

# **Analysis of the College Admissions Process Through Family Income**

Nick Mitchell

May 3, 2022

## Introduction

For this project, I was interested in the college admissions process in the United States. Over the past several decades, the college admissions process has drastically changed; college admissions have become far more competitive and the cost of college has become one of the biggest investments one will ever make in their lifetime. Recently, there has been extreme controversy regarding the college admissions process and how it “is a system that is already so distorted by money and privilege...” according to US District Judge Indira Talwani. The college admissions scandal of 2019, otherwise known as “Operation Varsity Blues”, shined light on such distortion and the corrupt college admissions system in America. This scandal included hundreds of the wealthiest families in America, including celebrities like Lori Loughlin, who ultimately cheated the admissions process by paying extremely large sums of money to pay off coaches, admissions officers, and standardized test proctors to increase their child’s chances of getting into a more prestigious and selective college, even though they may not actually be qualified for the school. The college admissions process may be unfair to lower income families after such a scandal. Lower income families do not have access to the same resources and capabilities as high income and upper class families, which extremely skews the chances of a lower income student getting into a more prestigious university. I wanted to focus on the financial backgrounds of students and where they are getting accepted, so I proposed these two questions:

1. Does wealth/income impact a prospective college student’s chances of admission to a more selective/prestigious college or university?
2. Do selective colleges accept a higher/lower percentage of students based on their family’s income?

After running a few different regressions, all with the independent variables of parental median income, parental mean income, and rejection rate, there was an apparent relationship between higher income families and the types of universities that students are going to. The dependent variables I looked at were college tier, college type, iclevel (number of years for the university/college), and tuition price.

## Background

The college admissions scandal in 2019 highlighted the extreme measures some will take to better their chances of getting into a more prestigious college. It began to raise more questions regarding the resources that higher income families have compared to lower income families for applying to college. Families with higher income are more likely to live in affluent areas where the public school systems tend to be better or they have the ability to send their children to private schools. Students from higher income families are also more likely to participate in extracurricular activities with more time and financial support. In regards to the actual application process, students from higher income families have better

access to tutoring for the SAT/ACT exams, college admissions counselors, as well as college essay tutors. Top tier colleges are known to look for prospective students who have high GPAs, high standardized test scores, a great essay, and are active in their community through extracurricular activities. Having access to better resources gives students a much greater chance of getting into a more selective and prestigious college.

I was interested in looking into the difference in access to such resources between high income and lower income families, however, my datasets would not necessarily provide us with the best information regarding such variables. Instead, I decided it would be more beneficial to look at the income of families and find the relationship between variables such as college tier (how selective/prestigious an institution is), the type of college (public vs. private), IC level (how many years the university is), and tuition price of the school. These variables will help determine whether or not a relationship exists between income level and where students are being admitted.

## **Data**

I utilized two datasets containing data on children who attended college between the ages 19 and 22 in the early 2000s up to 2013, by individual college. The data comes from Opportunity Insights, a group based at Harvard University collecting research in an effort to identify barriers to economic opportunity and help develop solutions to empower citizens to rise out of poverty. The first dataset contained information focused around the families of students attending universities, reporting baseline estimates of parents' income distribution by college. This dataset also includes a row for those who did not attend college between the ages of 19-22. The second dataset was more focused around each college's enrollment, reporting selected characteristics drawn from the Department of Education's database and the College Scorecard in 2013. I was able to join the two datasets together in R to analyze income distributions by the college statistics selected in the second dataset.

Before I get into the data, some of the limitations of the datasets must be identified. The biggest limitation of the datasets was the time period, as the most recent data available from Opportunity Insights comes from 2013. Another limitation of the data is the exclusions of international data. The data also only contains information on American families, another limitation due to the large population of international students attending universities in America. So, it is important to keep in mind the data is not entirely representative of a college's actual student population. Also, the first dataset contained data on roughly 2200 colleges while the second contained 2464 colleges, so data for 264 individual schools was lost upon joining the datasets. While the estimated number of colleges in the US varies, most estimates circle around 5,300 schools in the US, with 3,982 degree-granting postsecondary institutions in the U.S. The dataset may or may not account for some of the missing schools data via the variable "multi". In

some cases, multiple colleges share a single identifier in their tax records, for example the multiple campuses of University of Massachusetts or Babson's Wellesley and Miami campuses. These cases are grouped together, denoted by having a 1 value for "multi" and assigned locations based on the largest school/campus in the group, so it is difficult to estimate exactly how many colleges the dataset contains information on.

With that said, the independent variables I chose to look at were type, tier, iclevel, sticker\_price\_2013, par\_top1pc. Type refers to whether the institution is public, private non-profit, or for profit. Tier refers to the selectivity status of the school, from Ivy to schools that require less than two years, as seen below. Iclevel refers to the college's level of degree offering: four year, two year, or less than two year. Sticker\_price\_2013 refers to the average annual cost of attendance, including tuition and fees, in the year 2013. Par\_top[percentile]pc refers to the fraction of parents in the top selected percentile of the income distribution. For example, Par\_top1pc is the fraction of parents in the top 1% of the income distribution.

tier	Selectivity and type combination (see Table 6 for more detailed descriptions of these groups): 1 = Ivy Plus 2 = Other elite schools (public and private) 3 = Highly selective public 4 = Highly selective private 5 = Selective public 6 = Selective private 7 = Nonselective 4-year public 8 = Nonselective 4-year private not-for-profit 9 = Two-year (public and private not-for-profit) 10 = Four-year for-profit 11 = Two-year for-profit 12 = Less than two year schools of any type 13 = Attending college with insufficient data 14 = Not in college between the ages of 19-22
------	--

The dependent variables I wanted to look at to see the impact of the variables described above were par\_mean, par\_median, and scorecard\_rej\_rate\_2013. Par\_mean is the mean parental income at the college. Par\_median is the median parent household income rounded to the nearest hundred dollars. Scorecard\_rej\_rate\_2013 is the rejection rate in 2013, obtained from College Scorecard.

Lastly, the second dataset did contain data on minority's share of enrollment, such as asian, black, and hispanic share of enrollment. While I originally wanted to dive into these disparities, this data was only from the year 2000. The U.S. and more importantly the college admissions process have both radically changed since the year 2000, so I did not run regressions based on these variables.

## Method

For each set of regressions, I selected all 3 dependent variables (parental mean and median income, scorecard rejection rate) as explained by each independent variable individually. First, I wanted

to look at the dependent variables as explained by the selectivity of the college. For this regression I first had to develop dummy variables for all 14 levels of college. Next, I wanted to look at the dependent variables explained by college type. Again, I had to create dummy variables for all 3 college types. Next I choose IC level, creating dummy variables for the 3 levels of institutions. I wanted to look at the same variables explained by sticker price. For this set of regressions, I had to switch sticker price and rejection rate as the dependent and independent variable, as the coefficient results were smaller than 3 decimal places before switching the variables. I also looked at parental income distributions. For this data, I only looked at rejection rate as a dependent variable, using the top 0.1%, 1%, and 5% income percentiles.

## Results

The regressions that I ran in R gave insightful results. When looking at the tier of the university, the regression showed that as the tier, and therefore selectivity, lowered so did all three of the independent variables, as seen below:

	meanincomebytier	medianincomebytier	rejectionratebytier
(Intercept)	72905.271+	50500.000**	0.219***
	(42375.445)	(18947.900)	(0.025)
Ivy	358659.988***	124841.667***	0.689***
	(44105.762)	(19721.599)	(0.051)
Elite	236199.888***	103287.500***	0.502***
	(42705.220)	(19095.356)	(0.031)
HighlySelective	140815.198***	71607.216***	0.242***
	(42593.316)	(19045.319)	(0.029)
Selective	39698.673	33648.905+	0.104***
	(42397.533)	(18957.776)	(0.025)
Nonselective	29331.122	22211.111	0.128***
	(42513.702)	(19009.720)	(0.030)
Twoyearorless	550.427	10900.131	0.054
	(42403.169)	(18960.296)	(0.044)
Forprofitfittier	79.495	5714.667	
	(42516.462)	(19010.955)	
notincollegeby22	-20970.995	-11250.000	

The intercept values of 72905.271 and 50500.000 mean that the mean and median average income of all rows in the dataset are \$72,905 and \$50,500 respectively. The coefficients for Ivy tier are 358659.988 and 124842.667, which means the parental mean and median income increases by \$358,659 and \$124,842 as compared to average, respectively. The coefficients for not in college are -20970.995 and -11250.000,

which means that the parental mean and median income decreases by \$20,970 and \$11,250 as compared to the average, respectively. For both of these dependent variables, tiers Ivy, elite, and highly selective had a statistically significant correlation with mean and median income. For the rejection rate dependent variable, the intercept of 0.219 shows that the average rejection rate across all schools was 21.9%. The coefficient of 0.689 for Ivy means that for schools of this tier, the rejection rate increases by 68.9%, making it 90.8%. Tiers for profit and not in college cannot be explained by rejection rate; those not in college are not getting rejected and for profit colleges are pay-to-learn, so they do not reject any students. All tiers except the two above as well as two year or less schools have a statistically significant correlation with rejection rate.

The next regression I ran was on college type, and the results are shown below.

	<b>meanincomebytype</b>	<b>medianincomebytype</b>	<b>rejectionratebytype</b>
(Intercept)	72464.402***	55827.368***	0.214***
	(4306.768)	(1791.774)	(0.030)
public1	12218.921**	13195.489***	0.123***
	(4637.858)	(1929.520)	(0.031)
privatenonprof	76312.431***	39669.579***	0.163***
	(4780.299)	(1988.780)	(0.031)

The intercept values of 72464.402, 55827.368, and 0.214 align closely with the results found by the college tier. As tier and type are similarly segregating variables, this is a reassuring consistency in the data. Both variables for public and private non profit colleges have a statistically significant relationship with mean income, median income, and rejection rate.

Regression based on IC level shown below:

	<b>meanincomebyiclevel</b>	<b>medianincomebyiclevel</b>	<b>rejectionratebyiclevel</b>
(Intercept)	56753.580***	40976.790***	-0.156
	(11850.130)	(4796.353)	(0.198)
fouryear	72209.641***	47375.741***	0.518**
	(11966.977)	(4843.647)	(0.199)
twoyear	5643.562	9331.543*	0.289
	(9367.116)	(3791.351)	(0.193)
Twoyearorless	11398.351	11653.644***	0.156*
	(7613.519)	(3081.580)	(0.066)

This regression showed four year colleges have a statistically significant correlation with parental mean, all levels have a significant relationship with median income, and schools that require two years or less have a slightly significant relationship with rejection rate.

The next set of independent variables I looked at was having a parent in the top 0.1, 1, and 5 percent of the income distribution. The results are shown below.

	<b>rejectionrateby.1%</b>	<b>rejectionrateby1%</b>	<b>rejectionrateby5%</b>
(Intercept)	0.314***	0.300***	0.274***
	(0.006)	(0.006)	(0.007)
par_toppt1pc	16.846***		
	(1.004)		
par_top1pc		2.377***	
		(0.138)	
par_top5pc			0.794***
			(0.048)

The coefficient of 16.846 means that for each 100 percent increase in the percentage of parents in the top 0.1% of income distribution, rejection rate increases by about 16%, on average. All three percentiles have a statistically significant relationship with rejection rate.

The regression below looks at the college's sticker price.

	<b>meanincomebystickerprice</b>	<b>medianincomebystickerprice</b>	<b>stickerpricebyrejectionrate</b>
(Intercept)	55410.955***	57010.924***	16768.831***
	(1651.018)	(696.015)	(793.993)
sticker_price_2013	3.412***	1.387***	
	(0.079)	(0.033)	
scorecard_rej_rate_2013			18031.398***
			(1957.184)

Sticker price has a significant relationship with mean and median income, and rejection rate has a significant relationship with sticker price.

Lastly, and most importantly, I ran a regression for parental mean income by rejection rate.

<b>Par_mean</b>	
(Intercept)	70463.113***
	(4348.406)
scorecard_rej_rate_2013	177283.999***
	(10748.829)

Scorecard rejection rate has a significant relationship with parental mean income.

### Discussion

The college tier regression gave impactful results, confirming many assumptions I had when building the research questions. As seen in the results, the coefficients that explain both mean and median income show an incremental decrease as the selectivity of the school decreases. I expected this to be the case, but seeing the cleanly segregated incremental increase caught my attention. It is especially interesting to note that for those not in college, their parental mean and median income was below the average. This could be because of two reasons: the family cannot afford to send their child to college, or similarly because the student is either forced to enter the workplace right after highschool to help their family or, chooses to for any number of other reasons that I cannot measure. Furthermore, the regression showed that highly selective, elite, and ivy level schools have a statistically significant relationship with mean and median income, confirming my hypothesis that families of higher incomes typically end up at colleges of higher selectivity.

The income percentile regressions show just how significant parental income distributions can affect rejection rates. I looked at the top 0.1, 1, and 5 percent of the income distribution. It was eye opening to see the coefficients for each variable, especially as they all have a significant relationship with rejection rate. For each 100 percent increase in the percentage of parents in the top 0.1, 1, and 5% of income distribution, rejection rate increases by about 16, 2, and 0.8% on average. As mentioned above, schools of higher prestige have higher rejection rates. Connecting that fact to these regressions, having more parents within higher percentile income distributions at a school means the school is more prestigious.



I can connect this back to the original research regarding the access to resources that higher income families have compared to those of lower income, such as better education, tutoring, and extracurricular activities. However, these are variables that I did not actually look at and run regressions on, so that is a limitation to my conclusions because I cannot actually determine a relationship between these variables. Regardless, this highlights one of the biggest problems facing society regarding college admissions. If the highest tier universities are accepting more students who come from affluent families compared to lower income families, this only leaves the lower income students to even consider going to less selective/prestigious schools. In order to fix this problem, we need to look to fix the college admissions process as a whole. Nowadays, more students are applying to colleges than ever before, yet acceptance rates for top tier colleges are lowering. This is an issue because it shows that less students can actually get into a top tier college and the first to be rejected are most likely the lower income students. The lower income students again, do not have access to the same resources as more affluent students do, which decreases their chances of getting into a top tier university. Another issue with the college admissions process is that college applications cost money, which some students cannot afford. In 2021, 700,000 students from low income families or who were first generation college students opened the Common App, but never actually filled out an application because of factors such as cost of application or overall tuition costs of schools they wanted to go to.

When using rejection rate as the dependent variable, the results similarly showed the rejection rate decreasing as the selectivity decreases, with an exception between selective and nonselective schools. It is important to note that each tier level aside from less than two year colleges has a statistically significant relationship with rejection rate, so tier of the school significantly explains the rejection rate of the school. This was mainly used to show the increase in rejection rate as selectivity increases, which will be important in the income percentile regression analysis.

Regarding type of college, the regression results again confirmed some assumptions made entering the project. Typically speaking, private colleges are regarded as more prestigious than public schools. For public colleges, mean parental income increases by roughly \$12,000 compared to for-profit schools, while private non profit schools' parental mean increases by roughly \$76,000. Both variables have a significant relationship with parental mean income, confirming the hypothesis that students with higher familial income typically end up at more prestigious universities. As public schools receive federal funding, they are typically less expensive than private schools. This helps explain the disparity in familial income between public and private schools, as families with lower income cannot afford some private schools. On top of this, as seen in the results, public schools have a lower rejection rate compared to private schools. Both variables have a significant relationship with rejection rate, further explaining why families with lower income do not have the same opportunity to attend private school.

For IC level, or the number of years that a student will complete college in, results were similar. Parental mean income increases by \$72,209 compared to average for four year schools, while it only increases \$11,000 or less for all other schools. This is to be expected, as a four year undergraduate college degree is the most expensive type of education, but also tends to be some of the best education you could get. Tuition rates are continuing to rise across the country for both private and public schools and many families struggle to afford such a four year college. This shows that the college system in America is made for the wealthy and privileged. A two year college is much less expensive compared to a four year college, yet the degrees from such colleges tend to be worth “less” than one of a four year college. The rising costs of tuition combined with less prestigious certifications traps families of lower income into a cycle; having low familial income forces them to attend a less prestigious school, this limits their opportunities post graduation, leading them to raise a family with low income and continue the cycle. The IC level regressions also showed public schools have a significant relationship with rejection rate. Four year institutions have higher rejection rates than two year institutions, again emphasizing the point that it is harder to get into more prestigious universities.

The sticker price regressions are impactful in answering the research questions. The coefficient for the first model means that for every dollar increase in sticker price, there is a 3.4 dollar increase in parental mean income. A higher tuition price implies a more prestigious school, so this confirms the hypothesis that students with higher familial income typically end up at schools of higher prestige. For rejection rate, the coefficient of 18031.398 implies that for every 1% increase in rejection rate, sticker price increases by about \$180, on average. As mentioned, a higher rejection rate and a higher sticker price implies a more prestigious university. Scorecard rejection rate has a statistically significant relationship with sticker price, this shows that families with lower income are at a disadvantage when it comes to getting into more prestigious schools, beyond the rejection rate itself is a tuition price these students cannot afford.

In order to solve these problems, four year colleges across America must look to lower their tuition. Each year, it is more common for a college to raise the cost of tuition than it is to lower. In 2021, only 11 schools actually lowered their tuition and in looking at the schools that did so, none of them were a highly selective, elite, or ivy tier university. Some schools have started to look into transparent pricing strategies and tuition caps, which would help make college affordable for more families and not just for the wealthy and privileged. Also, it is important to note that 65% of students in 2021 eliminated colleges based on the price, so lowering tuition should be a priority for admissions.

Lastly, the regression model explaining parental mean income by rejection rate encapsulates the argument. The coefficient of 177283.999 means that for every one percent increase in rejection rate, parental mean income increases by \$1772.83 dollars, on average. As evident in the tier regressions,

rejection rate is directly correlated with how prestigious a school is. As the results show a statistically significant relationship between rejection rate and parental income, this answers the research question; more prestigious colleges accept a higher percentage of students based on a higher familial income.

## **Conclusion**

The college admissions scandal of 2019 began to raise many questions regarding the fairness of the college admissions process in America. This scandal highlighted extreme wealth and cheating, and I became most interested in answering the questions: does wealth/income impact a prospective college student's chances of admission to a more selective/prestigious college or university? Do selective colleges accept a higher/lower percentage of students based on their family's income? After running many regressions, I was able to conclude that wealth and income correlate to the types of universities that students are being accepted to. I found that when a student comes from a more affluent background, they are accepted into more prestigious universities, which tend to be ivy league/elite/highly selective, private, and four year institutions with higher rejection rates (and lower acceptance rates). This is due to a disparity in resources and capabilities surrounding the college admissions process. Higher income families can afford to send their children to better high schools, have tutors for the SAT and ACT exams, and take part in extracurricular activities. These factors play a major role in college acceptance; the more a student has to offer on their application, the more likely they are to get accepted into an institution of higher tier. Overall, the regressions confirmed there is a disparity between the chances a student is accepted into a school based on their familial income. As such, the college admissions process and American education system as a whole must look to change to maintain integrity and equality.

## Sources

- <https://opportunityinsights.org/wp-content/uploads/2018/04/Codebook-MRC-Table-2.pdf>
- <https://opportunityinsights.org/wp-content/uploads/2018/04/Codebook-MRC-Table-3.pdf>
- <https://thebestschools.org/resources/for-profit-vs-non-profit-colleges/>
- <https://www.forbes.com/sites/robertfarrington/2015/07/30/the-simple-strategy-to-stop-rising-tuition-costs/?sh=179f69ae6e7e>
- <https://www.usnews.com/education/best-colleges/articles/how-many-universities-are-in-the-us-and-why-that-number-is-changing#:~:text=When%20it%20comes%20to%20determining,National%20Center%20for%20Education%20Statistics.>
- <https://www.educationunlimited.com/blog/how-many-universities-colleges-are-in-the-us/>
- <https://www.insidehighered.com/admissions/article/2019/03/18/look-many-legal-ways-wealthy-applicants-have-edge-admissions#:~:text=An%20analysis%20by%20the%20Jack,with%20incomes%20less%20than%20%2450%2C000.>
- <https://hechingerreport.org/another-way-college-applications-are-rigged-against-low-income-students/>