

Who goes to college?

By: Eric Krog

This paper is looking at the factors of if someone attends college. It looks at race, gender, income, and other data from the 2024 CPS ASEC survey taking only adults 25 and older. The main question it is answering is how does income, race, gender, marital status, and household role influence college attendance or years of education.

This is important allowing us to take a deeper look into which demographics attend college, and what affects one's years of education. There are debates in society regarding accessible education, the concerns of student debt and the value of college. Additionally on the financial side with the rising costs of tuition and the debate of student loan forgiveness were factors in choosing this dataset and question. All of these put together caused me to investigate schooling as a whole and which individuals are more likely to go.

To answer these questions, I made 2 sets of models in stargazer, one is unadjusted showing just race, while the other is adjusted to factor income, gender, marriage and head of household. This allows us to isolate the effect of each factor

The data shows that higher income has a strong correlation with college attendance. In the adjusted data Black individuals are more likely to attend in comparison to white people a reverse of the unadjusted result. Women and married people are more likely to have gone to or go to college, the head of a household is less likely to attend college. These findings show how variables can change our educational outcomes and suggest that certain variables such as race can show a deeper socioeconomic issue.

Education is a key driver for innovation and the economy, although one can be smart without a large amount of schooling, it is fundamental for economic and scientific growth. Although it is important, it is not always easily accessible, factors such as income, race, gender, and family life could affect one's access to education and that is what we are exploring in this paper.

In total, this papers goal is to help out the ongoing conversation around educational equity and equality by finding which factors whether demographic or socioeconomic are the most predictive towards college attendance. As well as how these relationships shift when income and social contexts are controlled.

Breaking down these factors let's begin with income, higher income families can afford college making it easier to attend (Bailey & Dynarski, 2011). Beyond paying for school, it can also allow them to get in easier with tutoring, counselors and other services that can boost GPA and factors that admit you into school as well. Next is race, we are looking if systemic inequality

affects one's opportunity to attend college. As for gender women currently have a higher enrollment in college than men (Buchmann & DiPrete, 2006), and although it may not be in this report there are factors with this as well such as male dominated and female dominated fields. We will also be looking at marital status to see if that has any effects on schooling and education levels as well.

This framework connects to broader ideas of social issues that can be overlooked. Although there may be race and income affects in regard to admissions there are still other factors that could be overlooked such as AP courses, school funding, and exposure to pro college propaganda. If one grew up in a culture of anti-college it may be considered wrong in their culture or thought process.

Other research has highlighted this being Bailey and Dynarski showing the inequality in college entry by income and Buchmann and DiPrete showing the rise of female advantage in completing college. By using both an adjusted and unadjusted model this paper will show changes in outcome when these factors are controlled.

This paper uses data from the 2024 Current Population Survey Annual Social and Economic Supplement (CPS ASEC). This is data collected by the U.S. Census Bureau and the Bureau of Labor Statistics which has 95,587 respondents aged 25 and older. It has persons demographic, income, and all of which we are using in this analysis. The CPS ASEC is a nationally representative survey collected each year using interviews and telephones to gather responses. Its depth and coverage make it a good source for understanding the labor market and education difference across U.S. population. It has an extremely large sample size at over 95,000 individuals which allows it to have meaningful comparisons that can cover income, education, household structure, and employment. One of the few limitations is that the data is self-reported which can cause inaccuracy. It also does not have GPA or other specific data. This data is ideal for what we are looking for which is education in comparison to socioeconomic variables in the United States, for the case of testing we are only looking at those aged 25 years and older so we can assess a completed education compared to those who have yet to go or are in progress.

The variables used are education (both continuously as years and binary for attended college or not), race (White, Black, American Indian, Asian, Pacific Islander, Multiracial, and Other), gender, income is right-skewed with extreme high-end values, so we use the natural log of income plus one to compress the scale and better capture marginal effects (e.g., a \$10,000 difference matters more at lower income levels), marital status, and finally Household role (Head of household).

Variable	Mean	Std. Dev.	Min	Max	N
----------	------	-----------	-----	-----	---

Years of Education	15.04	3.49	0	21	95,587
Log(Income + 1)	~10.5*	~1.7*	~0	~14	95,587
Age	51.75	16.44	25	85	95,587

*log(income+1)

The average respondent in the sample had about 15 years of education, suggesting many had some college education. Income levels are highly skewed which is why log was used to make the distribution more normal. The average age in this is 51.75 which is consistent with the adult sample restriction.

This paper uses two types of regression models to show how demographic and socioeconomic factors influence education attainment: (1) an ordinary least squares regression to show the dependent variable is years of education and a logistic regression model where the variable is attended college with 1 meaning yes and 0 meaning no.

The analysis has 2 parts

1. The unadjusted model includes only race as the independent variable. This gives us a baseline comparison of group differences between races.
2. The adjusted model has multiple factors
 - a. Income is made log(income+1) is a continuous variable for taxable income. The log being used to reduce the skewness. It is also used as 10,000 dollars has a bigger impact at lower incomes.
 - b. Race- has 7 categorical variables (white, black, Asian, American Indian, pacific islander, multiracial, and other)
 - c. Sex- binary with male or female as the options
 - d. Marital status: binary yes or no for being married or single
 - e. Head of household- binary yes or no for being the head of the house or not.

I decided to use both regressions as the first shows the continuous measure of how many years they have attended school. While the other shows the have you gone to college. Both go hand in hand and can show a lot about our society as a whole. The OLS regression treats years of education as a continuous variable which lets us show the differences in total education. The logistic regression in comparison just has a “Yes” or “No” for attending college. These both help show the differences in the adjusted and the unadjusted model.

Results

Here are the results from the two types of models. One being an OLS regression showing total years of education and logistic regressions predicating college attendance. The logistic models have both an adjusted and unadjusted version with race only and an adjusted version that adds other factors such as income, sex, marital status, and household role. Below is the reports for coefficients, statistical significance and other key data.

Everything within the above table is in comparison to a baseline group which is made of the following- White, Male, Unmarried, not head of household, Income = 0.

Variable	Unadjusted Coef.	(SE)	Adjusted Coef.	(SE)
Log(Income + 1)	—	—	0.124***	00.00
Black	−0.081**	00.03	0.102***	00.03
American Indian	−0.442***	00.08	−0.268***	00.08
Asian	0.194***	00.04	0.129***	00.04
Pacific Islander	−0.017	00.14	0.077	00.15
Multiracial	−0.395***	00.13	−0.289**	00.13
Other	−0.153	00.12	−0.060	00.12
Female	—	—	0.285***	00.02
Married	—	—	0.670***	00.02
Head of Household	—	—	−0.592***	00.02
Constant (everything baseline)	2.283***	00.01	1.421***	00.02

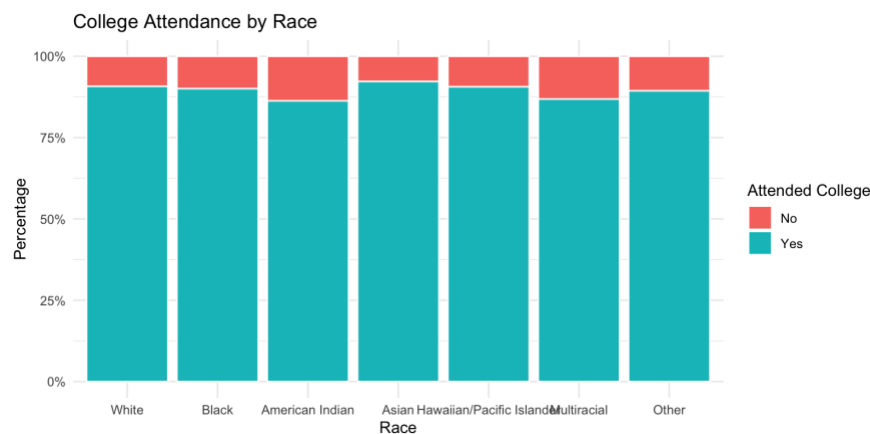
*** p< 0.01

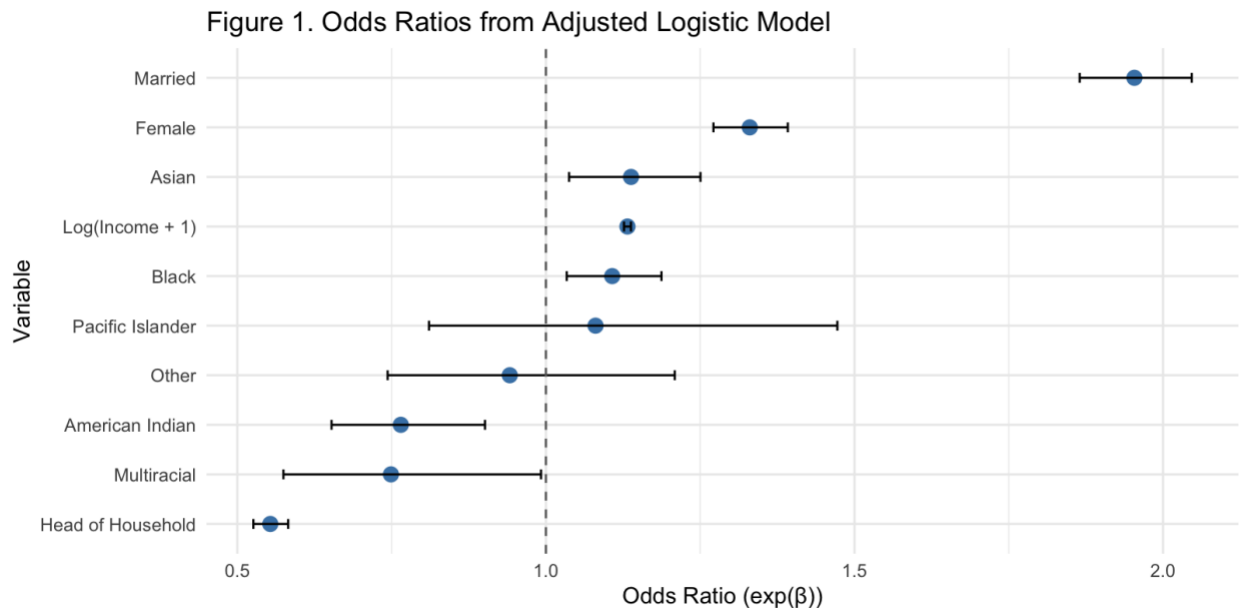
R² = .07

Sample size 94,783

Let's speak on the key coefficients. Income has a coefficient of .124 meaning there is a 13.2 percent increase in the odds of attending college for each one unit increase in $\log(\text{income} + 1)$. Females have 33% higher odds of attending school in comparison males, Married individuals have 95 percent higher odds of attending college in comparison to unmarries individuals. Heads of households have around 45 percent lower odds of attending college.

Comparing the adjusted and the unadjusted results, the effect for Black individuals changes drastically going from negative to positive (both of which are statistically significant) this could show socioeconomic issues within our schooling systems. This suggests that controlling income, gender, marital status, and household role, Black Individuals have structural disadvantages that change their chances of attending school. Both pacific islander and other in terms of race are statistically insignificant within both models, this suggests little to no association or affect in schooling. Multiracial and American Indian remain negative even after adjustment which is an interesting find.





To better show the findings, I've included two figures. These help us see what's going on in the data beyond just the regression numbers.

Figure 1 shows the proportion of people who attended college by race. While most groups have pretty high attendance, you can see that American Indian and Multiracial respondents have visibly lower college attendance rates. This backs up what we saw in the unadjusted model, where those two groups had large negative coefficients. Meanwhile, Asian and White groups show the highest proportions attending.

Figure 2 breaks down the adjusted logistic model into odds ratios. Each dot is an odds ratio for a variable, and the lines show the confidence intervals. It helps us understand how each factor is related to college attendance when all others are controlled for:

- Income: As income goes up, the odds of attending college go up. A one-unit increase in $\log(\text{income} + 1)$ is linked to about 13% higher odds of college attendance.
- Married people are nearly twice as likely to have gone to college compared to single people.
- Women are about 33% more likely to attend than men.
- Black individuals switch from a negative coefficient in the unadjusted model to a positive one in the adjusted model. That shows how important it is to control for income and other factors—otherwise, the racial differences can look misleading.
- Heads of household are significantly less likely to have attended college, which was one of the more surprising findings for me.

These visuals bring the data results to life. The first one shows raw group differences by race, and the second one shows how each variable affects college attendance when everything else is accounted for. One important takeaway is how the unadjusted and the adjusted model can show

different stories and outputs. In the unadjusted Black individuals seem less likely to attend college but once we control for income, gender, and family status the relationship flips and becomes not only positive but remains statistically significant. That may suggest that socioeconomic variables explain much of the initial gap and shows how raw comparisons could be misleading or misinformative. Along with that the negative coefficient for heads of households shows that a head of household may have had bigger responsibilities that make it harder to receive a higher education.

In summary, the results contribute to the idea that educational attainment has factors beyond ambition, it is changed by factors that influence opportunity before someone may even have the chance to apply.

This paper looked at who goes to college by analyzing data from the 2024 CPS ASEC. By running both unadjusted and adjusted models, we were able to isolate how race, income, gender, marital status, and household role influence college attendance. The biggest takeaway is that income plays a huge role—higher-income individuals are much more likely to attend college. That part wasn't too surprising, but what stood out more was how adjusting for income, and other factors changed the race-related results. For example, Black individuals went from being less likely to attend in the unadjusted model to being more likely in the adjusted one. That shift suggests that income and structural disadvantages—not race alone—may explain a large part of the gap in college attendance.

Other demographic factors also played a clear role. Married individuals were significantly more likely to attend college, which could reflect increased stability or access to resources within a household. Women were also more likely to attend college than men, continuing a broader trend that's been happening over the past few decades. On the flip side, people identified as the head of household were less likely to attend college. That could suggest that people taking on that responsibility early whether due to parenthood, caregiving, or economic pressure could face trade offs that make school less accessible.

Overall, the results show that who attends college is shaped by more than just academic effort, skill, or talent. Social and economic factors, including income, gender roles, and systemic barriers, all come into play. These findings support the argument that access to education isn't equal, and that efforts to improve equity in higher education should focus on reducing the structural disadvantages that certain groups face. This might include targeted financial aid, support services for nontraditional students, or policies that account for family responsibilities and economic hardship. Future research could expand on this by looking into field of study, graduation rates, or even the long-term effects of education access on employment and income.

Bibliography

Bailey, M., & Dynarski, S. (2011). *Gains and Gaps: Changing Inequality in U.S. College Entry and Completion*. <https://doi.org/10.3386/w17633>

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review*, 71(4), 515–541. <https://doi.org/10.1177/000312240607100401>