provide a better understanding of national trends. Furthermore, there are some other factors in the past research are shown also influence students' college application decisions as the impact of school marketing strategies(Szekeres, 2010), high school environment(Roderick et al., 2011) and personal connections to a university (Mothkvich, 2018). In addition, this study does not consider students' demographic characteristics and personal preferences, which can be very different. For example, students from different economic backgrounds and ethnicities may have different college application preferences.

Future research should aim to address these limitations by incorporating a more diverse and representative dataset, examining additional factors that influence college applications, and using more advanced statistical techniques to deal with outliers and heteroscedasticity. In addition, future research could also explore the diverse preferences and backgrounds of students to develop more nuanced models that reflect the complexity of college application decisions. By improving our understanding of the factors that drive college applications, institutions can develop more effective strategies for attracting and retaining students, ultimately contributing to the broader goal of expanding access to higher education.

**References**


Alter, M., & Reback, R. (2014). True for Your School? How Changing Reputations Alter Demand for Selective U.S. Colleges. *Educational Evaluation and Policy Analysis*, *36*(3), 346–370. https://doi.org/10.3102/0162373713517934

Carnevale, A. P., Rose, S. J., & Hanson, A. R. (2012). *Certificates: Gateway to Gainful Employment and College Degrees*.

Hoxby, C. M. (2004). College Choices: The Economics of Where to Go, When to Go, and How to Pay For It. *NBER Books*.

Kim, C., & Tamborini, C. (2019). Are They Still Worth It? The Long-Run Earnings Benefits of an Associate Degree, Vocational Diploma or Certificate, and Some College. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(3), 64. https://doi.org/10.7758/rsf.2019.5.3.04

McDonald, G. (2013). Does size matter? The impact of student–staff ratios. *Journal of Higher Education Policy and Management*, *35*(6), 652–667. https://doi.org/10.1080/1360080x.2013.844668

Mothkovich, T. A.(2018). *Making Cents Of It: How The Focus On Financial Aid By Selective Universities May Not Be Enough To Attract More Low-Income, High-Achieving Students.* (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/4705

Pavlov, O. V., & Katsamakas, E. (2020). Will Colleges Survive the Storm of Declining Enrollments? A Computational Model. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3841287

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the Road to College: High School Effects in Shaping Urban Students' Participation in College Application, Four-year College Enrollment, and College Match. Sociology of Education, 84(3), 178-211. https://doi.org/10.1177/0038040711411280

Schumacher, R. M. (2015). *What Attracts Students To A Small, Private University?*

Szekeres, J. (2010). Sustaining student numbers in the competitive marketplace. *Journal of Higher Education Policy and Management*, *32*(5), 429–439. https://doi.org/10.1080/1360080x.2010.511116

The Integrated Postsecondary Education Data System. (n.d.). Nces.ed.gov. https://nces.ed.gov/ipeds

Womble, C. C. (2018). Faculty salary (in) equity: A review of the literature. *Trends in Diversity*, *1*(1), 1.