Education and Income

1. Motivation and purpose

Income inequality is one of the hottest topics in the social sciences, with important economic and political implications. It is usually said that today more than ever a college degree is indispensable for an individual to move upward in the socioeconomic ladder.

The purpose of this study is to analyze what impact education has on individual income. I use linear regression analysis to predict an individual's income as a function of his/her level of education, while controlling for other variables such as race and sex. More specifically, I am interested in answering questions like: how much extra income can be expected from earning a bachelor degree? what's the premium (if any) of holding a doctorate? Is it different for men and women? etc. These questions are addressed via hypothesis testing.

The scope of the study is delimited to New York City so that no adjustments have to be made to address purchasing power parity and other issues. The source of data is the Current Population Survey, a database sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. The sample is from 2011 and contains the following information from 20,428 individuals:

- Individual income
- Education, understood as highest level of education attained. It ranges from kindergarten to doctoral studies, with one academic year increments.
- Race
- Marital status
- Sex
- Age

Not all of the above variables are utilized, and some manipulations are applied to a few of them in order to simplify the analysis and make it more relevant. For example, the different categories within education are bundled into five: pre-high school, high school diploma, bachelor's degree, master's degree, doctoral degree. Likewise, race is simplified as black/non-black and marital status is reduced to single/married.

The author is aware that important variables may have been left out of the analysis. When considering individual income, a number of factors not considered here may have a substantial weight in individual income. For one thing, it would be interesting to distinguish between academic majors: for example, a degree in STEM or business is more likely to yield its holder a higher income than one in, say, liberal arts. Another factor of interest sociologists and public policy researchers look at is interpersonal skills and network capital, which crucially affect an individual's career advancement and upward mobility. Since a comprehensive database including all desired variables was not available, the CPS was deemed the best among the resources at hand given the purpose of the study.

2. Analysis

On a first approach, a regression model is run that uses education, race, marital status, sex and age as predictors of individual income. Dummy variables are employed for all qualitative variables. In particular, four dummy Boolean variables "high school", "bachelor", "master", PhD are used for education, leaving the pre-high school level out of the set to avoid perfect multi-colinearity. Race takes on the values 0 for non-black and 1 for black; marital status can be 0 for single and 1 for married; and male is 1 when the individual is male. The resulting model has the following form:

Income = β0 + β1*high school + β2*bachelor + β3*master + β4*PhD +

+ β5*black + β6*married + β7*male + β8*age

Excel regression tool returns the following results:

Regression Statistics	
Multiple R	0.41242
R Square	0.17009
Adjusted R Square	0.16976
Standard Error	69724
Observations	20429

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	2.0345E+13	2.54312E+12	523.1277177	0
Residual	20420	9.92693E+13	4861375962		
Total	20428	1.19614E+14			

Variables	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-14702	2199	-6.7	2.36E-11	-19012	-10392	-19012	-10392
highschool	11943	1722	6.9	4.15E-12	8568	15318	8568	15318
bachelor	49041	1786	27.5	5.37E-163	45539	52542	45539	52542
masters	64976	1984	32.7	3.26E-229	61086	68866	61086	68866
PhD	76215	3639	20.9	2.09E-96	69083	83347	69083	83347
black	-9384	1156	-8.1	4.94E-16	-11650	-7119	-11650	-7119
married	15566	1151	13.5	1.81E-41	13310	17823	13310	17823
male	23756	989	24.0	8.38E-126	21818	25695	21818	25695
age	520	40	13.1	8.37E-39	442	598	442	598

Notice that, according to this definition, the β-coefficients of each of the four education-related dummies must be interpreted as the premium in income relative to not having high school diploma, rather than as an absolute value. For instance, β3 = 64,976 means people with master's degree are expected to have an individual income $64,976 higher than those who didn't get a high school diploma; it does not mean they are expected to earn an absolute $64,976 even if black, married, male and age were zero, nor does it mean a master's degree brings an expected $64,976 additional income. The difference between consecutive coefficients can be interpreted as the premium income each degree confers on expected value. For example, β2 - β1 = 49,041-11,943 = 37,098 means a college degree leads to an average $37,098 additional income over a high school diploma. Likewise,

β1 = 11,943 = premium income of high school diploma above no high school.
β2 - β1 = 37,098 = premium income of bachellor's degree above high school diploma.
β3 – β2 = 15,935 = premium income of master's degree above bachellor's degree.
β4 – β3 = 11,239 = premium income of doctoral degree above master's degree.

All p-values and Significance F are well below 0.05, which suggests all coefficients variables considered have a slope different from zero. R2 is modest at 17%, although values above 30% are uncommon in social sciences due to the myriad of underlying factors and relationships usually involved. As mentioned above, additional variables such as academic major would make the model more complete.

Heterogeneous effects

In order to study closely the premiums derived from degrees and the possible differential effects of race and sex, a different model is considered. This part of the analysis focuses on bachelor and doctoral degrees since they represent the greatest leaps in an individual's academic life. Before proceeding to any model formulation, some simple summary statistics provide some color as to what results can be expected:

	high school	bach	Δ	% Δ
no black	39,512	77,309	37,797	96%
black	29,670	50,893	21,223	72%
Δ	-9,843	-26,417		
% Δ	-25%	-34%		

	high school	bach	Δ	% Δ
women	28,617	55,283	26,665	93%
men	42,031	90,227	48,196	115%
Δ	13,413	34,944		
% Δ	47%	63%		

Average individual income by se and race for individuals with and without bachellor's degree

Clearly college graduates earn higher incomes than high school graduates, regardless of sex and race. At the same time, men and non-black individuals realize higher incomes than women and black individuals, respectively. This is also true of PhD graduates vs. Bachellor's/Master's degrees:

	Bach/Master	PhD	Δ	% Δ
no black	83,744	114,205	30,461	36%
black	57,394	86,535	29,141	51%
Δ	-26,350	-27,670		
% Δ	-31%	-24%		

	Bach/Master	PhD	Δ	% Δ
women	60,967	82,671	21,704	36%
men	99,955	135,642	35,687	36%
Δ	38,988	52,971		
% Δ	64%	64%		

Average individual income by se and race for individuals with and without Phd degree

Now the following questions arise:

- Does earning a bachellor's degree (vs staying with a high school diploma) lead to the same income gain for both black and non-black individuals? Or is the effect different? The increase in income is 96% for non-black individuals vs 72% for black individuals. Is this statistically significant or just a coincidence?
- Does earning a bachellor's degree (vs staying with a high school diploma) lead to the same income gain for both men and women? Men get a 115% bump vs 93% of women.
- Does earning a doctoral degree (vs staying with a bachellor's/master's degree) lead to the same income gain for both black and non-black individuals?
- doctoral degree (vs staying with a bachellor's/master's degree) lead to the same income gain for both men and women?

In short, are there heterogeneous effects linked to race and sex? To answer these questions, interactions are introduced as variables to detect differential effects. At the bachellor's degree level, the following model is formulated:

$$\text{Income} = \beta 0 + \beta 1 * \text{bachellor} + \beta 2 * \text{black} + \beta 3 * \text{married} + \beta 4 * \text{male} + \\ + \beta 5 * \text{age} + \beta 6 * \text{bachxmale} + \beta 7 * \text{bachxblack}$$

where the variable bachelor is 1 if the highest educational level attained is a bachellor's degree and 0 if the individual gave up studying after obtaining his/her high school diploma. PhD, Master's and pre high school degrees are ruled out of the subsample. Notice that, by construction, β1 represents the premium income a non-black woman derives from earning a bachellor's degree. For a non-black man, the degree premium is β1+β6.

The corresponding regression returns the following results:

Regression Statistics	
Multiple R	0.392
R Square	0.154
Adjusted R Square	0.153
Standard Error	62908
Observations	14572

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	1.04668E+13	1.49525E+12	377.8409174	0
Residual	14564	5.7635E+13	3957361530		
Total	14571	6.81018E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-591.7	1943.0	-0.3	0.760727914	-4400	3217	-4400	3217
bachelor	31124.1	1690.7	18.4	7.88213E-75	27810	34438	27810	34438
black	-4581.0	1467.4	-3.1	0.001801194	-7457	-1705	-7457	-1705
married	15034.8	1239.5	12.1	1.07045E-33	12605	17464	12605	17464
male	11752.0	1401.3	8.4	5.46671E-17	9005	14499	9005	14499
age	592.2	43.5	13.6	5.12547E-42	507	677	507	677
bachxmale	18583.9	2116.6	8.8	**1.81632E-18**	14435	22733	14435	22733
bachxblack	-16164.3	2532.4	-6.4	**1.78931E-10**	-21128	-11200	-21128	-11200

where the low p-values of the two interaction variables indicate they are statistically significant (their slopes are significantly different from 0). Therefore it can be concluded:

Expected income premium for Bachellor's…

…if non-black woman	= β1	= $31,124
…if non-black man	= β1+β6	= $49,708
…if black woman	= β1+β7	= $14,960
…if black man	= β1+β6+β7	= $33,544

At the PhD degree level, the following model is formulated:

Income = β0 + β1*PhD + β2*black + β3*married + β4*male +

+ β5*age + β6* PhDxmale + β7* PhDxblack

where the variable PhD is 1 if the highest educational level attained is a PhD degree and 0 if the individual gave up studying after obtaining his/her Bachellor's/Master's degree. High school diploma and pre high school individuals are ruled out of the subsample. Notice that, by construction, β1 represents the premium income a non-black woman derives from earning a PhD's degree. For a non-black man, the degree premium is β1+β6.

The corresponding regression returns the following results:

Regression Statistics	
Multiple R	0.295
R Square	0.087
Adjusted R Squa	0.086
Standard Error	90489
Observations	10175

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	7.93821E+12	1.13403E+12	138.494	1.051E-195
Residual	10167	8.32503E+13	8188283314		
Total	10174	9.11885E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29079.9	3039.9	9.6	1.37E-21	23121	35039	23121	35039
PhD	12032.8	6650.9	1.8	7.04E-02	-1004	25070	-1004	25070
black	-18085.4	2481.5	-7.3	3.38E-13	-22950	-13221	-22950	-13221
married	23988.4	2056.2	11.7	3.00E-31	19958	28019	19958	28019
male	34091.9	1857.8	18.4	5.03E-74	30450	37733	30450	37733
age	586.3	76.4	7.7	1.80E-14	437	736	437	736
PhDxblack	109.3	14729.4	0.0	9.94E-01	-28763	28982	-28763	28982
PhDxmale	10586.1	8725.2	1.2	2.25E-01	-6517	27689	-6517	27689

where the p-values of the two interaction variables are above 0.05, which suggests this variables may actually be not statistically significant. In order to ascertain this, a new regression is run ignoring the interactions:

Regression Statistics	
Multiple R	0.29482
R Square	0.08692
Adjusted R Squa	0.08647
Standard Error	90487
Observations	10175

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	7.92613E+12	1.58523E+12	193.606746	1.1284E-197
Residual	10169	8.32624E+13	8187860888		
Total	10174	9.11885E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	28750	3027	9	2.6444E-21	22816	34683	22816	34683
PhD	17768	4384	4	5.08818E-05	9175	26360	9175	26360
black	-18043	2446	-7	1.76569E-13	-22839	-13248	-22839	-13248
married	23999	2056	12	2.81225E-31	19968	28029	19968	28029
male	34559	1817	19	2.81615E-79	30998	38121	30998	38121
age	589	76	8	1.37478E-14	439	738	439	738

and the corresponding F-test is performed:

SSE (RM)	8.32624E+13
SSE (FM)	8.32503E+13
n	10175
p	7
k	5
d.f.(num)	3
d.f.(den)	10167
alpha	0.05
Critical F	2.606
Observed F	0.492
p-value	0.688

Observed F < Critical F → null hypothesis $\beta 6 = \beta 7 = 0$ is accepted

Therefore it can be concluded that the expected income premium for a PhD degree is $17,768 regardless of race and sex.

3. Conclusions

On a first approach a simple model shows that increasing levels of education lead to increasing levels of individual income. Specifically, this analysis predicts that

…a high school diploma returns an expected additional $11,943,
…a bachelor's degree returns an expected additional $37,098,
…a master's degree returns an expected additional $15,935, and
…a PhD's degree returns an expected additional $11,239.

However, a closer inspection reveals potential heterogeneous effects that can be studied via interaction variables. The interactions of earning a bachelor's degree with race and sex show unequal effects of attaining a bachellor's degree for men vs women and for black individuals vs. non-black individuals. While a bachellor's degree leads to substantial increases in individual income, men and non-black individuals come out benefitted vs women and black individuals, respectively. Expected income premium derived from a Bachellor's degree is…

…if non-black woman	= $31,124
…if non-black man	= $49,708
…if black woman	= $14,960
…if black man	= $33,544

This phenomenon does not seem to occur for PhD graduates, who get an expected premium income over bachellor's/master's graduates equal to $17,768 regardless of race and sex. One reason for this might be that academia is more meritocratic and sets more objective compensation schemes based on qualifications/merits.

It is worth noticing here that other variables such as academic major (STEM, liberal arts, business, etc.), interpersonal skills, network/connections, family conditions during childhood/teenage (single mother, divorced parents, etc.), family income during childhood/teenage, etc. most likely play an important role in shaping an individual's academic and professional performance, and thus his/her income. The more of those variables could be incorporated into a model, the more complete the picture.