Tegveer Ghura and Andrew Zheng Present:

THE DAILY QUANT

QM222 Section G1 Project Part B

Returns To College: Does College Pay?

Documenting and analyzing differences in earnings between people who go and don't go to college.

Meet The Editors

Tegveer Ghura is an international student from India and is majoring in Economics at Boston University's College of Arts and Sciences. Andrew Zheng is a local US student from New Jersey and is majoring in Business Administration with a concentration in Finance at Boston University's Questrom School of Business. They met each other in a class known as SM131 (Business Society and Ethics), Questrom's entry level course for students looking to major or minor in the business school. For their QM222 class, they have come together to publish a quantitative modelling report regarding Returns to College. We hope you enjoy reading it!

Picture Above: Tegveer Ghura (left) and Andrew Zheng (right) after giving their SM131 Final Presentations.

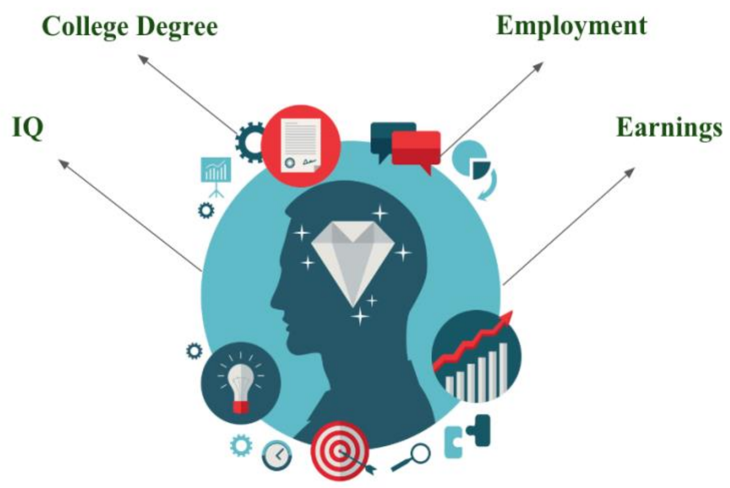
THE DAILY QUANT

Introduction

In the United States, the number of individuals obtaining a college degree has significantly increased. As a result, there is an ongoing debate on the importance and benefit of receiving a college education. This report aims to solve the question if going to college is a good investment by analyzing the effects (in terms of annual earnings) of a college degree. Part A of the report focuses on exploring the source of the gap between non-college graduates and college graduates, using descriptive statistics and graphing commands.

To make the dataset unique, we identified a member of our group that went **first** when our **last** names were ordered alphabetically. Therefore, Tegveer Ghura, who's ID number is U73744276, went first and, hence, we erased 300 observations starting from row 737 through row 1036 of our dataset.

Part B of this report will analyze the effect of a college degree on earnings. The effect of a college degree will be measured by analyzing its effect when holding other variables constant. These other variables include gender, area of residence, cognitive ability, and hours worked. This will be done by creating regressions A, B, C, and D to measure the impact of each variable.



Going Beyond the Requirements

“Increasing the average number of years of schooling attained by the labor force boosts the economy only when increased levels of school attainment also boost cognitive skills. In other words, it is not enough simply to spend more time in school; something has to be learned there.”

– Eric A. Hanushek

Hanushek’s research article about education and economic growth extends our analysis of earnings between college graduates and non-graduates. To understand this further, we have created an additional section named “Application of Hanushek’s Report”, which discusses a regression of two interaction terms we created, CollegeGrad*IQ and NotCollegeGrad*IQ, on Earnings. We also briefly discuss implications of the Human Capital Model versus the Signaling Effect.

Picture Above: Eric A. Hanushek

Regression C

Regression C measures the effect of college graduation, rural residence, male dummy (gender), and IQ on Total Earnings. The intercept’s coefficient was -52,535.92. This number’s statistical significance is that it is our baseline. If you are a non-graduate male who lived in a rural area and had an IQ score of zero, you earn $52,535 on average. Additionally, each variable’s coefficient affects your overall earnings. From our regression we have concluded and, by holding all other variables constant and strictly the singular coefficient:

- A college graduate will earn $35,787.67 more than a non-college graduate, on average, keeping other variables constant.
- Someone living in a rural area will earn $2354.08 less than someone living in a non-rural area, on average.
- If you are male you will make $28,521.43 more than females, on average, keeping other variables constant.

Continues on P3

- For every additional point scored on the IQ test you will make $757.42, on average, keeping other variables constant.

You may question whether these figures are statistically significant. By looking at each of their t-statistics scores, we are 95% confident that college graduates, gender, and IQ are statistically significant. However, the absolute value for rural dummy variables is less than 2, we can conclude that this intercept was not statistically significant.

The Relationship Between Cognitive Ability, College Attendance and Earnings

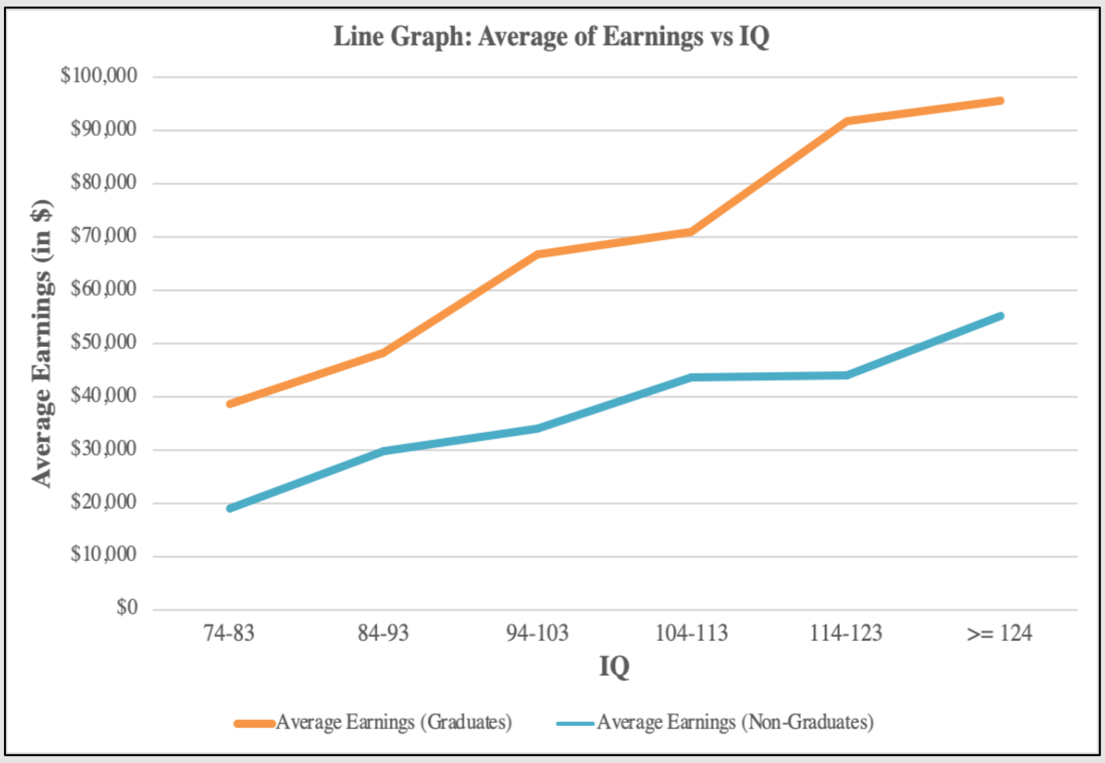
There is a positive relationship between attending college and earnings (Regression A or The Limited Model). The coefficient for college graduates is $35,127.45. This means, college graduates will make roughly $35,000 more than non-graduates. This statistic is significant because its t-statistic is 15.93 > 2.

There is a positive relationship between attending college and cognitive ability (Background Model). The coefficient for college graduate is 17.602. This means, college graduates have IQ's higher than non-graduates by approximately 18 points, on average. This statistic is significant because its t-statistic is 38.57 > 2.

In the Full Model, coefficient for college graduate is $35127.4501, which means, college graduates will make roughly $35127.4501 more than non-graduates, keeping cognitive ability constant. Coefficient for cognitive ability is $805.498905, which means that an additional IQ point leads to an increase in earnings by $805.498905, keeping college graduate constant. Both coefficients are statistically significant at the 95% confidence level.

The omitted-variable-bias formula is $c_1 = b_1 + a_1 * b_2$. This formula measures the combined effect of the direct effect (b_1) and the biased term (a_1b_2). From the formula, we have c1 = 49306.0006, b1 = 35127.4501, a1 = 17.602197, and b2 = 805.498905. Therefore: 49306.0006= 35127.4501+ (17.602197)(805.498905)). Because c1 > b1, the bias is Positive and, hence, we overestimate the effect of College Graduate by failing to include Cognitive Ability.

IQ ▾	Average of Earnings in $ (Non-Graduates)
74-83	$ 20,144.01
84-93	$ 29,335.48
94-103	$ 35,329.31
104-113	$ 41,302.74
114-123	$ 44,058.68
>= 124	$ 51,227.58
Average Grand Total	**$ 32,465.29**



IQ ▾	Average of Earnings in $ (Graduates)
74-83	$ 38,733.33
84-93	$ 48,323.84
94-103	$ 66,924.73
104-113	$ 71,139.36
114-123	$ 91,978.12
>= 124	$ 95,704.72
Average Grand Total	**$ 81,771.29**

The Effect of Self-Selection Bias

Self-selection bias occurs when individuals select themselves to participate in a group. This will cause the entire poll to suffer because there will be biased data. Individuals who choose to participate may not necessarily represent the total population because they volunteer the data, therefore, it may have different characteristics than the average. This can cause misrepresentative data.

Self-selection bias can impact these regressions because the survey claims to be a representative sample of the total population. However, we do not know how the participants were selected, if they were volunteered it could cause self-selection bias because more income rich individuals can volunteer their data opposed to the average total population.

Additionally, there can be factors that skew the data. For example, people do not want to participate in releasing their financial records. These factors can cause self-selection bias. This paper uses the information received and assumes the data is representative of the population.

We included the following variables in Regression D:

- College education. Which is the variable this report measures.
- Male Dummy. This was included because there is a gap between males and females*
- Rural residence. This was included because of the different cost of living in certain areas and salary differences in certain locations.*
- IQ. This was included because logically people with higher cognitive intelligence and ability should have higher potential in many occupations.*
- Hours. This was included because it is important to know the number of hours worked to produce a certain amount of earnings. This will help account for salary and individuals who are currently unemployed or in the process of looking for a job.
- Marriage status. This was included because we wanted to see if being in a relationship affected total earnings. If being married made an individual more inclined to work harder.

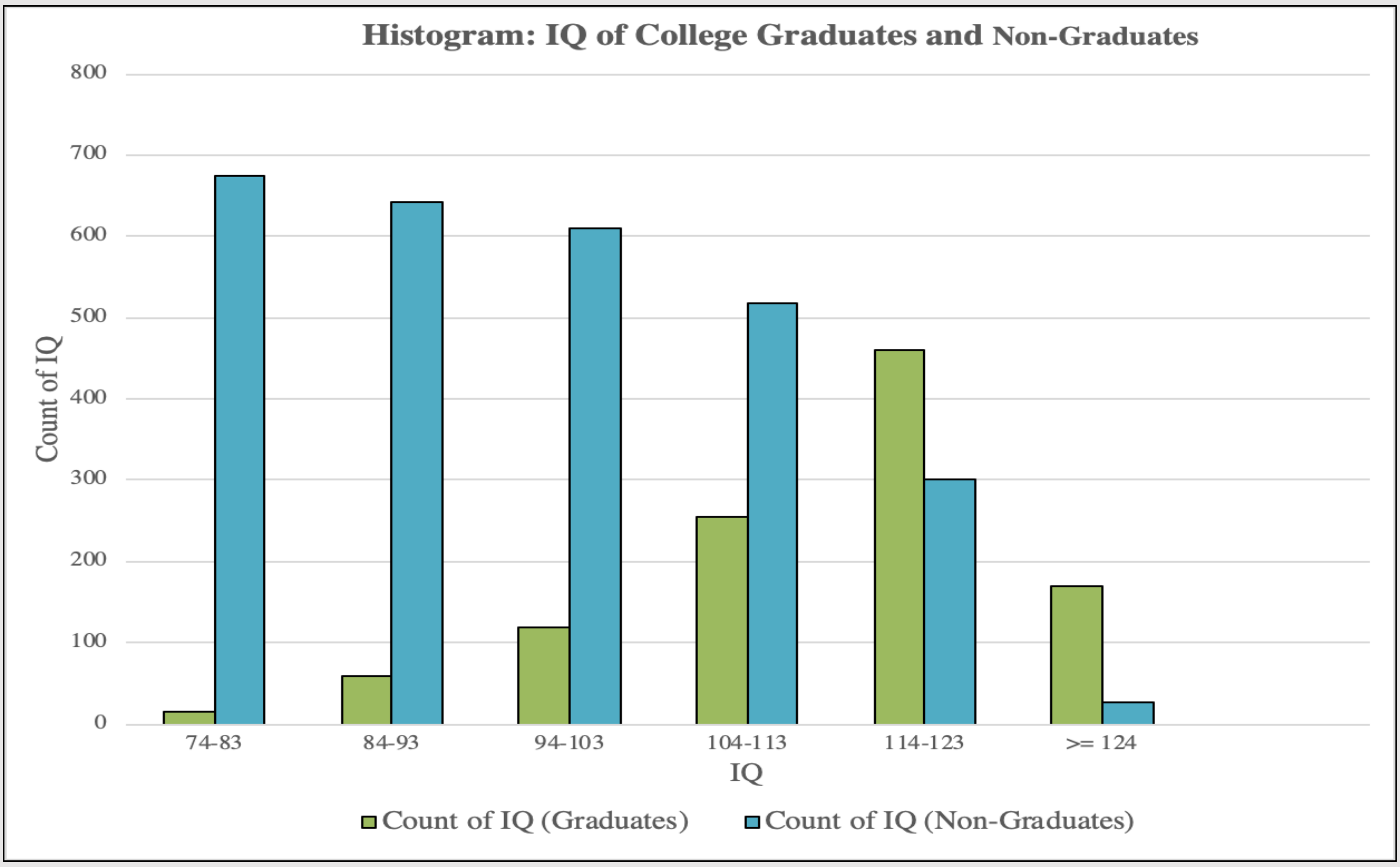
Lastly, by looking through regressions A through D, we noted for any significance difference between the coefficients for college education. We wanted to make sure there was least possible bias. We believed no other variables given would create a change in the coefficient for college education. This demonstrates that coefficient for college education in regression D does not contain much bias.

The variables marked with * additionally increased the adjusted R square value. Even though the variables hours and marriage were not significant in improving the R square value, they were still important to note.

Regression D

For regression D, we decided to use college graduation, rural, male dummy (gender), IQ, number of hours worked, and if they were married or not. These variables help explain the effects on total earnings. These variables were selected on what we thought were logical factors that may influence total earnings. We then made sure that these variables created the highest adjusted R square and had least bias.

Application of Hanushek's Report



Continues On P5

The key takeaway from Hanushek's research is that an increased level of cognitive skills leads to higher economic growth, translating to higher earnings and better standards of living. However, number of years of schooling only increases economic growth if cognitive skills are increased. To further reinforce this outcome, we used our data to produce a regression of two interaction terms, CollegeGrad*IQ and NotCollegeGrad*IQ, on Earnings. Our results did, in fact, support the outcome of Hanushek's paper. The coefficient of CollegeGrad*IQ was 1055.9, signifying that if IQ is boosted by 1 point for a college graduate, he/she would earn $1055.9 more. On the other hand, the coefficient of NotCollegeGrad*IQ was 729.5, , signifying that if IQ is boosted by 1 point for a non-college graduate, he/she would earn $729.5more, much less than that of a graduate.

This conclusion leads us to understand critical concept known as The Human Capital Model and the Signaling Model (Sheepskin Effect). Although both models support that additional schooling improves earnings, the Human Capital Model believes that additional schooling makes a person more productive (in terms of skills learnt) and, hence, increases earnings; however, the signaling model states that additional education does not make a person more productive but just acts as a signal to employers and provides them information about that person's **underlying or inherent productivity**. This is because only those who are productive to begin with will get more education. The Sheepskin Effect can help better relate to the Signaling Model. Take two students, for example, with similar GPA. One student completed graduation but one could not finish graduation because he/she had one class to graduate. Eventually, who gets paid more? The graduate does and this proves that a degree alone does play a signaling role. Therefore, it is not certain which model is more applicable or "correct" than the other, but according to our regression results and Hanushek's paper, it is certain that cognitive ability, along with increased schooling or more number of graduates, is essential to foster average earnings across a country.

It's evident from the histogram above that IQ for college graduates is highly skewed to the left or negatively skewed and IQ for non-graduates is highly skewed to the right or positively skewed. As a result, the mean IQ of non-graduates is greater than their median and median IQ for graduates is greater than their mean. Also, it should be noted that the number of observations for non-graduates was more than double than that of graduates. Therefore, this implies that graduates have more number of people, on average, with higher IQ scores than non-graduates do.

Response to The NY Times Quote

> "
>
> *Some employers are rethinking whether going to college is even necessary: 14 percent of hires at Google have no college degree, according to the company's senior human-resources officer. Nearly half of Americans surveyed last year by Public Agenda – a nonpartisan policy organization that focuses on education and other topics – said a higher education is no longer necessarily a good investment. And about the same proportion of graduates in a Gallup poll released last year said they were less than certain their degrees were worth the money.*
>
> "

The regressions ran in part B and the statistics shown in part A demonstrate that having a college degree will increase your earning potential on average.

By using variables like IQ, gender, rural background, number of hours, and marriage to receive the highest adjusted R squared, we know that 27% of total earnings is explained through these variables. These variables are shown in regression D. The regression will show that a college degree will increase earnings by $34,829.49 per year on average. Furthermore, the coefficient for college graduates is statistically significant because the t-statistics are greater than the absolute value of 2. This is shown through regression A to D. Therefore, we can be 95% confident that obtaining a college degree will obtain a positive effect on earnings.

From a financial stance, many individuals may get scared that the cost of tuition may not be covered after graduation, however from what we have gathered and calculated through regressions. One can conclude that going to college will likely lead to greater earnings. Additionally attending universities will allow for networking opportunities that people who do not attend college to have. These meaningful connections can surprise you if you are friends with the next Steve Jobs.

It is possible to obtain success/earnings without higher education; however, set yourself up with a higher chance of success through college.

Appendix

1. Descriptive Statistics

Descriptive Statistics of Annual Earnings	*College Graduates*	*Non-Graduates*
Mean	$81,771.28	$32,465.28
Median (50th Percentile)	$60,000	$26,000
Minimum	$0	$0
Maximum	$312,324	$312,324
Range (Maximum - Minimum)	$312,324	$312,324
Standard Deviation	$82,124.76	$36,104.31
25th Percentile (First Quartile)	$33,000	$4,500
75th Percentile (Third Quartile)	$100,000	$47,000
InterQuartile Range (75th Percentile - 25th Percentile)	$67,000	$42,500

2. Table of Regression Results

	A	B	C	D
College degree (y/n)	49306.000*** (25.84)	49125.829*** (26.65)	35787.672*** (16.81)	34829.493*** (16.61)
Male Dummy (1 for Male / 0 for Female)		29352.48703*** (17.78)	28521.439*** (17.58)	26389.338*** (16.38)
Rural (y/n)		-2312.95334 (-1.31)	-2354.080 (-1.36)	-2952.362* (-1.72)
IQ			757.421*** (11.89)	703.917*** (11.14)
Hours				307.528*** (10.90)
Married (y/n)				6102.190*** (3.53)
Constant (integer)	32465.289 (32.14)	19438.532 (13.90)	-52535.925 (-8.46)	-61443.932 (-9.98)
# observations	3851	3851	3851	3851
SEE	53173.15605	51111.57703	50203.35003	49389.73237
Adjusted r2	0.147629	0.212443	0.240183	0.264611

t-statistics in parentheses; *p<.1 **p<0.05 ***p<0.01

3. Regression A (also the Limited Model for relationship between cognitive ability, college attendance and earnings)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.38451367
R Square	0.14785076
Adjusted R Square	0.14762937
Standard Error	53173.156
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.88817E+12	1.8882E+12	667.814475	6.325E-136
Residual	3849	1.08826E+13	2827384524		
Total	3850	1.27708E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	32465.2886	1009.940586	32.1457411	4.808E-201	30485.2188	34445.3584	30485.2188	34445.3584
college_graduate	49306.0006	1907.971408	25.8421066	6.325E-136	45565.269	53046.7321	45565.269	53046.7321

4. Regression B

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.46158031
R Square	0.21305638
Adjusted R Square	0.2124427
Standard Error	51111.577
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	2.7209E+12	9.0696E+11	347.177725	1.61E-199
Residual	3847	1.005E+13	2612393307		
Total	3850	1.2771E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	19438.5324	1398.301	13.9015366	6.6526E-43	16697.0503	22180.0146	16697.0503	22180.0146
college_graduate	49125.8293	1842.99251	26.6554688	8.541E-144	45512.4935	52739.1651	45512.4935	52739.1651
Rural	-2312.9533	1764.17896	-1.311065	0.18991405	-5771.7688	1145.86211	-5771.7688	1145.86211
Male Dummy	29352.487	1650.10726	17.7882298	4.3389E-68	26117.3184	32587.6557	26117.3184	32587.6557

5. Regression C

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.49088943
R Square	0.24097244
Adjusted R Square	0.24018302
Standard Error	50203.35
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	3.0774E+12	7.6935E+11	305.252416	2.524E-228
Residual	3846	9.6934E+12	2520376354		
Total	3850	1.2771E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-52535.924	6205.57219	-8.4659274	3.5667E-17	-64702.451	-40369.397	-64702.451	-40369.397
college_graduate	35787.6723	2129.4851	16.8057867	3.2199E-61	31612.6443	39962.7003	31612.6443	39962.7003
Rural	-2354.0802	1732.83384	-1.3585147	0.1743801	-5751.4413	1043.28088	-5751.4413	1043.28088
Male Dummy	28521.4385	1622.29121	17.580961	1.3028E-66	25340.8052	31702.0718	25340.8052	31702.0718
IQ	757.420901	63.6845922	11.8933148	4.6464E-32	632.5621	882.279701	632.5621	882.279701

6. Regression D

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.51551661
R Square	0.26575737
Adjusted R Square	0.26461131
Standard Error	49389.7324
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	3.3939E+12	5.6565E+11	231.887774	1.824E-253
Residual	3844	9.3768E+12	2439345663		
Total	3850	1.2771E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-61443.932	6155.60979	-9.9817783	3.4994E-23	-73512.506	-49375.359	-73512.506	-49375.359
college_graduate	34829.4931	2096.96272	16.6094956	6.9424E-60	30718.2272	38940.759	30718.2272	38940.759
Rural	-2952.3625	1712.46774	-1.7240398	0.08478106	-6309.7947	405.069765	-6309.7947	405.069765
Male Dummy	26389.3388	1610.70454	16.3837241	2.2782E-58	23231.4216	29547.2561	23231.4216	29547.2561
IQ	703.91658	63.2043983	11.1371455	2.2384E-28	579.999218	827.833942	579.999218	827.833942
Hours	307.528307	28.2170308	10.8986771	2.9177E-27	252.206524	362.850091	252.206524	362.850091
married	6102.19037	1727.70291	3.53196741	0.00041733	2714.88833	9489.4924	2714.88833	9489.4924

7. Background Model for the relationship between cognitive ability, college attendance and earnings

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.52798614
R Square	0.27876936
Adjusted R Square	0.27858198
Standard Error	12.7182706
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	240643.927	240643.927	1487.71172	1.779E-275
Residual	3849	622592.714	161.754407		
Total	3850	863236.641			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	95.5627706	0.24156358	395.600912	0	95.0891657	96.0363754	95.0891657	96.0363754
college_graduate	17.602197	0.45635991	38.5708662	1.779E-275	16.7074667	18.4969273	16.7074667	18.4969273

8. Full Model for the relationship between cognitive ability, college attendance and earnings

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.42365322
R Square	0.17948205
Adjusted R Square	0.17905558
Standard Error	52183.725
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2.2921E+12	1.1461E+12	420.860336	5.076E-166
Residual	3848	1.0479E+13	2723141150		
Total	3850	1.2771E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-44510.418	6397.31516	-6.9576717	4.0482E-12	-57052.871	-31967.966	-57052.871	-31967.966
college_graduate	35127.4501	2204.84171	15.9319601	2.1658E-55	30804.6801	39450.2202	30804.6801	39450.2202
IQ	805.498905	66.1352591	12.1795683	1.6405E-33	675.835395	935.162416	675.835395	935.162416

9. Regression for the Application of Hanushek's Report

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.4270217
R Square	0.18234753
Adjusted R Square	0.18192256
Standard Error	52092.525
Observations	3851

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2.3287E+12	1.1644E+12	429.077955	6.057E-169
Residual	3848	1.0442E+13	2713631164		
Total	3850	1.2771E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-37382.032	6557.58857	-5.700576	1.2837E-08	-50238.713	-24525.35	-50238.713	-24525.35
CollegeGrad*IQ	1055.91295	59.0970768	17.8674311	1.1709E-68	940.048362	1171.77754	940.048362	1171.77754
NotCollegeGrad*IQ	729.524178	68.0750181	10.7164742	2.0028E-26	596.057613	862.990742	596.057613	862.990742