Hamsa Bandi, Maya Evans

Professor Taylor Weidman

Economic Data Analysis 0150

09 December 2025

Final Project Written Report (Bandi, Evans)

1) Introduction and question

Our project seeks to evaluate the relationship between level of education and expected income, specifically in the year 2023. We are interested in answering whether people with more education are more likely to make higher salaries, and we are testing against the null hypothesis that there is no relation.

2. Data description and sources

We sourced our data set from IPUMS USA, or the Integrated Public Use Microdata Series. We do not foresee any credibility issues with this source as it has the Core Trust Seal Certification, indicating international consensus that it contains trustworthy data.

We used the dataset for the 2023 US Census Data, which contains categories such as age, gender, and race, as well as income and education level. The first step of the data cleaning process was to filter for the only two categories we needed, which was income and education level. Secondarily, we filtered for only the year 2023, which eliminated every data point with a 2001 year tag. This ensured that all data points were within the correct time range and could be accurately compared against one another without any confounding variables. Thirdly, we filtered out the education code ‘99’, which indicated missing data on education. Finally, we filtered for

any outliers outside of the IQR, or middle 50% of the data set. This ensured that we would be working with a smaller, more averaged data set to try and find any more accurate relationships.

3. Methodology

Our general linear model graphs each data point according to the education label key, where they could fall into 12 categories ranging from 00 (no education past 4th grade) to 11 (more than 5 years of college education). Potential limitations of this model would be the fact that it does not account for partial years of education, as categories move in integer increments. Due to the bin sorting system, there is clear division between levels of education, and viewers can see the spread of potential salaries on the vertical axis of each category.

4. Results and analysis

While initially it may be hard to discern, the application of a regression line makes it clear that there is a positive relationship at work in the data. The regression output gives a Beta 0 of -1.131e+04, which hypothetically represents the yearly salary of someone with 0 years of education past 4th grade. However, in reality this would be a nonsensical interpretation, as you cannot have a negative salary. As such, we know that the true Beta 0 would be $0 annual salary. The Beta 1 education coefficient is 7431.29, which indicates that for every additional unit of education, we would expect to earn $7,431.29 more for an annual salary.

5. Conclusions

When testing the relationship between years of education and expected yearly salary, we found that for every additional unit of education, we would expect to earn $7,431.29 more in annual salary. This means there is a positive correlation, and helps us reject the idea that there is no relation.

6. References

Team, MPC UX/UI. 2023. “IPUMS USA.” *Ipums.org*. https://usa.ipums.org/usa/ (December 7, 2025).