

Statistical Analysis Report:
The Effect of Parent-Family Income on Student
Likelihood of Attending College

Kobey Workman

Brigham Young University

ME 275: Computational Methods

Abstract

This report analyzes and discusses the correlation between a student's family income and the likelihood of that student attending college between the ages of 18 and 21. Descriptive statistics and plots reveal the distribution and spread of the data. To analyze the hypothesis, a two-sample t-test is performed, revealing that there is a significantly higher likelihood for students with high-income families (>\$60K/yr) to attend college. Further analysis is conducted concerning the percentage of students per family who attend college, regardless of family income. The hypothesis for this secondary analysis is that more than 55% of all students attend college, regardless of income. A one-sample t-test shows that this hypothesis is accepted. The implications and consequences of these conclusions are then discussed, along with the limitations of this study and discussion of future studies.

Introduction

My family has strong ties with education, since both my mom and dad work in public education. Because of this, I decided to analyze the relationship between a grade school student's family income and the likelihood of that student attending college. My hypothesis for this analysis is that students whose parents have comparatively high income have a higher likelihood of attending college. I will also perform further analysis on the overall percentage of students who attend college to compare and contrast this with the previous question and hypothesis. The results of this study can reveal much about the imbalance in higher education access in the United States, and can be used to better understand how one might go about solving this problem. This report applies not only to prospective college students, but also to the parents of

those students, because of the perpetuation of the income imbalance that is analyzed and presented here.

Methods

In order to analyze this question, I first chose a dataset that would accurately represent the variables involved. I decided to use a dataset titled Statistics By Parent or Child Income Percentile, from the website <https://opportunityinsights.org/data/>. The 9,867,751 rows were divided into 100 bins, each with either 98677 or 98678 rows. The relevant columns for most of the conclusions and analysis are derived from “par_fam_inc” (the average parent family income) and “college” (the share of kids ever attending college during age 18-21). Because the threshold of “high income” compared to “low income” is arbitrary, I chose to split the data at \$60K, which evenly separated the data into two sets of 50. The revised hypothesis using this value is: Do students with families whose income is greater than \$60K/yr have a higher likelihood of attending college than those with income below \$60K/yr?

To describe the data, I decided to use the median of the data as the measure of central tendency, and standard deviation as the measure of spread. Median, in this case, is more helpful for determining central tendency because the data is significantly skewed right, towards higher income. The median value of the parent-family income is \$60,150. The standard deviation of the income data is \$148,202. This standard deviation reveals the skewness of the data towards higher income families. The descriptive statistics for both parent-family income and percentage of kids who attended college per family, divided into the 100 bins, are shown below in Table 1. To visualize the income data effectively by adjusting for the skewness, the highest 6 income rows were removed, and the corrected box plot and histogram were plotted. These two plots are shown

below in Figures 1 and 2. Table 1 is not adjusted for skewness in order to accurately represent all of the data.

Table (1) Descriptive statistics for the binned columns Parent Family Income (per year), and College Attendance (percentage of kids in the family who attended college from ages 18-21)

Statistic	Parent Family Income (\$)	College Attendance (%)
Count	100	100
Mean	87148	0.59
Standard Deviation	148202.8	0.2
Minimum	1700	0.25
25%	30650	0.42
50%	60150	0.58
75%	97500	0.75
Maximum	1408800	0.94

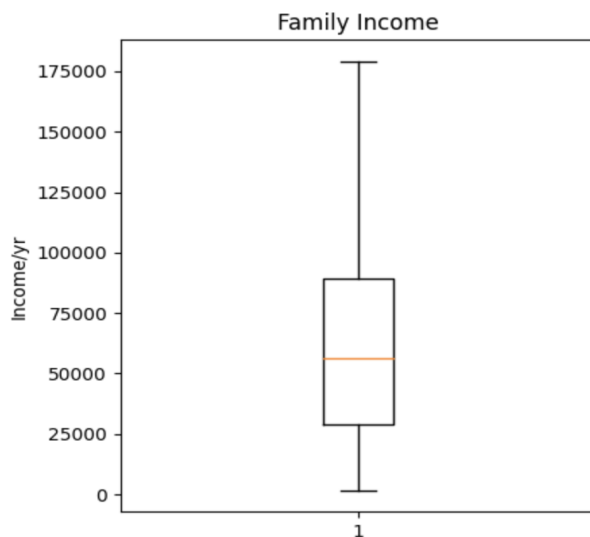


Figure (1) Box plot showing the mean, quartiles, and range of family income per year

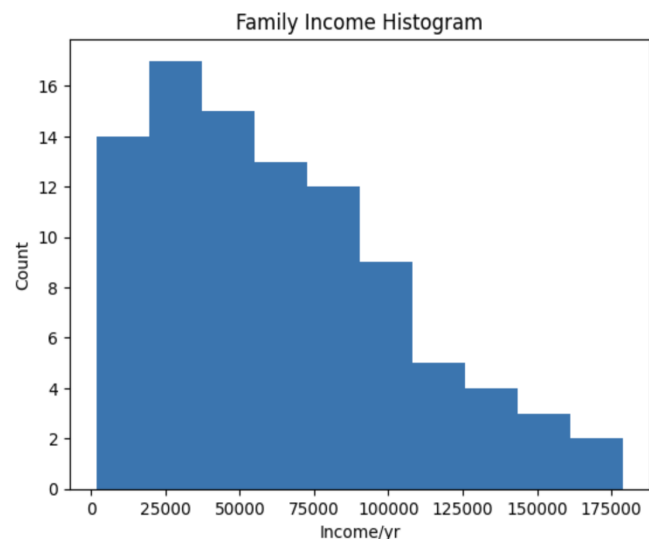


Figure (2) Histogram showing the distribution of family income per year, with 10 bins for ease of readability

Results

After cleaning, organizing, and describing the data, I analyzed the correlation between student family income and the students attending college. The apparent effect was a difference in means of 0.34. This effect size can be interpreted to mean that from the sample data, there is a difference of 34% between the percentage of kids per family who attended college that have high income families, and those that do not. I performed a two-sample t-test to calculate a p-value, and set the limit for a significant p-value as <0.05 , which is standard for similar tests. After running the test, the resultant p-value was $8.30e-30$. This is a significantly low p-value, indicating that we can reject the null hypothesis with great confidence. From this value, we conclude that there is a statistically significant increased likelihood for students with high-income families to attend college than those with low-income families.

After this initial conclusion, I ran a new test to further explore the college attendance of students. The new analysis question I tested was: Do more than 55% of students attend college, regardless of income? This question could help with related exploration of my initial question. My hypothesis for this test was that more than 55% of students end up attending college, regardless of income. The apparent effect was a percentage of students who attended college: 0.589, with a difference of 0.039. By running a one-sample t-test, the p-value was 0.024. Using a limit of <0.05 for the p-value again, this indicates that there is a statistically significant likelihood that more than 55% of students end up attending college, regardless of income.

Discussion

The results of this analysis reveal multiple conclusions. First, we learn that there is a statistically significant increased likelihood for students with high-income families to attend

college. This conclusion can be visualized in one way with the scatterplot of family income vs percentage of students who attended college per family, represented in Figure 3 below. There may be several covariate explanations for why students with high-income families have a higher likelihood of attending college, but the relationship shown from the analysis in this study is significant and strong. The R^2 value for the plot in Figure 3 is 0.97902968, indicating a strong linear correlation between parent-family income per year and college attendance.

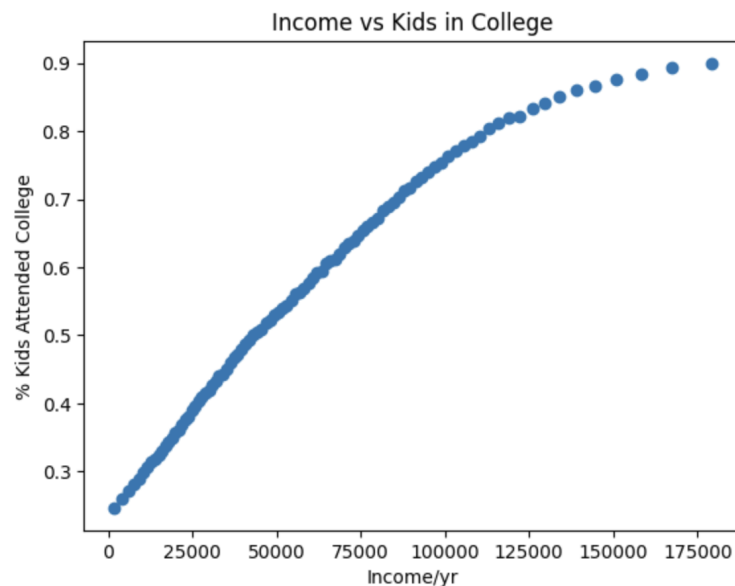


Figure (3) Scatterplot showing the trend of the percentage of students in each family that attended college vs the family income per year. The plot does not account for the top 6 rows of family income per year because these outliers can misrepresent the trend visually.

Another conclusion that can be drawn from the tests performed in this analysis is that there is a statistically significant likelihood that more than 55% of students per family end up attending college, regardless of income. This result, however, could be misinterpreted to mean that all students have equal chances of attending college despite income. But from the first analysis we can see that this is not the case.

The significance of these findings is revealing of the nationwide imbalance of higher education opportunities. This study demonstrates how rigid and predictable the trends of who has more access to college are. With these implications in mind, it is apparent that the severity of this problem is high, as is its consistency.

Further research could be conducted to analyze the other factors that affect the likelihood of attending college, such as race, gender, geographic location, whether parents attended college, etc. Some of these factors could be correlated with the income of the family to some extent, so this further analysis would be more revealing overall.