Analyzing Key Factors for Increasing University Applications: A Focus on Institution Type, Acceptance, and Enrollment Rates

Introduction

The goal of the study is to pinpoint the main variables affecting how many applications colleges receive. Information was gathered from 777 universities in the US, including both public and private ones. Acceptance rates, enrollment rates, and institution status (private vs. public) are among the factors included in the dataset. To provide colleges with useful information to boost application rates, the main objective is to investigate the connections between these variables and the volume of applications.

Summary Statistics

Summary Statistics for Number of Applications received

Quantiles

100.0%	maximum	48094
99.5%		19908.09
97.5%		14442.4
90.0%		7694.4
75.0%	quartile	3635
50.0%	median	1558
25.0%	quartile	776
10.0%		456
2.5%		244.45
0.5%		149.01
0.0%	minimum	81

Summary Statistics

Mean	3001.6384
Std Dev	3870.2015
Std Err Mean	138.8427
Upper 95% Mean	3274.1902
Lower 95% Mean	2729.0866
N	777
N Missing	0

Summary Statistics for Private and public Universities

Public Universities

Quantiles

100.0%	maximum	48094
99.5%		46385.15
97.5%		18814.65
90.0%		12791
75.0%	quartile	7781.5
50.0%	median	4307
25.0%	quartile	2178.25
10.0%		991
2.5%		447.5
0.5%		236.38
0.0%	minimum	233

Summary Statistics

Mean	5729.9198
Std Dev	5370.6753
Std Err Mean	368.8595
Upper 95% Mean	6457.0417
Lower 95% Mean	5002.7979
N	212
N Missing	0

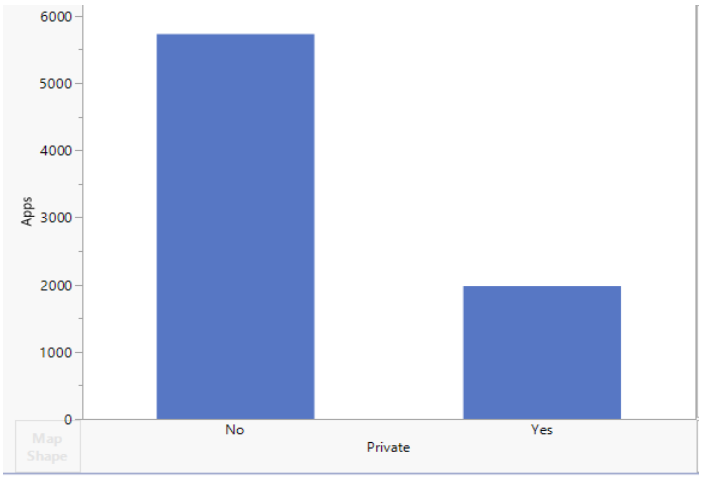
Private Universities

Quantiles

100.0%	maximum	20192
99.5%		13963.77
97.5%		10220.35
90.0%		4296.2
75.0%	quartile	2188
50.0%	median	1133
25.0%	quartile	617.5
10.0%		383
2.5%		223.5
0.5%		134.03
0.0%	minimum	81

Summary Statistics

Mean	1977.9292
Std Dev	2443.3413
Std Err Mean	102.79214
Upper 95% Mean	2179.8314
Lower 95% Mean	1776.027
N	565
N Missing	0



Methods

Study Design

This analysis employs multiple linear regression models to evaluate the relationships between the number of applications (Apps) and explanatory variables:

1. Quantitative: Accept (number of accepted students) and Enroll (number of enrolled students).

2. Qualitative: Private (university status as public or private).
3. Interaction Terms: Accept*Enroll.
4. Quadratic Terms: Accept*Accept and Enroll*Enroll.

Statistical Approach

1. **Model 1**: A quantitative model with Accept and Enroll as predictors.
    a. Purpose: To determine how acceptance and enrollment rates affect the number of applications.
2. **Model 2**: Includes the categorical variable Private alongside Accept and Enroll.
    a. Purpose: To investigate the effect of university type (public/private) on applications.
3. **Model 3**: Explores quadratic and interaction terms for the quantitative predictors.
    a. Purpose: To capture any nonlinear relationships between Accept and Enroll and their interactions.

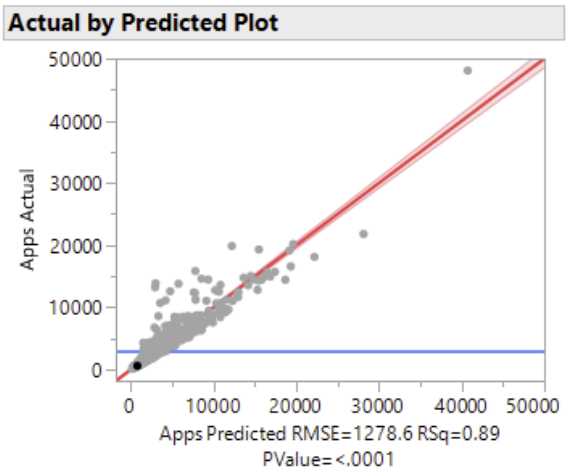
All models were evaluated using the following:

- **Goodness-of-Fit Metrics**: R-squared and Adjusted R-squared values.
- **Statistical Significance**: p-values for predictors to assess their contribution.
- **Residual Analysis**: To verify model assumptions.

JMP was used for data analysis and visualization.

Results

Model 1: Quantitative Variables



Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	21.206074	60.25667	0.35	0.7250
Enroll	-0.327017	0.120188	-2.72	0.0067*
Accept	1.6026795	0.045561	35.18	<.0001*

$$E(Apps) = \beta 0 + \beta 1 x_1 + \beta 2 x_2$$

- $\boldsymbol{E(Apps)}$: Expected mean of applications received.
- $\boldsymbol{\beta_0}$**:** Intercept (21.206), representing the baseline number of applications when both predictors (Accept and Enroll) are zero.
- $\boldsymbol{\beta_1}$**:** Coefficient (1.6027) estimating the increase in applications per additional accepted student ($p < 0.0001$).
- $\boldsymbol{\beta_2}$**:** Coefficient (−0.3270) estimating the decrease in applications per additional enrolled student ($p = 0.0067$).

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	1.0358e+10	5.179e+9	3168.032
Error	774	1265308319	1634765.3	Prob > F
C. Total	776	1.1623e+10		<.0001*

Analysis of Variance Test (F-Test):

- **F-Ratio:** 3168.032
- **P-Value:** < 0.0001
- Null Hypothesis (H_0): $\beta_1 = \beta_2 = 0$
- Alternative Hypothesis (H_A): At least one $\beta \neq 0$.
- At the 5% significance level, H_0 is rejected, providing strong evidence that at least one model coefficient is non-zero. This confirms the model's statistical usefulness for predicting the number of applications received.

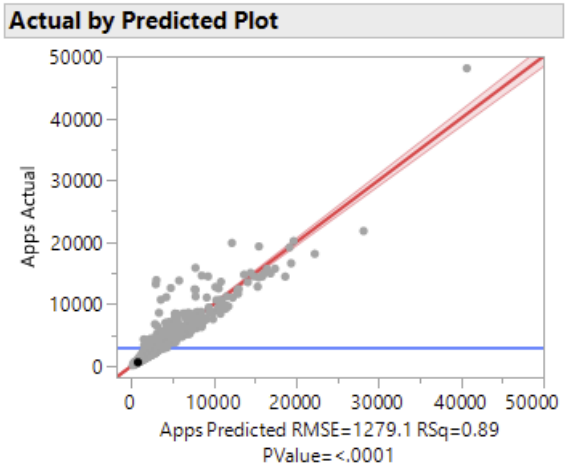
Summary of Fit	
RSquare	0.89114
RSquare Adj	0.890859
Root Mean Square Error	1278.579
Mean of Response	3001.638
Observations (or Sum Wgts)	777

R-Squared and Adjusted R-Squared

$R^2 = 0.8911$: 89.14% of the variation in applications is explained by the model.

$R^2_{adj} = 0.8908$: After adjusting for sample size and degrees of freedom, 89.08% of the variation is explained.

Model 2: Quantitative and Qualitative Variables



Prediction Expression

-10.99934

+ 1.5992794714 • Accept

+ -0.298582738 • Enroll

+ Match(Private, "No" ⇒ -37.1810208, "Yes" ⇒ 37.181020798, else ⇒ .)

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

- $\boldsymbol{E(y)}$: Expected mean of applications received.
- $\boldsymbol{\beta_0}$**:** Intercept (−10.999), representing the baseline for public universities when other predictors are zero.
- $\boldsymbol{\beta_1 x_1}$**:** Coefficient (1.5993), estimating the increase in applications per additional accepted student ($p < 0.0001$).
- $\boldsymbol{\beta_2 x_2}$**:** Coefficient (−0.2986), estimating the decrease in applications per additional enrolled student ($p < 0.0001$).
- $\boldsymbol{\beta_3 x_3}$**:** Coefficient (37.181), representing the advantage in applications received by private universities compared to public universities.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.0359e+10	3.4528e+9	2110.356
Error	773	1264739909	1636144.8	Prob > F
C. Total	776	1.1623e+10		<.0001*

Analysis of Variance (F-Test):

- F-Ratio: 2110.356
- P-Value: < 0.0001
- Null Hypothesis (H_0): $\beta_1 = \beta_2 = \beta_3 = 0$
- Alternative Hypothesis (H_A): $At\ least\ one\ \beta \neq 0$.
- At the 5% significance level, H_0 is rejected. The model is statistically useful for predicting applications received.

Summary of Fit	
RSquare	0.891189
RSquare Adj	0.890767
Root Mean Square Error	1279.119
Mean of Response	3001.638
Observations (or Sum Wgts)	777

R-Squared and Adjusted R-Squared

- $R^2 = 0.8912$: 89.12% of the variation in applications is explained by the model.
- $R^2_{Adj} = 8.908$: After adjusting for sample size and degrees of freedom, 89.08% of the variation is explained.

Model 3: Second order regression model of the two quantitative factors Acceptance and Enrollment.

Equation for the second order regression model :$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2$

$E(y)$: Expected mean of Applications received.

β_0 : Intercept.

$\beta_1 x_1$: β estamite of number of applications Accepted.

$\beta_2 x_2$:β estimate of number of applications Enrolled.

$\beta_3 x_1 x_2$: β estimate interaction of number of accepted and number applications Enrolled.

$\beta_4 x_1^2$: β estimate interaction of number of applications Accepted.

$\beta_5 x_2^2$: β estimate interaction of number of applications Enrolled.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	1.0501e+10	2.1002e+9	1443.078
Error	771	1122103197	1455386.8	Prob > F
C. Total	776	1.1623e+10		<.0001*

Analysis of Variance test (F-tetst)

F Raito = 1443.078

Pvalue < 0.0001

H_0: $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \neq 0$ H_A: $At\ least\ \beta_1, \beta_2, \beta_3, \beta_4\ or\ \beta_5 = 0$

(H_0) Initial hypothesis is rejected. At 5% significance level, there is strong evidence that at least one of the model coefficients is non-zero. This model is statistically useful for predicting the number of applications received.

T-test for Betas

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-229.9008	73.22317	-3.14	0.0018*
Accept	1.4073476	0.093068	15.12	<.0001*
Enroll	0.8156237	0.251381	3.24	0.0012*
Accept*Enroll	-0.000278	5.456e-5	-5.10	<.0001*
Accept*Accept	5.2627e-5	7.132e-6	7.38	<.0001*
Enroll*Enroll	0.0001626	9.528e-5	1.71	0.0883

t-ratio = -5.10

Pvalue < 0.0001

$$H_0: \beta_3 = 0 \quad H_A: \beta_3 < 0$$

(H_0) Initial hypothesis is rejected. At 5% significance level, there is strong evidence that the number of applications accepted interacts negatively with the number of applications enrolled. The applications increase slowly per unit increase in number of applications accepted and fastly per unit per unit increase in number of applications enrolled.

$$H_0: \beta_4 = 0 \quad H_A: \beta_4 < 0$$

T-ratio = 7.38

Pvalue < 0.0001

(H_0) Initial hypothesis is rejected. At 5% significance level, there is strong evidence to conclude that the number of applications is increasing slowly per unit increase in number of applications accepted.

$$H_0: \beta_5 = 0 \quad H_A: \beta_5 < 0$$

T-ratio = 1.71

Pvalue = 0.0883

(H_0) () Initial hypothesis failed to be rejected. At 5% significance level, there is strong evidence to conclude that the number of applications is increasing slightly fast per unit increase in number of applications enrolled.

Summary of Fit

RSquare	0.903461
RSquare Adj	0.902835
Root Mean Square Error	1206.394
Mean of Response	3001.638
Observations (or Sum Wgts)	777

R^2=0.9035: This shows that %90 of the sample variation in number of applications received is explained by the model.

R^2_{adj}=0.9028: This shows that %90 of the sample variation in number of applications received is explained by the model after adjusting sample size and degree of freedoms.

Model 4 : Regression Analysis model for Number of Application received with qualitative predictor of Public/private indicator.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2170135119	2.1701e+9	177.9147
Error	775	9453149477	12197612	Prob > F
C. Total	776	1.1623e+10		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3853.9245	140.6455	27.40	<.0001*
Private[No]	1875.9953	140.6455	13.34	<.0001*

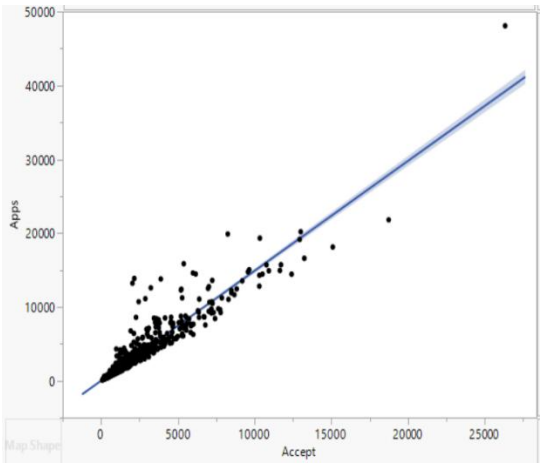
Analysis of Variance test (F-test)

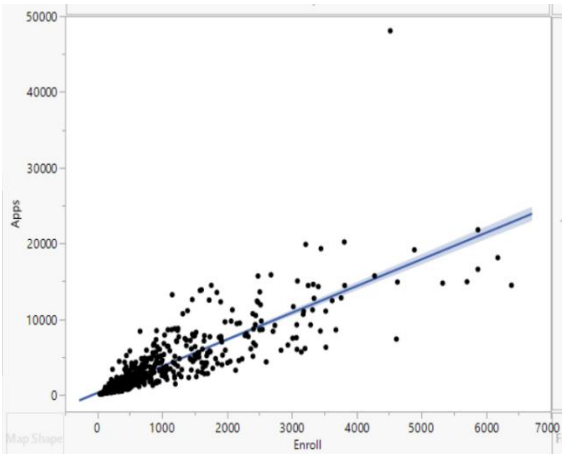
$$H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$$

(H_0) Initial hypothesis is rejected. At 5% significance level, there is strong evidence of the difference between the predicted number of applications received depending on the Public/private indicator, thus Public/private indicator is useful predictor of number of applications received.

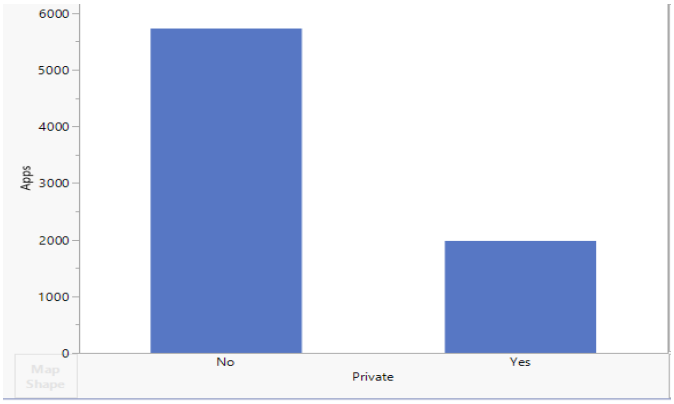
Conclusion

This report focused on how different factors affect the number of applications universities receive. Multiple regression analysis used to show the important patterns of what causing numbers of application received increase or decrease. There were three main predictors used from the dataset, Number of applications accepted, Number of applications enrolled and Private/public indicators. Number of applications accepted, and Number of applications enrolled are quantitative parameters of this problem. Both parameters are chosen stronger positive correlations.





Third indicator is Private/public which is considered qualitative predictor. I in the graph below we see how this indicator is decisive in predicting number of applications received, as shown in the graph the number of applications received with NO category is more than the number of applications with YES category.



There are four regression models on studying the three predictors, number of applications accepted, and number applications enrolled and Private/public indicator. The first model is to study the relationship between the two quantitative predictors for the number of applications received. The variance analysis (F-test) indicated that there is at least one of the coefficients (βs) has non nonzero effect on the number of applications received. The coefficient of determination R^2that %89 of the sample variation for the number of applications was explained by the model and can be relied upon for forecasting applications. However, the negative effect of enrolled students warrants

further investigation to understand its implications for policy or strategy. The second model improves upon the first model by incorporating the qualitative distinction between public and private universities, which proves to be a significant factor. The qualitative variable improves the model's explanatory power and provides understanding of the factors driving application numbers. The third model is a second-order model, it demonstrates importance of considering non-linear interactions and interactions between number of applications accepted and number of applications enrolled. The negative interaction between the two predictors shows a decline of application received when both predictors are high, while the quadratic terms indicate that growth in applications is not constant but varies depending on the predictor levels. In the quadratic interaction for number of applications accepted (β_4) showing that the number of applications grows at a decreasing rate as the number of Accepted students increases. In the quadratic interaction for number of applications enrolled (β_5)showing a slight non-linear decline in applications with increased number of applications enrolled. The coefficient of determination R^2 for this model shows that %90 of the sample variation for the number of applications was explained by the model the adjusted coefficient of determination of this model explains the same percentage of R^2. The fourth model demonstrates that the Public/Private indicator is a statistically significant qualitative predictor of the number of applications received. In summary universities depending on mor acceptance and enrollment rates will create a slow growth of number of applications, they should focus more in creating an extra incentive for students to apply, for example they can reduce tuition or connect with more scholarship funders. Also, they can work on introducing their majors and programs to high schools' juniors or seniors.