Have You Been Accepted Yet?: A Study of College Acceptance Rates

Introduction

The college application process is stressful, exciting, and difficult for the majority of high school seniors. Students dedicate hours of their time to present their best side to the various college admissions officers that read both their application and the millions of others that arrive at the admissions office around the world. Even so, a student’s best may not be enough. More people are applying and going to college every year, creating an increasing number of qualified applicants, thus allowing colleges to grow increasingly more selective throughout the years to some of the lowest college acceptance rates in history. During a conversation with a fellow classmate, I was told that one of the Ivy League schools that he was applying to had acceptance rates of less than six percent. Later on, I decided to compare the acceptance rates to the colleges to which I applied. I saw that the two Ivy Leagues that I applied to had considerably lower acceptance rates than any of the other schools. The Ivy League schools are known as some of the top universities in the United States, causing me to believe that there was a positive correlation between the quality of a college and an acceptance rate.

In order to answer this question, I will compile my raw data of college rankings and the acceptance rates which I will then organize into a data table. From this data, I will create a scatter plot and a frequency table. Next, I will calculate my line of best fit and the correlation coefficient from my scatter plot to find the relation between the two factors. With the data I acquire, I will perform a χ2 test in order to display the dependence of undergraduate acceptance rates on college ranking.

Collecting Data

Since I was trying to find the correlation between the two factors, both variables had to be quantitative. The quality of a college is hard to consistently measure accurately on a quantitative scale, so I decided to use college ranking as my measuring factor instead of quality. While finding my data for both the college ranking and undergraduate acceptance rate, I tried to make it more accurate by the using the same website for each set of data: using U.S. News for the college rankings and acceptancerate.com for the acceptance rates. This would allow my data to be decided on one set of materials and factors, making it more precise. Although I used this method to prevent error, it still caused error in other ways. The U.S. News website had multiple sets of rankings that separated schools into National Universities, National Liberal Arts Colleges, Regional Universities, and Regional Colleges. I only used the National Universities because there was no possible way to combine the lists to make a more complete and accurate list. Since each list did not encompass all universities and colleges the rankings are a skewed, so the differences in the rankings and, in turn, the admission rates can cause the data to be more linear than it actually is. Similarly, the website used for the acceptance rates for the math was a secondary site and the information was not directly from the universities in which I am studying. Using information from the universities’ actual websites would have been more accurate, but some of the universities did not openly list their acceptance rates and I could not find data from the same year on the university websites. Finally, I noticed when I created my raw data table, which can be found in Figure 1 of the appendix, the college rankings are not continuous since some of the colleges have the same ranking, causing gaps in between certain numbers. This shifts the data slightly so that the data is less linear.

Linear Regression

When I created my scatter plot, I placed the college ranking on the x-axis and the acceptance rates on the y-axis. I thought my data would have an unusual representation since the lower college rankings, numerically, correspond to higher performance levels by the college, but that was not the case. Instead, the data points created a weird pattern. The higher ranked universities had very low acceptance rates that formed a very linear pattern. As the rankings got lower, the percentages for the acceptance rates began to vary, so the range between the highest and lowest acceptance rates at nearby rankings began to increase. The fanning out of the data caused me to be unsure if the Pearson’s correlation coefficient would say if the scatter plot data had a strong enough relation.

I then solved for Pearson’s correlation coefficient on my calculator (Figure 2) which stated that r had a value of 0.820. Since the relationship was not perfect or strong and the relationship was so varied at towards the lower rankings, I decided to check to see if the relationship was quadratic (Figure 3). The value for r2 was equal to 0.737, since 0.820 is greater than 0.737, I concluded that the relation between the university ranking and undergraduate acceptance rate was linear. I then used the values for a and b to graph the line of best fit which was equal to 0.596x+13.477.

χ2 and Yates’ Continuity Correction

Since the χ2 test determines independence, I created the null and alternative hypothesis.

H0 = College ranking and undergraduate acceptance rate are independent.

H1 = College ranking and undergraduate acceptance rate are dependent.

I then used my raw data to build a frequency table and an expected contingency table. To calculate the expected frequency, I multiplied the corresponding row and column totals and divided by the total sum of the observed values. When I was calculating my expecting frequencies, I noticed that the values were not near what I had observed through my research, hinting me that my variables would be dependent on each other.

*Frequency Table*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Undergraduate Acceptance Rate** | |  |
| 0.0% - 49.9% | 50.0% - 99.9% | **Sum** |
| **College Ranking** | 1-50 | 45 | 16 | *61* |
| 51-100 | 7 | 34 | *41* |
| **Sum** | | *52* | *50* | ***102*** |

p = 0.01 d.o.f. = 1 critical value = 6.63

*Expected Contingency Table*

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Undergraduate Acceptance Rate** | |
| 0.0% - 49.9% | 50.0% - 99.9% |
| **College Ranking** | 1-50 | (61/102)(52) = 0.598(52) = 31.1 | (61/102)(50) = 0.598(50) = 29.9 |
| 51-100 | (41/102)(52) = 0.402(52) = 20.9 | (41/102)(50) = 0.402(50) = 20.1 |

I then calculated χ2 by subtracting the expected frequency from the observed frequency, squaring the difference, and dividing that value by the expected frequency. While I was calculating χ2 I used a p-value of 0.01 which had a critical value of 6.63, corresponding to a degree of freedom of one. In order to make all of my expected frequencies have a value of at least a value of five, I had to make my table a 2x2, thus the degree of freedom of 1. With this data I attain a χ2 value of 31.54.

χ2 = (45-31.1)2 + (7-20.9)2 + (16-29.9)2 + (34-20.1)2 = 6.21 +9.25 + 6.46 + 9.62 = 31.54

31.1 20.9 29.9 20.1

Since the χ2 value of 31.54 is greater than critical value of 6.63, the null hypothesis is incorrect, meaning college ranking and undergraduate acceptance rate are dependent.

To fix this issue with my degree of freedom and make my χ2 value more accurate, I used Yates’ Continuity Correction. I calculated χ2 with the continuity correction by subtracting 0.5 from the difference of the expected frequency and the observed frequency, squaring the differnce, and dividing that value by the expected frequency. After redoing the χ2 test with Yates’ Continuity Correction, I attained a χ2 value of 29.31.

χ2Yates= (|45-31.1|-.5)2 + (|7-20.9|-.5)2 + (|16-29.9|-.5)2 + (|34-20.1|-.5)2 = 5.78+8.59+6.01+8.94

31.1 20.9 29.9 20.9

= 29.31

Since the χ2Yates value of 29.31 is greater than the critical value of 6.63, the null hypothesis is incorrect, meaning college ranking and undergraduate acceptance rate are dependent.

In both instances my χ2 value was extremely larger than my critical value, so I was able to fully reject my null hypothesis which stated that university rating and college acceptance were independent values.

Conclusion

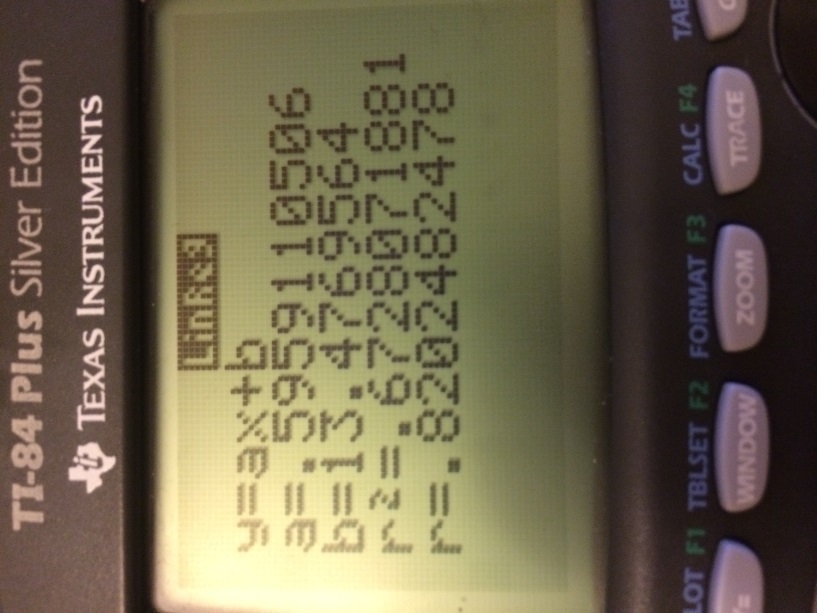
Through my investigation, I proved that there was a moderate positive correlation between college acceptance rates and college ranking. This means that as the college ranking increases the acceptance rate will be higher, but this will not always be the case and the values may fluctuate throughout the graph. This means that the colleges with rankings closer to one are harder to get into percentage-wise. Through the data I’ve collected, I believe that I can properly say that those two factors are dependent on each other since the χ2Yates value is higher than the critical value. Even though my frequency tables had degrees of freedom of one, the value of χ2 were over three times that of the critical value, so this dependence is verified. I should have also looked at the number of applicants and the number of accepted first-year students as well in this experiment so that I could determine if the higher ranked schools had an increased number of applicants which would have skewed the acceptance rate data. My sample population was of a decent size, but I would have liked to include liberal arts and regional universities in my data as well, but there were no college ranking it systems that compared them on an equal basis. Although this experiment is flawed, it gives some insight into the admissions process. Through the moderately positive correlation, shows that the higher ranked colleges are more selective when choosing their undergraduate applicants. By being more selective, they are receiving the best applicants for their school, making them more inclined to have a successful school through successful students who are willing to work harder and more intelligent.

Appendix

**(Figure 1)**

|  |  |  |
| --- | --- | --- |
| College Ranking  1 | University  Princeton Univ. | Acceptance Rate  7.9 |
| 2 | Harvard Univ. | 5.8 |
| 3 | Yale Univ. | 7.7 |
| 4 | Columbia Univ. | 6.9 |
| 4 | Stanford Univ. | 6.6 |
| 4 | Unvi. Of Chicago | 13.2 |
| 7 | MA Institute of Tech. | 8.9 |
| 8 | Duke Univ. | 14.0 |
| 8 | Univ. of PA | 12.6 |
| 10 | CA Institute of Tech. | 12.8 |
| 11 | Dartmouth College | 10.1 |
| 12 | Johns Hopkins Univ. | 18.4 |
| 13 | Northwestern Univ. | 18.0 |
| 14 | Washington Univ. in STL | 17.9 |
| 15 | Cornell Univ. | 16.6 |
| 16 | Brown Univ. | 9.6 |
| 16 | Univ. of Notre Dame | 23.3 |
| 16 | Vanderbilt Univ. | 14.2 |
| 19 | Rice Univ. | 16.7 |
| 20 | Univ. of CA - Berkeley | 21.6 |
| 21 | Emory Univ. | 26.7 |
| 21 | Georgetown Univ. | 18.1 |
| 23 | Univ. of CA - LA | 25.5 |
| 23 | Univ. of VA | 29.5 |
| 25 | Carnegie Melon Univ. | 27.8 |
| 25 | Univ. of Southern CA | 19.9 |
| 27 | Tufts Univ. | 21.4 |
| 27 | Wake Forest Univ. | 34.0 |
| 29 | Univ. of MI - Ann Arbor | 36.6 |
| 30 | Univ. of NC - Chapel Hill | 33.0 |
| 31 | Boston College | 28.8 |
| 32 | NY Univ. | 35.0 |
| 33 | College of William and Mary | 32.2 |
| 33 | Univ. of Rochester | 35.3 |
| 35 | Brandeis Univ. | 39.1 |
| 35 | GA Institute of Tech. | 51.2 |
| 37 | Univ. of CA - San Diego | 35.3 |
| 38 | Case Western Reserve Univ. | 54.3 |
| 38 | Univ. of CA - Davis | 48.3 |
| 40 | Lehigh Univ. | 32.6 |
| 40 | Univ. of CA - Santa Barbara | 46.3 |
| 42 | Boston Univ. | 49.4 |
| 42 | Northeastern Univ. | 31.9 |
| 42 | Rensselaer Polytechnic Institute | 43.6 |
| 42 | Univ. of CA - Irvine | 47.5 |
| 42 | Univ. of IL - Urbana-Champaign | 63.3 |
| 47 | Univ. of WI - Madison | 68.6 |
| 48 | PA State Univ. - Univ. Park | 52.4 |
| 48 | Univ. of FL | 44.1 |
| 48 | Univ. of Miami | 39.7 |
| 48 | Univ. of Washington | 58.4 |
| 48 | Yeshiva Univ. | 84.0 |
| 53 | Univ. of TX - Austin | 46.7 |
| 54 | George Washington Univ. | 33.0 |
| 54 | OH State Univ. - Columbus | 64.0 |
| 54 | Pepperdine Univ. | 38.4 |
| 54 | Tulane Univ. | 27.3 |
| 58 | Fordham Univ. | 42.9 |
| 58 | Southern Methodist Univ. | 53.8 |
| 58 | Syracuse Univ. | 51.3 |
| 58 | Univ. of CT | 44.7 |
| 62 | Brigham Young Univ. - Provo | 54.9 |
| 62 | Clemson Univ. | 63.3 |
| 62 | Purdue Univ. - West Lafayette | 61.5 |
| 62 | Univ. of GA | 63.0 |
| 62 | Univ. of MD - College Park | 46.6 |
| 62 | Univ. of Pittsburgh | 57.9 |
| 68 | TX A&M Univ. - College Station | 66.9 |
| 68 | Worcester Polytechnic Institute | 52.6 |
| 70 | Rutgers | 60.9 |
| 71 | American Univ. | 44.2 |
| 71 | Baylor Univ. | 60.7 |
| 71 | Univ. of IA | 78.4 |
| 71 | Univ. of MN - Twin Cities | 49.5 |
| 71 | VA Tech | 70.4 |
| 76 | Clark Univ. | 70.0 |
| 76 | Indiana Univ. - Bloomington | 74.4 |
| 76 | Marquette Univ. | 55.2 |
| 76 | Miami Univ. - Oxford | 72.8 |
| 76 | Stevens Institute of Tech. | 40.4 |
| 76 | SUNY College of Environmental Science and Forestry | 50.7 |
| 76 | TX Christian Univ. | 40.9 |
| 76 | Univ. of DE | 59.4 |
| 76 | Univ. of MA - Amherst | 62.6 |
| 85 | MI State Univ. | 70.6 |
| 85 | Univ. of CA - Santa Cruz | 67.3 |
| 85 | Univ. of VT | 76.7 |
| 88 | Binghamton Univ. - SUNY | 40.7 |
| 88 | CO School of Mines | 44.6 |
| 88 | Stony Brook Univ. - SUNY | 40.1 |
| 88 | Univ. of AL | 53.1 |
| 88 | Univ. of CO - Boulder | 83.6 |
| 88 | Univ. of Denver | 67.6 |
| 88 | Univ of Tulsa | 40.7 |
| 95 | Drexel Univ. | 74.9 |
| 95 | FL State Univ. | 53.7 |
| 95 | NC State Univ. - Raleigh | 52.6 |
| 95 | Univ. of San Diego | 42.6 |
| 99 | St. Louis Univ. | 64.4 |
| 99 | Univ. of MO | 81.5 |
| 99 | Univ. of NE - Lincoln | 64.4 |
| 99 | Univ. of NH | 77.9 |

**(Figure 2)**

****

**(Figure 3)**

