Predicting and Maximizing Reputation Scores in U.S. News and World Report: An Analysis of 50 Research Universities

For Rutgers School of Nursing

Stephen Chemsak

EXECUTIVE SUMMARY

College and university leaders often are interested in enhancing the reputations of their institutions to advance student recruitment initiatives, increase external funding, and improve research visibility. Using regression analysis and U.S. News and World Report data, this short paper investigates acceptance rates and reputation scores of universities across the United States. The analysis determines that there is a strong relationship between university reputation and acceptance rates. Universities with high reputation scores tend to have low acceptance rates and vice versa. Reputation scores can be predicted based on acceptance rates. Predictions about a school's reputation score can be made based solely on its acceptance rate. The accuracy of the predictions is demonstrated to vary depending on the difference between a particular school's acceptance rate and that of the group average. The precision of the prediction range becomes greater the closer one moves toward the average acceptance rate for schools in the sample; the average acceptance rate in the present case is approximately 53%. For a school close to the group average, any predictions for it likely would be robust.

Schools that would like to predict their reputation such as Penn State may draw on the techniques of regression and confidence intervals used in this analysis and the most recent U.S. News data available. In addition, in conjunction with other information, they may use such techniques in the development of more advanced "causal models" such as structural equation models, which could inform an understanding of which factors, and to what extent, might be most important to manipulate in attempts to increase their reputation scores.

OVERVIEW

In this memo, a technique known as simple linear regression is used to answer the question of whether there is a relationship, on average, between U.S. private research universities' reputations and one important variable that is controllable from an institutional policy standpoint, acceptance rates. I use a random sample of 50 universities from a U.S. News and World Report rankings list. In addition, I examine a set of "less selective schools" not included in my sample – the University of Hartford, DePaul University, Florida International University, and Adelphi University – and make a prediction on the average reputation of such schools, based solely on their common acceptance rate. Furthermore, I construct a "95% confidence interval" that would almost certainly cover their actual reputation score values. Finally, I answer the question of whether a university's being "very selective," "moderately selective," or "less selective" makes a difference, and if so, to what degree, in the accuracy of predictions I make about a school's reputation based solely on knowledge of its acceptance rate.

The reputation scores in the sample of universities studied in the present analysis ranged from 10 (lowest) to 50 (highest)[1], are denoted by the variable REPUT, and are based on U.S. News' 2001 rankings.[2] Acceptance rates (0-100%), taken from the same

[1] For the purposes of simplification, the original scores of 1-5 were multiplied by 10.
[2] The most complete, available data set, provided by the Harvard Graduate School of Education.

rankings, are reflected in the variable ACCEPT. When each university's reputation (REPUT) is plotted against its acceptance rate (ACCEPT), a moderately strong, negative, and apparently linear relationship between the two variables is evident. In other words, high reputation scores tend to be associated with low acceptance rates, and vice versa. For example, Harvard has a very high reputation score (49) and a very low acceptance rate (11) compared to most of the other schools. Clarkson, on the other hand, has a very low reputation score (26) and a very high acceptance rate (83). Some schools fit the pattern but not as well as Harvard and Clarkson. For example, the University of Chicago has a very high reputation score but a mid-range acceptance rate (44). Andrews has the lowest reputation score (20) but a mid-range acceptance rate (54). The average reputation score, across schools, is calculated to be 34.5.

Despite variation in the sample, because it has been randomly selected, predictions about the population of schools from which the sample has been drawn can be attempted by fitting a "least-squares regression model" line to the data. A "linear" model was chosen because it captures the data's relationship in a simple, straightforward manner. Such a model summarizes the relationship between reputation scores and acceptance rates more accurately and concisely than, for example, a curve would. A line that minimizes the vertical distances between itself and the data points was calculated using the statistical programming package SAS (Statistical Analysis System) and is reflected in the following equation:

Regression Model

$$\text{REPUT}\hat{} = 50.50 + -.31 * \text{ACCEPT}.$$ [3]

In this equation, .31 is the "predictor coefficient" (or "slope"). 50.50 is the "y-intercept," or point where ACCEPT = 0. The negative sign denotes that low rates of acceptance (ACCEPT) tend to be associated with high rates of reputation (REPUT). The equation reveals that a one-point difference in acceptance rate is, on average, associated with a .31 difference in reputation score. For example, inserting ACCEPT = 20 into the equation yields REPUT = 44.3; when ACCEPT = 80, REPUT = 24.8. In other words, when a school's acceptance rate is 20%, we would predict a reputation score of 44.3 (relatively high); when it is 80%, we would predict a reputation score of 24.8 (relatively low).

Note that using this equation or "least-squares regression line," reputation cannot be predicted based on acceptance rate with the same success for all universities in the sample. Many of the data points have larger or smaller vertical distances to the model regression line than others. Unlike in the model, actual acceptance rates and reputation scores have a relationship that is not constant. Such variation is not unexpected. Different universities, like different companies, have different reputations and, in the case of companies, different numbers of job offers per number of candidates. The linear equation attempts merely to show what is the case on average.

[3] REPUT^ denotes average reputation and * denotes multiplication. The values have been rounded to the nearest hundredth decimal place.

The vertical distance of the data points to the regression line is captured by an "R-square statistic." If all of the variation in reputation were explained by acceptance rate, R-square would equal 1. If none of the variation in reputation were explained by acceptance rate, R-square would equal 0. Based on the R-square calculated for this regression line using SAS, seventy-three percent of the variation in reputation scores can be explained by acceptance rates (R-square = .73). Twenty seven percent of the variation cannot be explained by acceptance rates. Greater variation may be explained by other factors like school revenue, location, etc. that are not included in the equation. There is no implication in the present analysis that differences in reputation cause differences in acceptance rates. However, there is clearly a relationship of association between the two variables.

The relationship between universities' reputation scores and their acceptance rates is furthermore "statistically significant." The evidence for this conclusion is based on a "t-value," and "p-value." Using SAS, the predictor coefficient's t-value is calculated to be 32.55, which is a relatively large value, and the p-value .0001, which is relatively small. The relationship is statistically significant because the p-value, the probability that we would have gotten a t-value this large if the hypothesis that there were no relationship is true, is less than .05, which is the conventional value below which the hypothesis that there is no relationship between the two variables is rejected. In other words, we can infer with 95% confidence, that in the population of schools from which the sample was selected, there is a relationship between REPUT and ACCEPT. To put this yet another way, it is highly unlikely that a population in which reputation and acceptance rates had no relationship gave rise to the data in the sample.

"Statistical significance" in this case means the following in non-statistical terms: based on an analysis of the reputation and acceptance rates of 50 American private universities, on average, acceptance rates and reputations of American private universities are associated with each other. In general, universities with low acceptance rates have high reputation scores, and those with high acceptance rates have low reputation scores. The data on some universities does not exhibit this pattern; on the whole, however, this analysis provides strong evidence that there is a relationship between the two variables not only in the sample, but in all American private research universities.

The University of Hartford, DePaul University, Florida International University, and Adelphi University are "less selective" schools admitting 70% of their applicants; they are not represented in the sample drawn. However, I am able to predict that the average reputation across all schools that admit 70% of their applicants is 29.14, by inserting 70 into the regression equation.

Calculation 1: Average REPUT at ACCEPT = 70[4]

$$\hat{y}_i = (-.3052)\,(70) + 50.5031$$
$$\hat{y}_i = 29.1391$$
$$\hat{y}_i = 29.14 \quad \text{(by rounding)}$$

[4] Values in first equation left to thousandth decimal place.

Furthermore, across repeated random samples of American research universities, 95% of confidence intervals constructed, such as the one in Calculation 2 below, will cover the true value of the average reputation of all U.S. research universities with acceptance rates of 70%; 5% of the confidence intervals will not. The confidence interval constructed based on the sample drawn is 27.54 to 30.74. It is highly likely that this confidence interval covers the true value of the average reputation of all U.S. research universities with acceptance rates of 70%.

Calculation 2: 95% Confidence Interval for Average REPUT at ACCEPT = 70[5]

$$29.14 \pm t_{n-2}\,\text{RMSE}\sqrt{\left[\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right]}$$

$$29.14 \pm 2.01\,(4.56469)\sqrt{\left[\frac{1}{50}+\frac{(70-52.56)^2}{28918.32}\right]}$$

$$29.14 \pm 9.1750\sqrt{\left[.02+\frac{304.1536}{28918.32}\right]}$$

$$29.14 \pm 9.1750\sqrt{\left[.02+\frac{304.1536}{28918.32}\right]}$$

$$29.14 \pm 9.1750\,(0.1747)$$

$$29.14 \pm 1.6028725$$

$$29.14 \pm 1.60$$

$$(27.54,\ 30.74)$$

Finally, given the confidence intervals for the other types of schools – "very selective" and "moderately selective" – it is clear that the confidence intervals differ. The three intervals are similar in that they are all centered on the predicted average value of reputation based on their respective rates of acceptance (table below). However, the

[5] RMSE = Root Mean Squared Error. n=sample size. t_{n-2} is t with n-2 degrees of freedom. x_i = acceptance rate of 70. $\bar{x}$ = average acceptance rate.

intervals are different in that they are not the same size. Schools with acceptance rates of 30% have the largest confidence interval range (3.56). Schools with acceptance rates of 70% have the second largest confidence interval range (3.20). Schools acceptance rates of 50% have the smallest confidence interval range (2.61).

Table: Average Reputation 95% Confidence Intervals at Given Acceptance Rates

ACCEPT	*RÊPUT*	*RÊPUT* 95% Confidence Interval	Range
30% (very selective)	41.35	39.57, 43.13	3.56
50% (moderately selective)	35.24	33.94, 36.55	2.61
70% (less selective)	29.14	27.54, 30.74	3.20

The differences in these intervals reflect the fact that the further a school's acceptance rate is from the average acceptance rate, the larger the confidence interval becomes around the regression line. In other words, the precision of the prediction range becomes greater the closer one moves toward the average acceptance rate for schools in the sample; the average acceptance rate in the present case is approximately 53%.

In summary, reputation scores and acceptance rates at U.S. private research universities tend to be related to each other. Universities with high reputation scores tend to have low acceptance rates and vice versa. Reputation scores can be predicted based on acceptance rates. Schools that want to predict their reputation may draw on the techniques of regression and confidence intervals used in this analysis and the most recent U.S. News data available. In addition, in conjunction with other relevant information, they can use such techniques in the development of more advanced "causal models," which could inform an understanding of which variables, and to what degree, would be important to manipulate in attempts to increase their reputation score.