**Econometrics Project 2**

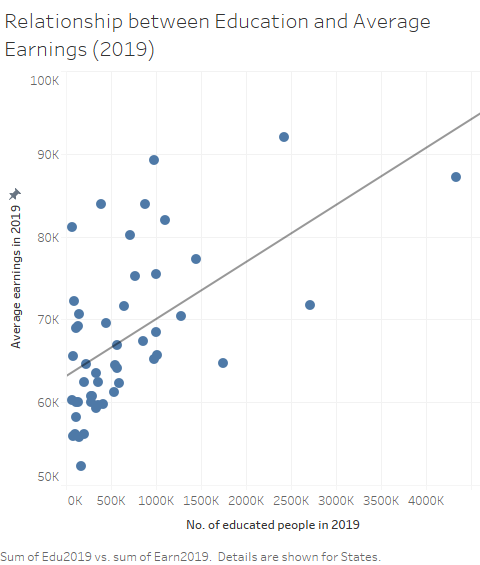
For the first project, I extracted two data sets – Average Earnings Per Job (Millions) in 2019 and Educational attainment across States in 2019 to see the impact of educational attainment on average earnings.

After filtering the data sets, I created 2 ‘Top 10’ tables in average earnings and educational attainment in 2019 which are as follows:

|  |  |
| --- | --- |
| **Top 10 states – 2019**  **(No. of educated people)** | |
| **State** | **2019** |
| California | 4,330,535 |
| Texas | 2,717,435 |
| New York | 2,418,300 |
| Florida | 1,745,405 |
| Illinois | 1,441,205 |
| Pennsylvania | 1,273,570 |
| New Jersey | 1,101,485 |
| Ohio | 1,015,845 |
| Georgia | 1,005,425 |
| Virginia | 1,001,990 |

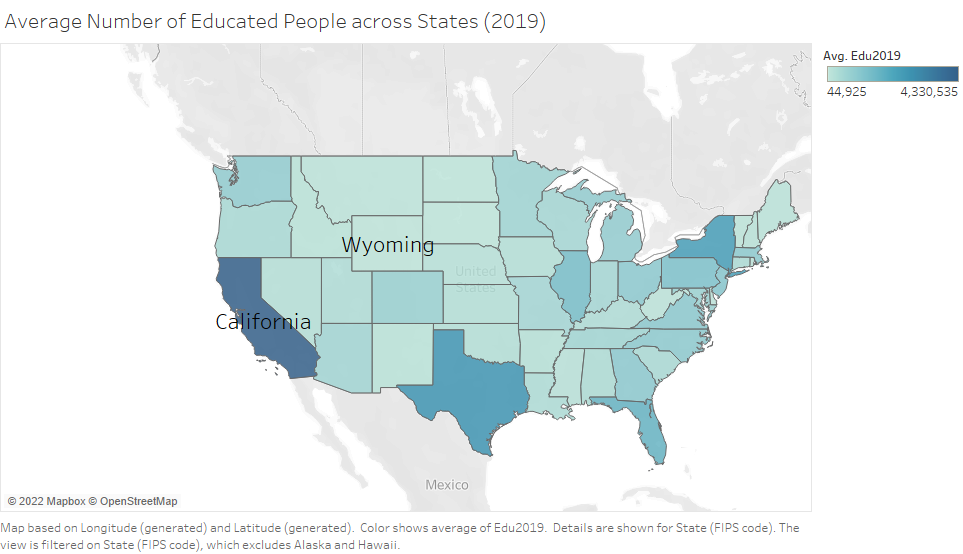
|  |  |
| --- | --- |
| **Top 10 states - 2019**  **(Average earnings per job)** | |
| **State** | **2019** |
| New York | $92061 |
| Massachusetts | $89258 |
| California | $87226 |
| Connecticut | $83968 |
| Washington | $83904 |
| New Jersey | $82047 |
| Alaska | $81213 |
| Maryland | $80157 |
| Illinois | $77248 |
| Virginia | $75465 |

The educational attainment data suggests that number of people with bachelor’s degree are the number of educated people in a state, which might create certain bias because a lot of individuals have master’s degree, PhD, etc. which is not included in the data set, so bachelor’s degree is not the best measure of educational attainment.



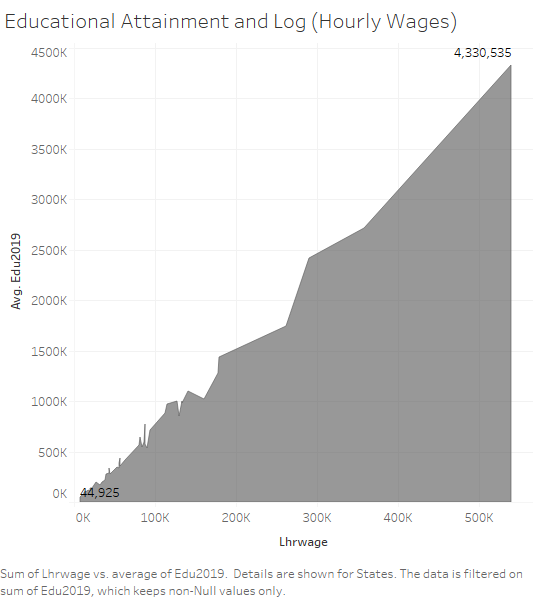
*Source*: U.S. Census Bureau (2021), Bureau of Economic Analysis (2021)

The above graph shows a relationship between number of educated people in a state and their average earnings per job in 2019. It seems to have a positive correlation which means that individuals who are getting bachelor’s degree earn more in general compared to those who do not get a bachelor’s degree.



*Source*: U.S. Census Bureau (2021)

The above graph shows the highest and lowest number of average educational attainment across states in 2019. Wyoming has the least average number of educated people and California had the most average number of educated people.



*Source*: U.S. Census Bureau (2021), IPUMS USA (2019)

The above graph shows relationship between Average number of educated people and Sum of log (hourly wages) which gives the average wage and income of people in those states.

It shows a somewhat positive correlation indicating as more people got their bachelor’s degree in 2019, their average wages increased.

**Table (1): Descriptive Statistics**

*Source*: IPUMS USA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| **Educational Attainment (2019)** | 1484185.57 | 1271835 | 44925 | 4330535 |
| **Years of Education** | 14.7542433 | 2.6446367 | 0 | 18 |
| **Log (Hourly Wages)** | 3.0473102 | 0.8492903 | -6.6592939 | 13.087772 |
| **Experience** | 28.6430553 | 15.0648549 | 1 | 94 |
| **Experiencesq** | 1047.37 | 936.097085 | 1 | 8836 |

The above table shows the descriptive statistics of the variables that I have used in my dataset.

Educational attainment across states in 2019 shows the number of individuals between age 25-44 with bachelor’s degree.

Years of Education = Calculated with the vague idea of how much an individual has studied.

Log (HourlyWages) is calculated first by calculating hourly wages and then taking the log of it.

Experience = Age – Years of education

Experiencesq = Experience^2

**Table (2): Regression Model 1**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number of Observations Read - 1450390** | | | | | |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** | **Mean** | **F Value** | **Pr > F** |
| **Squares** | **Square** |
| **Model** | 1 | 123042 | 123042 | 193322 | <.0001 |
| **Error** | 1.45E+06 | 923115 | 0.63646 |  |  |
| **Corrected Total** | 1.45E+06 | 1046157 |  |  |  |
| **Root MSE** | 0.79778 | **R-Square** | 0.1176 |  |  |
| **Dependent Mean** | 3.04731 | **Adj R-Sq** | 0.1176 |  |  |
| **Coeff Var** | 26.17997 |  |  |  |  |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** | **Standard** | **t Value** | **Pr > |t|** |
| **Estimate** | **Error** |
| **Intercept** | **1** | 1.42238 | 0.00375 | 378.84 | <.0001 |
| **educyr** | **1** | 0.11013 | 0.00025 | 439.68 | <.0001 |

Source: IPUMS USA

Regression Equation: *log (HourlyWages)* = 1.42238 + 0.11013*educyr* + *u*

Table (2) shows the results from first regression model where log of hourly wage is the dependent variable and years of education is the independent variable. The regression model is shown below the table, which shows that wages would increase by 11.01% for every additional year of education. I do not believe that this is a good model, because there are a lot of important variables which are not considered while running this regression. The R-square of this model is 0.1176 which explains 11.76% of total variation in log of hourly wages, and the years of education variable is significant at 1% level.

**Table (3): Regression Model 2**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Number of Observations Read - 1450390** | | | | | |
| **Number of Observations Used - 1450390** | | | | | |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** | **Mean** | **F Value** | **Pr > F** |
| **Squares** | **Square** |
| **Model** | 3 | 199895 | 66632 | 114198 | <.0001 |
| **Error** | 1.45E+06 | 846262 | 0.58347 |  |  |
| **Corrected Total** | 1.45E+06 | 1046157 |  |  |  |
|  |  |  |  |  |  |
| **Root MSE** | 0.76385 | **R-Square** | 0.1911 |  |  |
| **Dependent Mean** | 3.04731 | **Adj R-Sq** | 0.1911 |  |  |
| **Coeff Var** | 25.06652 |  |  |  |  |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** | **Standard** | **t Value** | **Pr > |t|** |
| **Estimate** | **Error** |
| **Intercept** | **1** | 0.6706 | 0.00415 | 161.62 | <.0001 |
| **educyr** | **1** | 0.11053 | 0.00025 | 448.74 | <.0001 |
| **Experience** | **1** | 0.04727 | 0.00017 | 277.48 | <.0001 |
| **Experiencesq** | **1** | -0.00058044 | 2.8E-06 | -210.33 | <.0001 |

*Source*: IPUMS USA

Regression Equation*: log (HourlyWages)* = 0.6706 + 0.11053*educyr* + 0.04727*experience* – 0.00058044*experiencesq* + *u*

Table (3) shows the results from second regression model where log of hourly wage is the dependent variable and years of education, experience, and experiencesq are the independent variables. The regression model is shown below the table, which shows that wages would increase by 11.05% for every additional year of education, 4.73% for every additional year of experience, and decrease by 0.058% for every experiencesq. Experiencesq seems to be having a negative relationship with wages. The R-square of this model is 0.1911 which explains 19.11% of total variation in log of hourly wages which shows that it is a better model than the previous one in Table (1) but it still has a lot of omitted variables which can have a huge importance in determining how exactly are wages affected considering multiple variables. The variables years of education, experience, and experiencesq are significant at 1% level.

**Table (4): Regression Model 3**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number of Observations Read - 1450390** | | | | | | |
| **Number of Observations Used - 1446590** | | | | | | |
| **Number of Observations with Missing Values - 3800** | | | | | | |
| **Analysis of Variance** | | | | | |  |
| **Source** | **DF** | **Sum of** | **Mean** | **F Value** | **Pr > F** |  |
| **Squares** | **Square** |  |
| **Model** | 4 | 204164 | 51041 | 88047.6 | <.0001 |  |
| **Error** | 1.45E+06 | 838582 | 0.5797 |  |  |  |
| **Corrected Total** | 1.45E+06 | 1042746 |  |  |  |  |
| **Root MSE** | 0.76138 | **R-Square** | 0.1958 |  |  |  |
| **Dependent Mean** | 3.04613 | **Adj R-Sq** | 0.1958 |  |  |  |
| **Coeff Var** | 24.99497 |  |  |  |  |  |
| **Parameter Estimates** | | | | | | |
| **Variable** | **Label** | **DF** | **Parameter** | **Standard** | **t Value** | **Pr > |t|** |
| **Estimate** | **Error** |
| **Intercept** | Intercept | **1** | 0.60072 | 0.00421 | 142.69 | <.0001 |
| **educyr** |  | **1** | 0.11041 | 0.00025 | 449.05 | <.0001 |
| **Experience** |  | **1** | 0.04721 | 0.00017 | 277.7 | <.0001 |
| **Experiencesq** |  | **1** | -0.0006 | 2.8E-06 | -209.89 | <.0001 |
| **edu2019** | edu2019 | **1** | 4.70E-08 | 4.98E-10 | 94.35 | <.0001 |

*Source*: IPUMS USA

Regression Equation*: log (HourlyWages)* = 0.60072 + 0.11041*educyr* + 0.04721*experience* – 0.0006*experiencesq* + 4.70E-08*edu2019* + *u*

Table (4) shows the results from third regression model where log of hourly wage is the dependent variable and years of education, experience, and experiencesq, edu2019 are the independent variables. The regression model is shown below the table, which shows that wages would increase by 11.04% for every additional year of education, 4.72% for every additional year of experience, decrease by 0.06% for every experiencesq, and increase by 4.70E-08 for every additional number of people educated (between 25-44 with bachelor’s degree). The R-square of this model is 0.1958 which explains 19.58% of total variation in log of hourly wages which shows that it is a better model than the previous one in Table (1) & Table (2) but educational attainment doesn’t seem to have much effect on wages because the R-square only increase from 0.1911 to 0.1958, and this model can have a lot of relevant variables determining how wages are affected. The variables years of education, experience, experiencesq, and educational attainment in 2019 are significant at 1% level.

As seen in the above table and explanation, state educational attainment has little to no influence on log (hourly wages) even after controlling for the person’s own educational attainment. The variable of educational attainment is so small that it has an exponent of -08, so it does not have much impact on wages.

**Table (5): Regression Model 4**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Number of Observations Read - 1450390** | | | | | | |
| **Number of Observations Used - 1446590** | | | | | | |
| **Number of Observations with Missing Values - 3800** | | | | | | |
| **Analysis of Variance** | | | | | |  |
| **Source** | **DF** | **Sum of** | **Mean** | **F Value** | **Pr > F** |  |
| **Squares** | **Square** |  |
| **Model** | 5 | 216639 | 43328 | 75870.7 | <.0001 |  |
| **Error** | 1.45E+06 | 826107 | 0.57107 |  |  |  |
| **Corrected Total** | 1.45E+06 | 1042746 |  |  |  |  |
| **Root MSE** | 0.75569 | **R-Square** | 0.2078 |  |  |  |
| **Dependent Mean** | 3.04613 | **Adj R-Sq** | 0.2078 |  |  |  |
| **Coeff Var** | 24.80837 |  |  |  |  |  |
| **Parameter Estimates** | | | | | | |
| **Variable** | **Label** | **DF** | **Parameter** | **Standard** | **t Value** | **Pr > |t|** |
| **Estimate** | **Error** |
| **Intercept** | Intercept | **1** | 1.09535 | 0.00535 | 204.6 | <.0001 |
| **educyr** |  | **1** | 0.10556 | 0.00024624 | 428.67 | <.0001 |
| **Experience** |  | **1** | 0.03913 | 0.0001774 | 220.56 | <.0001 |
| **Experiencesq** |  | **1** | -0.00048 | 0.00000281 | -170.57 | <.0001 |
| **edu2019** | edu2019 | **1** | 4.83E-08 | 4.94E-10 | 97.79 | <.0001 |
| **MaritalStatus** |  | **1** | -0.20511 | 0.00139 | -147.8 | <.0001 |

*Source*: IPUMS USA

Regression Equation*: log (HourlyWages)* = 1.09535 + 0.10556*educyr* + 0.03913*experience* – 0.0048*experiencesq* + 4.83E-08*edu2019* – 0.20511*maritalstatus* + *u*

Table (5) shows the results from the fourth regression model where log of hourly wage is the dependent variable and years of education, experience, experiencesq, edu2019, and marital status are the independent variables. I chose Marital status as one of the variables to see how being married or unmarried affects an individual’s wages. The regression model is shown below the table, which shows that wages would increase by 10.56% for every additional year of education, 3.93% for every additional year of experience, decrease by 0.05% for every experiencesq, increase by 4.70E-08 for every additional number of people educated (between 25-44 with bachelor’s degree), and decrease by 20.5% for being married. The R-square of this model is 0.2078 which explains 20.78% of total variation in log of hourly wages which shows that it is a better model than the all the previously discussed models. Apparently, marital status plays a big role in determining wages because as seen in the gender wage gap project last semester, married women tend too be underpaid due to discrimination, fewer skill set, uncertainty of taking maternity leave, etc. It can be close to a good model to show the variation in wages. The variables years of education, experience, experiencesq, educational attainment in 2019, and marital status are significant at 1% level.

Works Cited

“Bureau of Economic Analysis.” *U.S. Department of Commerce*, 23 Feb. 2022, https://www.commerce.gov/bureaus-and-offices/bea.

Team, MPC UX/UI. “U.S. Census Data for Social, Economic, and Health Research.” *IPUMS USA*, https://usa.ipums.org/usa/.

“U.S. Census Bureau.” *USAGov*, https://www.usa.gov/federal-agencies/u-s-census-bureau.