

Appendix

Table of Contents

POTENTIAL EXPLANATORY VARIABLES:	2
CHECKING MODEL ASSUMPTIONS:	2
EXPLANATION FOR OUTLIERS:	4
PREDICTED RANKING IS LOWER THAN ACTUAL RANKING:	4
PREDICTED RANKING IS HIGHER THAN ACTUAL RANKING:	4

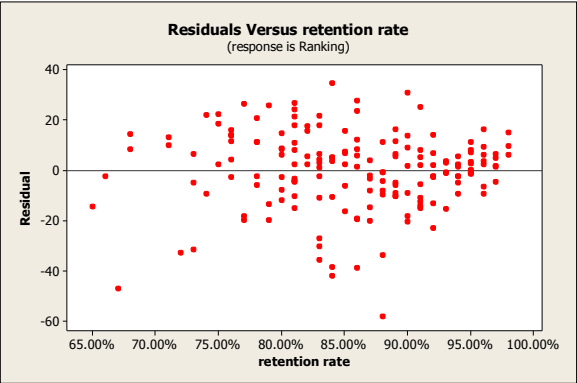
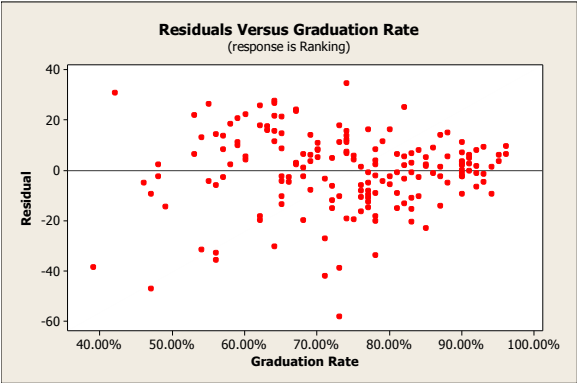
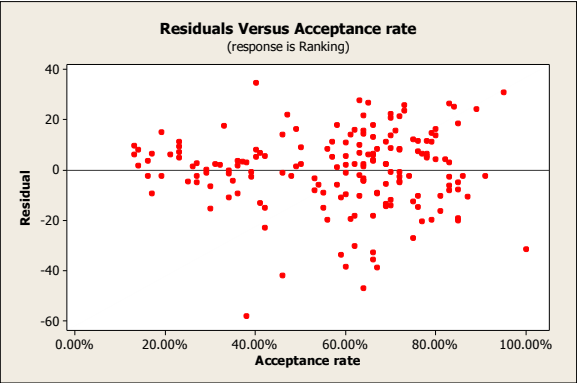
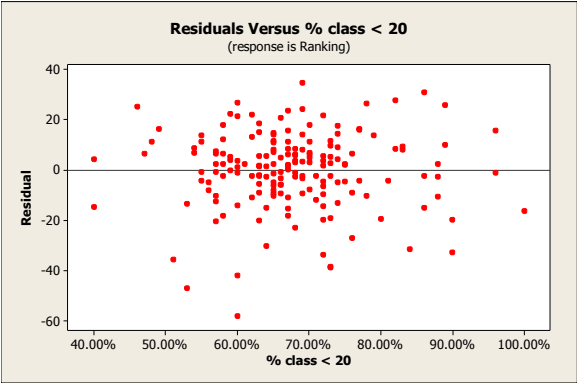
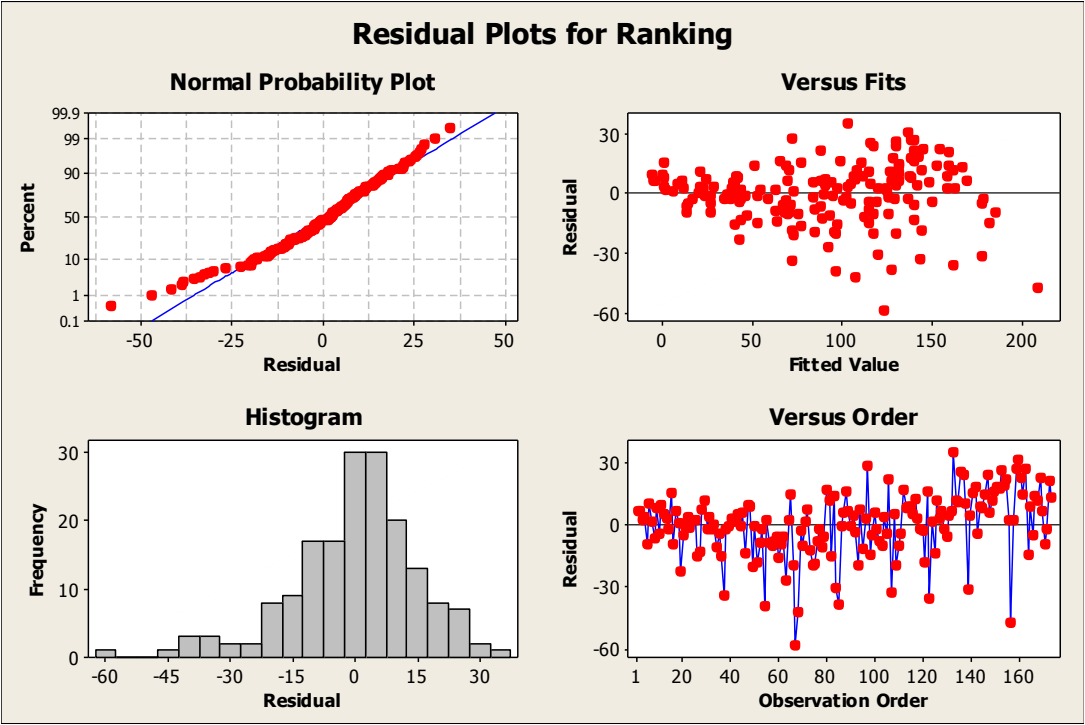
Potential explanatory variables:

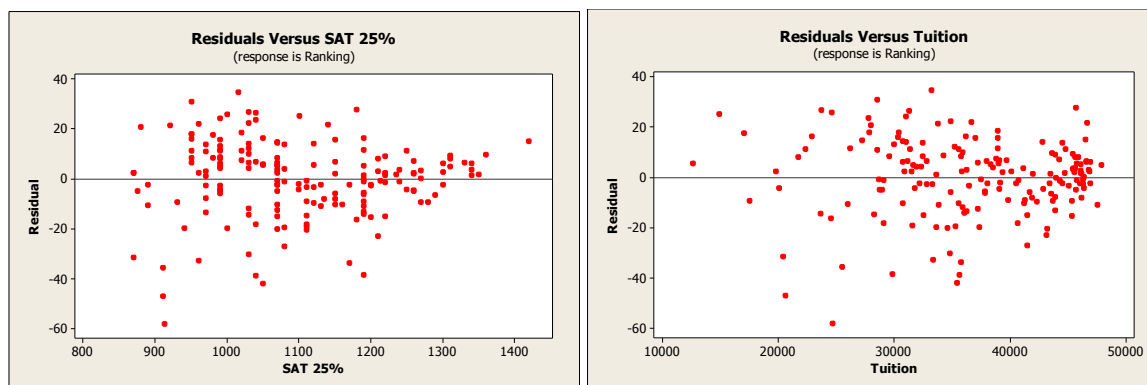
Variable	Description	Type of Variable
Ranking	U.S. News & World Report 2014 Liberal Arts Ranking based on data from 2012-13	Response
Tuition and fees	2013-2014 Tuition and fees (no room and board nor financial aid included)	Explanatory
Enrollment	Student population	Explanatory
Setting	City, urban, suburban or rural	Explanatory
SAT 25th percentile ('12)	SAT Score on a scale of 300-800	Explanatory
SAT 75th percentile ('12)	SAT Score on a scale of 300-801	Explanatory
Average freshman retention rate	Percent of students return after first year	Explanatory
2012 Graduation Rate	Actual percentage of students finishing college within 6 years	Explanatory
Acceptance rate ('12)	Percent of applicants who were offered admission	Explanatory
% of classes of 50 or more ('12)	NA	Explanatory
% of classes under 20 ('12)	NA	Explanatory
Freshmen in top 10% of HS class ('12)	Percent of freshmen who ranked top 10% at their high school	Explanatory
Student/Faculty Ratio	The student/faculty ratio indicates how much attention each student can get.	Explanatory
Location	Location of the college	Explanatory
% Accept Offer	Percent of admitted students accepted the college's offer	Explanatory
% receiving Financial Aid	Percent of students receiving financial aid	Explanatory

Checking Model Assumptions:

We have to check the following assumptions for the regression model to be valid:

- 1) The normality of the data set is checked by the normal probability plot (outliers are explained later). Neglecting the several outliers, the data is approximately normal.
- 2) Equal variances assumption is checked with the plot of the residuals versus fits and each explanatory variable. We can see from each graph that residuals are approximately equally distributed across residual = 0 line, which make the assumption valid.
- 3) The independence of the data is shown by residuals versus order plot, as there are no clear clusters of data and no clear pattern exists in the graph.





Explanation for outliers:

Predicted ranking is lower than actual ranking:

- Sewanee - University of the South charges a much lower tuition than schools with similar ranking, and has a relatively high acceptance rate.
- New College of Florida has an extremely low rate of 6-year graduation - 39% compared to other schools of similar ranking.
- Spelman College charges a very low tuition compared to schools of similar ranking and its SAT 25% is much lower compared to similar ranking schools.
- Xavier University of Louisiana charges a relatively low tuition, has a relatively low SAT 25% value, as well as relatively low retention rate, graduation rate and percentage of class < 20 students. Everything added up together makes it an outlier.
- Wabash College has a relatively low SAT 25% compared to schools with similar a SAT 75%. It also has a relatively low graduation rate.
- Hollins University has a relatively low retention rate and very low graduation rate.
- Morehouse College charges a relatively low tuition and has a relatively low graduation rate and % of class < 20 students.

Predicted ranking is higher than actual ranking:

- Gordon College has a relatively low acceptance rate and high graduation rate.
- Salem College has a high percentage of classes with 20 students or fewer.
- American Jewish University has a very high retention rate and a relatively high percentage of classes with 20 students or less.

*relatively means comparing to the schools with similar ranking in the U.S. News' data.