University of Nevada Las Vegas

Lee Business School

Department of Economics

Seminar in Economic Research

Research Portfolio for Econ 495

Tyler Pohlmann

Spring 2022

Acknowledgment: I would like to thank all my teachers and classmates that have helped me achieve my academic goals. All help, whether through lectures or small tips, have helped me acquire the skills necessary to put together this portfolio.

Table of Contents

Preface

Project 1

Project 2

Project 3

Term Paper (Life Expectancy)

Video Link

Preface

This portfolio is the collection of work I have completed throughout my economic capstone class during the Spring 2022 semester. This portfolio is comprised of a resume, 3 class projects, one term project, and a presentation on the final term project.

I am extremely proud of the portfolio I am putting forth today. I am still uncertain of what I will do for graduate school, but I believe this class and the rest of my experience at UNLV has provided me a fantastic base in order to begin my graduate career. This class in specific has provided me with a great deal of skills. Throughout the class, I have learned how to effectively breakdown data and interpret it. While these skills are applicable to the economic field, I believe that this skill suits many different career fields as well as communication in everyday life. I believe the ability to breakdown complex statistics into easy-to-read language is a skill that will help me excel in whatever career field I choose to go into.

The topic for my term project regards life expectancy and the different variables associated with longevity. I chose this topic because I was extremely interested on just how life expectancy is calculated and if there were certain variables that could be attributed to its rise. By the end of that paper, it is clear on the variable(s) that are correlated with life expectancy and what policy decisions can be made in order to raise it.

Project 1:Hedonic Housing Price Model

ECON 495 Section 1001

Spring 2022

Tyler Pohlmann, Bryce Ruggeroli, Jimmy Gamboa

Abstract:

This report will detail a list of properties and the characteristics that drive the price. The report will try and determine whether certain listing conditions can influence the end price of a house.

Introduction

This report pertains to a list of data given that contain a great amount of information regarding housing prices and characteristics. The data contains a house's price, stories, square footage, number of bathrooms, number of bedrooms, quantity of cars in garage, pool, fireplace, property, property age, vacancy, days on market, foreclosure and short sale.

The report will seek to determine if the variables: vacancy, days on market, foreclosure and short sale influence the overall housing price. In order to achieve this goal, we will use a P-value approach comparing the full model to the reduced model. The full model will contain all the variables discussed above, while the reduced model will omit the variables: vacancy, days on market, foreclosure, and short sale. Based on the significance of each model, it will be determined on whether the variables in question have an effect on housing price.

Model

Table 1: Variables, Variable Definition, and Expected Relationship with Price

Variables	Definition	Expected Sign
Price	Final price of property	n/a
Stories	Number of Stories	+
SqFt	Square footage of house	+
Baths	Number of Bathrooms	+
Bedrooms	Number of Bedrooms	+
Garage	Number of Car Garages	+
Pool	If property has pool	+

Fireplace	If property has fireplace	+
Age	Age of Property	-
Vacant	Occupancy of Property	+
Dom	Number of Day on Market	-
Foreclose	If house was Foreclosed	-
Shortsale	If house was sold during shortsale	-

The variables that will be discussed in the model are: price, stories, square footage, baths, bedrooms, garage, pool, fireplace, age, vacant, days on market, foreclose, shortsale. The price is the final property and has no expected sign because it is the dependent variable. The stories variable details the amount of stories the house has and it has an expected sign of positive because multiple story houses are typically bigger which means a higher price. Sqft details the square footage of a house and has an expected sign of positive because bigger houses typically have a higher price. Baths is the number of bathrooms in a house and has an expected positive sign because more bathrooms can accommodate more people making the house more valuable. Bedrooms is the number of bedrooms and is expected to be positive because it means the house can hold more people. Garage details the amount of cars the house's garage can hold which is expected to be positive. Pool and fireplace are both expected to be positive because they increase the value of a house. Age, days on market, foreclose, and shortsale all have expected negative sign because they show a house is less desirable by others. Finally, vacancy has an expected positive sign because it may allow the new buyers to have more freedom with the arrangement of the house.

Descriptive Statistics

Table 2: Summary Statistics of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
price	61,811	305936.7	127487.7	72000	865000
stories	61,811	1.504748	.5238736	1	3
sqft	61,811	1968.394	709.4849	452	7104
baths	61,811	2.464588	.6954588	1	8
bedrooms	61,811	3.402097	.7774559	1	9
garage	61,811	2.069745	.7747045	0	12
pool	61,811	.2064034	.4047267	0	1
fireplace	61,811	.5638317	.4959128	0	1
age	61,811	15.28328	12.49517	2	78
vacant	61,811	.6760447	.467987	0	1
dom	61,811	63.28459	63.77904	0	847
foreclose	61,811	.2257689	.418091	0	1
shortsale	61,811	.0425976	.2019497	0	1

Price has an Obs of 61,811 which means the average is 305936. While the standard deviation is 127487.7. With a Min. of 72000 and a Max of 865000. **Stories** has an Obs of 61,811. Which means the average is 1.504748. While the standard deviation is .5238736. With a Min. of 1 and a

Max of 3. **Sqft** has an Obs of 61,811. Which means the average is 1968.394. While the standard deviation is 709.4849. With a Min. of 452 and a Max of 7104. **Baths** has an Obs of 61,811. Which means the average is 2.464588. While the standard deviation is .6954588. With a Min. of 1 and a Max of 8. **Bedrooms** has an Obs of 61,811. Which means the average is 3.402097. While the standard deviation is .7774559. With a Min. of 1 and a Max of 9. **Garage** has an Obs of 61,811. Which means the average is 2.069745. While the standard deviation is .7747045. With a Min. of 0 and a Max of 12. **Pool** has an Obs of 61,811. Which means the average is .2064034.

While the standard deviation is .4047267. With a Min. of 0 and a Max of 1. **Fireplace** has an Obs of 61,811. Which means the average is .5638317. While the standard deviation is 4959128. With a Min. of 0 and a Max of 1. **Age** has an Obs of 61,811. Which means the average is 15.28328. While the standard deviation is 12.49517. With a Min. of 2 and a Max of 78. **Vacant** has an Obs of 61,811. Which means the average is .6760447. While the standard deviation is .467987. With a Min. of 0 and a Max of 1. **Dom** has an Obs of 61,811. Which means the average is 63.28459. While the standard deviation is 63.77904. With a Min. of 0 and a Max of 847. **Fereclose** has an Obs of 61,811. Which means the average is .2257689. While the standard deviation is .418091. With a Min. of 0 and a Max of 1. **Shortsale** has an Obs of 61,811. Which means the average is 0.425976. While the standard deviation is .2019497. With a Min. of 0 and a Max of 1.

Table 3: Correlation Matrix

	Price	Stories	SqFt	Baths	Bedrooms
Price	1				

Stories	0.0981	1			
SqFt	0.6916	0.3645	1		
Baths	0.4590	0.6523	0.7028	1	
Bedrooms	0.3411	0.3538	0.6294	0.5003	1
Garage	0.4719	0.1833	0.5237	0.3976	0.2829
Pool	0.3150	0.0371	0.2565	0.1368	0.1965
Fireplace	0.3179	-0.0288	0.3064	0.1359	0.1837
Age	-0.1972	-0.3470	-0.2836	-0.4068	0.0799
Vacant	-0.2155	0.1289	0.0607	0.1084	0.0485
Dom	0.0956	0.0203	0.1016	0.0668	0.0246
Foreclose	-0.3484	0.1429	0.1021	0.1378	0.0884
Short Sale	-0.0603	0.0303	0.0311	0.03096	0.0273

Stories has a correlation coefficient of 0.0981 which means the amount of stories a home has a weak but positive relationship with the price of a home. Square footage has a correlation coefficient of 0.6916 which means square footage has a strong and positive relationship with the price of a house. Bath has a correlation coefficient of 0.4590 which means that the number of bathrooms has a moderate positive relationship with the price of a house. Bedrooms have a correlation coefficient of 0.3411 which means bedrooms have a weak and positive correlation with the price of a house. Garage has a correlation coefficient of 0.4719 which means the number of car garages has a moderate positive relationship with house price. Pool has a coefficient of 0.3150 which means whether or not a house has a pool has a weak positive correlation with the price of a house. Fireplace has a coefficient of 0.3179 which means whether or not a house has a fireplace has a weak positive relationship with the price of a house. Age has a coefficient of -0.1972 which means the age of a house has a weak negative relationship with the price of a house. Vacancy has a coefficient of -0.2155 which means the occupancy of a house has a negative weak correlation with the price of a house. Days on market have a coefficient of 0.0956 which means days on market have a weak but positive relationship with the price of a house. Foreclose has a coefficient of -0.3484 which means that foreclose has a weak negative relationship with house pricing. Short sale has a coefficient of -0.0603 which means short sale has a weak negative correlation with house pricing.

Empirical Results

Table 4: Regression Results

Variables	Full Model	Reduced Model
Stories	-41169.8***	-49443.5***
	(-58.48)	(-54.92)
Sqft	119.0***	114.2***
	(181.38)	(134.18)
Baths	26371.8***	23487.8***
	(39.12)	(26.74)
Bedrooms	-17191.4***	-20058.2***
	(-38.17)	(-34.22)
Garage	14118.0***	19288.3***

	(30.04)	(31.56)
Pool	34391.1***	42533.1***
	(48.52)	(46.41)
Fireplace	12504.8***	14495.8***
	(20.28)	(18.47)
Age	-665.5***	-126.9
	(-23.26)	(-3.42)
Vacant	-25171.8***	
	(-40.84)	
Dom	-15.86***	
	(-3.78)	
Foreclos	-119079.6***	
	(-172.04)	
Shortsale	-84404.8***	
	(-64.12)	
_cons	13544***	
	(73.94)	
N	61,811	61,811
R-Squared	0.738	0.555

F	(61798) 14519.39	(61802) 9638.24

In the full model above, a 100 square footage increase is expected to increase a home value by $11,900. This is statistically significant at all conventional levels of alpha. An increase of one bedroom to a house is expected to decrease the home value by $17,191.40. This is statistically significant at all conventional levels of alpha. An increase of one bathroom to a home is expected to increase the price by $26,371.80. This is statistically significant at all conventional levels of alpha. An addition of a pool is expected to increase the property value by $34,391.10. This is statistically significant at all conventional levels of alpha. Additional stories added to a house should decrease the value of it by $41,169.80. This is statistically significant at all conventional levels of alpha. A house containing a garage increases its value by $14,118.00. This is statistically significant at all conventional levels of alpha. A house containing a fireplace increases its value by $12,504.80. This is statistically significant at all conventional levels of alpha. The age of when the house was built can decrease its value by $66,550.00 per year. This is statistically significant at all conventional levels of alpha. If a property is vacant, it can decrease its value by $25,171.80. This is statistically significant at all conventional levels of alpha. Every day the property is on the market for sales the price drops $1,586.00. This is statistically significant at all conventional levels of alpha. If a property is foreclosed its value drops $119,079.60. This is statistically significant at all conventional levels of alpha. If it is a short sale property, its value drops $84,404.80. This is statistically significant at all conventional levels of alpha.

In the reduced model, a 100 square footage increase is expected to increase a home value by $11400. This is statistically significant at all conventional levels of alpha. An increase of one bedroom to a house is expected to decrease the home value by $20,058.20. This is statistically significant at all conventional levels of alpha. An increase of one bathroom to a home is expected to increase the price by $23,487.80. This is statistically significant at all conventional levels of alpha. An addition of a pool is expected to increase the property value by $42,533.10. This is statistically significant at all conventional levels of alpha. Additional stories added to a house should decrease the value of it by $49,443.50. This is statistically significant at all conventional levels of alpha. A house containing a garage increases its value by $19,288.30. This is statistically significant at all conventional levels of alpha. A house containing a fireplace increases its value by $14,495.80. This is statistically significant at all conventional levels of alpha. The age of when the house was built can decrease its value by $12,690per year. This is not statistically significant at all.

All predictions from the beginning of the report are accurate except the number of bedrooms and stories. In both models, it is shown that an increase of bedrooms/stories shows a decrease of the overall home value. It should also be noted that these coefficients are significant on some level in each of the models. Besides the bed/stories variables, all other variables seem to show a positive correlation with the value of the home.

R-squared in the full model is 0.738. This means that all the variables can explain 73.8% of the variation in the model. R-squared in the reduced model is 0.555. This means that all the reduced model variables can explain 55.5% of the variation in the reduced model. To compare the overall, we will also be using the F statistic. In the full model, F_c (61798)=14519.39 which is compared to F_α = 3.697. Because $F_c > F_{\alpha,}$ we reject the null and declare that the overall model is significant. For the reduced model, F_c (4,159) =9638.24 compared to F_α = 4.864. Because $F_c > F_{\alpha,}$ we reject the null and conclude the overall model is significant.

Conclusion

In conclusion, there are many different variables that influence the price of a property. This report set to determine how much each variable directly affected home prices. Upon data analysis, it was concluded that a foreclosed property had the largest decrease in value and an addition of a pool had the largest increase. Upon completion of data analysis, it was concluded that a home's value can increase or decrease vastly depending on certain characteristics of the home. I would recommend that homeowners consider these variables when buying or selling a home to receive the greatest benefit.

Appendix

```
Sunday March 13 18:55:18 2022 Page 1
___ ____ ____ ____ ____(R) /__ / ____/ / ____/ ___/ / /___/ / /___/ Statistics/Data analysis
  ___ ____ ____ ____ ____ ©
/__ / ____/ / ____/ 17.0
___/ / /___/ / /___/ SE—Standard Edition
Statistics and Data Science Copyright 1985-2021 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC https://www.stata.com
979-696-4600 stata@stata.com
Stata license: 40-student lab perpetual
Serial number: 401706336008
Licensed to: UNLV
UNLV
Notes:
1. Unicode is supported; see help unicode_advice.
2. Maximum number of variables is set to 5,000; see help set_maxvar.
1 . *(13 variables, 61812 observations pasted into data editor)
```

2 . ssc install asdoc checking asdoc consistency and verifying not already installed... all files already exist and are up to date.

3 . sum price stories sqft bths bdrs garage pool firepl age vacant dom foreclos shortsale

Variable Obs Mean Std. dev. Min Max

price 61,811 305936.7 127487.7 72000 865000 stories

61,811 1.504748 .5238736 1 3

sqft 61,811 1968.394 709.4849 452 7104

bths 61,811 2.645888 .6954588 1 8 bdrs

61,811 3.402097 .7774559 1 9

garage 61,811 2.069745 .7747045 0 12

pool 61,811 .2064034 .4047267 0 1

firepl 61,811 .5638317 .4959128 0 1 age

61,811 15.28328 12.49517 2 78 vacant

61,811 .6760447 .467987 0 1

dom 61,811 63.28459 63.77904 0 847 foreclos

61,811 .2257689 .418091 0 1 shortsale 61,811

.0425976 .2019497 0 1

4 . corr price stories sqft bths bdrs garage pool firepl age vacant dom foreclos shortsale (obs=61,811)

price stories sqft bths bdrs garage pool firepl age vacant do

price 1.0000 stories 0.0981 1.0000 sqft 0.6916

0.3645 1.0000 bths 0.4590 0.6523 0.7028

1.0000 bdrs 0.3411 0.3538 0.6294 0.5003

1.0000

garage 0.4719 0.1833 0.5237 0.3976 0.2829 1.0000 pool

0.3150 0.0371 0.2565 0.1368 0.1965 0.1535 1.0000

firepl 0.3179 -0.0288 0.3064 0.1359 0.1837 0.2756 0.2609 1.0000 age -0.1972 -0.3470 -0.2836 -0.4068 -

0.0799 -0.5086 0.1461 0.1293 1.0000 vacant -0.2155 0.1289

0.0607 0.1084 0.0485 -0.0097 -0.1396 -0.0809 -0.1175 1.0000 dom 0.0956 0.0203 0.1016 0.0668 0.0246

0.0606 0.0009 0.0357 -0.0414 0.0384 1.000

foreclos -0.3484 0.1429 0.1021 0.1378 0.0884 0.0429 -0.0392 -0.0262 -0.1299 0.3730 -0.101shortsale -0.0603

0.0303 0.0311 0.0396 0.0273 0.0320 0.0108 -0.0004 -0.0506 -0.0469 0.096

foreclos shorts~e

foreclos 1.0000 shortsale -

0.1135 1.0000

5 . reg price stories sqft bths bdrs garage pool firepl age vacant dom foreclos shortsale

Source SS df MS Number of obs = 61,811

F(12, 61798) = 14519.39

Model 7.4158e+14 12 6.1798e+13 Prob > F = 0.0000

Residual 2.6303e+14 61,798 4.2563e+09 R-squared = 0.7382

Adj R-squared = 0.7381

Total 1.0046e+15 61,810 1.6253e+10 Root MSE = 65240

price Coefficient Std. err. t P>|t| [95% conf. interval]

stories -41169.77 692.1182 -59.48 0.000 -42526.33 -39813.22

sqft 118.998 .6560685 181.38 0.000 117.7121 120.2839 bths

26371.78 674.0829 39.12 0.000 25050.58 27692.98 bdrs -

17191.39 450.3961 -38.17 0.000 -18074.17 -16308.61 garage

14118.01 469.9239 30.04 0.000 13196.96 15039.06 pool

34391.13 708.7371 48.52 0.000 33002.01 35780.26 firepl

12504.79 602.5384 20.75 0.000 11323.81 13685.76 age -

665.4682 28.61171 -23.26 0.000 -721.5472 -609.3892 vacant -

25171.83 616.3148 -40.84 0.000 -26379.81 -23963.85 dom -

15.86132 4.196421 -3.78 0.000 -24.08632 -7.636324 foreclos -

119079.6 692.146 -172.04 0.000 -120436.2 -117723 shortsale

-84404.8 1316.432 -64.12 0.000 -86985 -81824.59 _cons

137663.7 1774.261 77.59 0.000 134186.2 141141.3

6 . est sto m1

7 . reg price stories sqft bths bdrs garage pool firepl age

Source SS df MS Number of obs = 61,811

F(8, 61802) = 9638.24

Model 5.5764e+14 8 6.9705e+13 Prob > F = 0.0000

Residual 4.4696e+14 61,802 7.2322e+09 R-squared = 0.5551

Adj R-squared = 0.5550

Total 1.0046e+15 61,810 1.6253e+10 Root MSE = 85042

price Coefficient Std. err. t P>|t| [95% conf. interval] stories -49443.48 900.3198 -54.92 0.000 -51208.1 -47678.85

sqft 114.2428 .8514404 134.18 0.000 112.574 115.9116 bths

23487.82 878.4596 26.74 0.000 21766.03 25209.6 bdrs -

20058.24 586.2144 -34.22 0.000 -21207.23 -18909.26 garage

19288.25 611.1591 31.56 0.000 18090.38 20486.13 pool

42533.12 916.3763 46.41 0.000 40737.02 44329.22 firepl

14495.82 784.8705 18.47 0.000 12957.47 16034.17 age -

126.9497 37.13038 -3.42 0.001 -199.7253 -54.17408

_cons 106622.1 2247.828 47.43 0.000 102216.3 111027.8 Sunday

March 13 18:55:18

2022 Page 3 8 . est sto m2

9 . esttab m1 m2, r2 scalar(f)

(1) (2) price price

stories -41169.8*** -49443.5*** (-

59.48) (-54.92)

sqft 119.0*** 114.2*** (181.38)

(134.18)

bths 26371.8*** 23487.8***

(39.12) (26.74)

bdrs -17191.4*** -20058.2*** (-

38.17) (-34.22)

garage 14118.0*** 19288.3***

(30.04) (31.56)

pool 34391.1*** 42533.1***

(48.52) (46.41)

firepl 12504.8*** 14495.8***

(20.75) (18.47) age -665.5*** -

126.9*** (-23.26) (-3.42)

vacant -25171.8*** (-

40.84)

dom -15.86*** (-

3.78)

foreclos -119079.6*** (-

172.04)

shortsale -84404.8***

(-64.12)

_cons 137663.7*** 106622.1***

(77.59) (47.43)

N 61811 61811 R-sq 0.738

0.555 f

t statistics in parentheses *

p<0.05, ** p<0.01, *** p<0.001

10 .

Project 2: Wage Equation

ECON 495 Section 1001

Spring 2022

Tyler Pohlmann, Bryce Ruggeroli, Jimmy Gamboa

Abstract:

The purpose of this project is to introduce the econometrics of the determinants of wage using a sample of the 2002 Current Population data for the state of Nevada. The model is based on Jacob Mincer's human capital wage function and the Oaxaca-Blinder decomposition of wage differentials.

Introduction

Two issues observed in this report are Jacob Mincer’s human capital wage function and the Oaxaca-Blinder decomposition of wage differentials between men and women. The Mincer’s wage equation is a single-equation model that investigates the human capital earnings function through years of schooling and experience. The Blinder-Oaxaca’s decomposition divides variations in mean outcomes between two groups into two parts: one due to group differences in explanatory variable levels, and the other due to differences in regression coefficient magnitudes.

Paper 1, “Gender Differences In Pay” by Blau and Kahn, is about the gender pay gap. One discovery is how the wage gap between men and women narrows as they get older. Gender-specific characteristics play a significant impact in the gender gap; for example, women have less work experience than males. Due to that women's salaries and their vocations might be affected by labor market discrimination. This model's wage structure addresses the fact that women work in different industries than men. Paper 2, “On The Decomposition of Wage Differentials” by Cotton, addresses a commonly used method for breaking pay differentials into human capital and discrimination components in order to assess both the disadvantage discrimination imposes on black and minority wage earners and the benefit discrimination bestows on white and majority wage earners. Paper 3, “A Grand Gender Convergence: Its Last Chapter” by Goldin, discusses how the gender wage gap may disappear if companies didn't have a reason to disproportionately reward people who worked long and irregular hours. The salary is a summary statistic for an individual's years of education, training, prior labor force experience, and predicted future participation. Paper 4, “Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education” by Jagger and Page, demonstrates the importance of degrees while observing the Sheepskin Effect in educational returns. The SE is an applied economics theory that persons with an academic degree earn more money than people who have studied for the same length of time but do not have an academic degree. Paper 5, “Pay Differences Among the Highly Trained: Cohort Differences in the Sex Gap in Lawyers” by Noonan/Corcoran/Courant, studies the sex gap in earnings for cohorts of lawyers. In the study men consistently made more than women with similar characteristics. Hours worked differences have grown over time, explaining more of the sex-based earnings disparity, whereas job settings and years spent in private practice have reduced, explaining less of the gap.

Model

Table 1: Variables, Variable Definition, and Expected Signs

Variables	Definition	Expected Sign
Dependent Variable		
WAGE	=Weekly Wage	
Independent Variables		
AGE	=AGE	+
AGE ^{2}	=AGE squared	-
EDUC	=years of education	+
BACHDEG	=1 if bachelor degree and 0 if not	+
MARRIED	=1 if married and 0 if not	+
FEMALE	=1 if female and 0 if not	-
UNION	=1 if union member and 0 if not	+
PARTIME	=1 if part time and 0 if not	-

The natural log function is used to represent non negative skewed dependent variables. This function investigates the factors that influence a person's weekly pay. The independent variables are age, age squared, education, bachelor's degree, married, female, union, and part time. The wage is dependent on the age, years, bachelor's degree, whether the individual is married, whether the individual is a female, whether the individual is a union member, and whether the individual works part time. AGE is expected to be positive because the older people get the more they are expected to earn. AGE squared is expected to be negative because individuals can work up until around retirement. EDUC is expected to be positive because an individual earns more money based on years of education. BACHDEG is expected to be positive because those with a bachelor's degree typically earn more than those without one. MARRIED is expected to be positive because married couples who file their taxes jointly will have less withheld from their paychecks than singles. FEMALE is expected to be negative because women typically earn less than their counterparts. UNION is expected to be positive because workers receive better benefits and pay. PARTIME is expected to be negative because people who work part time typically earn less than a full time worker.

Descriptive Statistics

Table 2: Summary Statistics of Variables (Mean, Standard Deviation)

	All (Mean,Std)	Female (Mean, Std.)	Male (Mean, Std.)
Wage	(7.247, .514)	(7.152, .4755)	(7.328, .5320)

Age	(39.177, 11.652)	(39.666, 11.7667)	(38.757, 11.5456)
Age2	(1670.51, 937.967)	(1711.58, 954.09)	(1635.20, 923.20)
Education	(13.127, 2.453)	(13.240, 2.374)	(13.029, 2.516)
Bachelor Degree	(.1412, .3484)	(.1442, .3516)	(.1387, .3459)
Married	(.4482, .4975)	(.4554, .4985)	(.4421, .4970)
Female	(.4623, .4988)	n/a	n/a
Union	(.1754, .3805)	(.1688, .3750)	(.1811, .3854)
Part-Time	(.1412, .3484)	(.1860, .3894)	(.1028, .3039)

The table above shows the mean and standard deviations of all the variables in the full, female, and male models. The following will focus mainly on the contrasting elements of the male and female models and what those differences mean in terms of the gender pay gap. The average wage of females is about $7.15 while males average pay is $7.33. Males in the sample data are younger by a year on average with a slightly lower amount of education. In addition to the education variable, 13.87% of males have a bachelor's degree while 14.42% of females have a degree. 45.54% of the women are married while only 44.21% of males are married. Male unionization rates are higher by about 1.3% while women are 8% more likely to hold a part-time rather than full time job. While women typically have a higher education level than males, they are less likely to hold union jobs and work full time. Because of this, a conclusion can not be reached regarding the gender-pay gap when relying solely on summary statistics.

Empirical Results

Table 3: Regression Results

	Full	Female	Male
Age	0.0559 (7.39)***	0.0461 (4.39)***	0.0651 (5.99)***
Age2	-0.0006 (-6.34)***	-0.0005 (-3.77)***	-0.0007 (-5.15)***
Education	0.0735 (12.33)***	0.0698 (8.02)***	0.0746 (9.11)***
Bachelor Degree	0.0943 (2.26)*	0.0550 (0.93)	0.132 (2.34)*
Married	-0.0808 (-2.98)**	-0.0450 (-1.18)	-0.118 (-3.06)**
Female	-0.195 (-7.55)***		
Union	0.0910 (2.69)**	0.119 (2.44)*	0.0718 (1.52)
Part-Time	-0.0156 (-0.41)	-0.0852 (-1.79)	0.0814 (1.35)
_cons	5.173 (31.21)***	5.235 (21.86)***	4.973 (21.52)***
N	1140	527	613
R-Sq	0.306	0.251	0.321
F	62.38	24.87	40.88

The table above shows regression results for the full model, female, and male model. The highest correlation coefficient is in the male model for bachelor’s degree with a correlation coefficient of 13.2 which is significant at 95% significance level. While this is the highest correlation in all of the models, it is still a weak correlation. All other variables have extremely low correlations while compared to wage. All variables seem to significant on some level excluding part-time in all models, bachelor degree and married in the female model, and union in the male model. Because all of the correlations coefficients are low we can assume that many of the variables have little to no influence over wage. This is also supported by the R-squared number for each of the models. The R-squared for the full model is .306. This means that the variables can explain 30.6% of the variation in the model. The R-squared for the female model is 0.251. This means that 25.1% of the variation in the model can be explained by the variables. Finally, the R-squared for the male model is 0.321. This means that 32.1% of the variation can be explained by the variables. The R-squared in all models shows that the variables have little influence on the overall wage of the workers.

Oaxaca-Blinder Decomposition

Blinder-Oaxaca decomposition Number of obs = 1,140

Model = linear

Group 1: female = 0 N of obs 1 = 613

Group 2: female = 1 N of obs 2 = 527

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
overall						
group_1	7.328461	.0215725	339.71	0.000	7.28618	7.370743
group_2	7.152294	.0208182	343.56	0.000	7.111491	7.193097
difference	.1761674	.0299796	5.88	0.000	.1174086	.2349263
endowments	-.0108754	.01524	-0.71	0.475	-.0407453	.0189946
coefficients	.2034052	.0262659	7.74	0.000	.1519251	.2548854

interaction	-.0163625	.0090848	-1.80	0.072	-.0341684	.0014435
endowments						
age	-.0418898	.0333262	-1.26	0.209	-.107208	.0234284
age2	.0368865	.028685	1.29	0.198	-.0193349	.093108
educ	-.0147107	.01029	-1.43	0.153	-.0348787	.0054574
bachdeg	-.0003053	.0011863	-0.26	0.797	-.0026304	.0020198
married	.0005999	.0014259	0.42	0.674	-.0021948	.0033947
union	.0014525	.0027523	0.53	0.598	-.0039419	.0068468
partime	.0070915	.0043399	1.63	0.102	-.0014145	.0155974
coefficients						
age	.7559589	.5996742	1.26	0.207	-.4193809	1.931299
age2	-.3556041	.3176842	-1.12	0.263	-.9782537	.2670455
educ	.0634979	.158288	0.40	0.688	-.2467408	.3737366
bachdeg	.0111674	.0120925	0.92	0.356	-.0125336	.0348683
married	-.0332208	.0247562	-1.34	0.180	-.081742	.0153004
union	-.0079943	.0114747	-0.70	0.486	-.0304842	.0144956
partime	.0309875	.0145465	2.13	0.033	.0024769	.0594981
_cons	-.2613872	.3327853	-0.79	0.432	-.9136345	.3908601
interaction						
age	-.0173257	.0190593	-0.91	0.363	-.0546813	.0200299
age2	.0158708	.0183154	0.87	0.386	-.0200268	.0517683
educ	-.0010104	.0026129	-0.39	0.699	-.0061315	.0041108
bachdeg	-.0004298	.0016708	-0.26	0.797	-.0037046	.002845
married	.0009717	.002275	0.43	0.669	-.0034873	.0054306
union	-.0005773	.0013507	-0.43	0.669	-.0032247	.0020701
partime	-.0138617	.0072745	-1.91	0.057	-.0281194	.0003961

As you can see in the model above, the difference between the male and female pay is 17.6 cents. The endowments portion of the model is -0.0108754. This means if you apply the male distribution to the female group, their pay will actually decrease 10 cents. The coefficient effect portion of the model is 0.2034052. This means that the unexplained difference between males and females is 20.3 cents. This unexplained difference can be attributed as discrimination towards women. Because of this, we can conclude that there is a gender pay gap of 20.3 cents due to gender discrimination.

Conclusion

Through the observations of this report, many of the continuous variables were highly significant on their effect of weekly wage. The age coefficients and education were of high significance all across, between both genders. The effect of a bachelor's degree is larger for females than males with the coefficient being more significant for women. The effect for men is still positively skewed but less significant. The marriage coefficient is also much more significant for women than men in this report. Overall for this report, women on average earn much less based on several different factors.From the observations we were able to conclude the absolute number of the wage gap is 20.3 cents due to gender discrimination. Differences in endowment and labor market discrimination are blamed for the wage disparity. Given the data used in this study, certain variables may not accurately depict effects on wages as greatly. Some other factors that could have given a better resolution to this wage gap could be further pursuing the effects of even higher education and observing the differences in wage disparity among different races.

References

-Blau, F. and Kahn, L. “Gender differences in Pay” *Journal of Economics Perspectives*. Vol. 14 (4), Fall 2000: 75-99.

-Cotton, J. “On the Decomposition of Wage Differentials” *The Review of Economics and Statistics*. Vol. 69, 1988:236-243.

-Claudia Goldin. “A grand Gender Convergence: Its Last Chapter.” *The American Economic Review*. Vol. 104 (04), 2014: 1091-1119.

-Jaeger D. and Page, M. “Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education” *The Review of Economics and Statistics.* Vol. 78, 1996: 733-740.

-Noonan M., Mary, Corcoran M, and Courant, P. “Pay Differences Among the Highly Trained: Cohort Differences in the Sex Gap in Lawyers’ Earnings.” *Social Forces*, Vol. 84, (2), 2005: 851-870.

Appendix

sum lwage age age2 educ bachdeg married female union partime

1

Variable	Obs	Mean	Std. dev.	Min	Max
lwage	1,140	7.247023	.5140347	5.769882	9.576583
age	1,140	39.17719	11.65235	18	64
age2	1,140	1670.511	937.9671	324	4096
educ	1,140	13.12675	2.452881	1	20
bachdeg	1,140	.1412281	.3484095	0	1
married	1,140	.4482456	.4975325	0	1
female	1,140	.4622807	.498794	0	1
union	1,140	.1754386	.3805087	0	1
partime	1,140	.1412281	.3484095	0	1

1 sum lwage age age2 educ bachdeg married union partime if female==1

2

Variable	Obs	Mean	Std. dev.	Min	Max
lwage	527	7.152294	.4755182	5.769882	8.713898
age	527	39.66603	11.76668	18	64
age2	527	1711.586	954.0943	324	4096
educ	527	13.24004	2.374166	1	20
bachdeg	527	.1442125	.3516389	0	1
married	527	.455408	.4984807	0	1
union	527	.1688805	.3750023	0	1
partime	527	.1859583	.3894426	0	1

3

4 . sum lwage age age2 educ bachdeg married union partime if female==0

Variable	Obs	Mean	Std. dev.	Min	Max

lwage	613	7.328461	.5320256	5.839915	9.576583
age	613	38.75693	11.54621	18	64
age2	613	1635.197	923.1967	324	4096
educ	613	13.02936	2.51644	2.5	20
bachdeg	613	.1386623	.345876	0	1
married	613	.4420881	.4970405	0	1
union	613	.1810767	.3853962	0	1
partime	613	.1027732	.3039105	0	1

10 . reg lwage age age2 educ bachdeg married female union partime

Source SS df MS Number of obs = 1,140

F(8, 1131) = 62.38 Model

92.1416551 8 11.5177069 Prob > F = 0.0000

Residual 208.818165 1,131 .184631446 R-squared = 0.3062

Adj R-squared = 0.3013 Total

300.95982 1,139 .264231624 Root MSE = .42969

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0559057	.007563	7.39	0.000	.0410666	.0707449
age2	-.0005879	.0000928	-6.34	0.000	-.00077	-.0004059
educ	.0735326	.0059631	12.33	0.000	.0618326	.0852325
bachdeg	.0943234	.041809	2.26	0.024	.0122916	.1763553
married	-.0807778	.0270995	-2.98	0.003	-.1339488	-.0276067
female	-.1946096	.0257856	-7.55	0.000	-.2452025	-.1440166
union	.0910351	.0338763	2.69	0.007	.0245676	.1575026
partime	-.0156106	.037661	-0.41	0.679	-.0895038	.0582826
_cons	5.17279	.1657357	31.21	0.000	4.847606	5.497973

10 . reg lwage age age2 educ bachdeg married union partime if female==1

Source SS df MS Number of obs = 527

F(7, 519) = 24.87 Model

29.8739795 7 4.26771136 Prob > F = 0.0000

Residual 89.0638726 519 .171606691 R-squared = 0.2512

Adj R-squared = 0.2411

Total 118.937852 526 .22611759 Root MSE = .41425

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0460783	.0104972	4.39	0.000	.025456	.0667006
age2	-.0004829	.0001282	-3.77	0.000	-.0007347	-.0002311
educ	.0698266	.0087095	8.02	0.000	.0527163	.0869369
bachdeg	.0550045	.0589323	0.93	0.351	-.0607707	.1707796
married	-.0450407	.0382207	-1.18	0.239	-.120127	.0300457
union	.1190906	.0487785	2.44	0.015	.023263	.2149183
partime	-.0852492	.0475536	-1.79	0.074	-.1786705	.0081721
_cons	5.234853	.2394516	21.86	0.000	4.764439	5.705267

11 . reg lwage age age2 educ bachdeg married union partime if female==0

Source SS df MS Number of obs = 613

F(7, 605) = 40.88 Model

55.6229257 7 7.94613225 Prob > F = 0.0000

Residual 117.604416 605 .194387464 R-squared = 0.3211

Adj R-squared = 0.3132

Total 173.227342 612 .283051212 Root MSE = .44089

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0651364	.0108768	5.99	0.000	.0437756	.0864971
age2	-.0006906	.0001341	-5.15	0.000	-.0009541	-.0004272
educ	.0746225	.0081897	9.11	0.000	.0585389	.0907061
bachdeg	.1324412	.0590806	2.24	0.025	.0164132	.2484693
married	-.1179881	.0384985	-3.06	0.002	-.193595	-.0423812
union	.0717535	.0470772	1.52	0.128	-.0207011	.1642082
partime	.0813876	.0602215	1.35	0.177	-.0368809	.1996561
_cons	4.973466	.2311039	21.52	0.000	4.519603	5.427329

12 . reg lwage age age2 educ bachdeg married female union partime

Source SS df MS Number of obs = 1,140

F(8, 1131) = 62.38 Model

92.1416551 8 11.5177069 Prob > F = 0.0000

Residual 208.818165 1,131 .184631446 R-squared = 0.3062

Adj R-squared = 0.3013 Total

300.95982 1,139 .264231624 Root MSE = .42969

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0559057	.007563	7.39	0.000	.0410666	.0707449
age2	-.0005879	.0000928	-6.34	0.000	-.00077	-.0004059
educ	.0735326	.0059631	12.33	0.000	.0618326	.0852325
bachdeg	.0943234	.041809	2.26	0.024	.0122916	.1763553
married	-.0807778	.0270995	-2.98	0.003	-.1339488	-.0276067
female	-.1946096	.0257856	-7.55	0.000	-.2452025	-.1440166
union	.0910351	.0338763	2.69	0.007	.0245676	.1575026
partime	-.0156106	.037661	-0.41	0.679	-.0895038	.0582826
_cons	5.17279	.1657357	31.21	0.000	4.847606	5.497973

13 . est sto m1

14 . reg lwage age age2 educ bachdeg married union partime if female==1

Source SS df MS Number of obs = 527

F(7, 519) = 24.87 Model

29.8739795 7 4.26771136 Prob > F = 0.0000

Residual 89.0638726 519 .171606691 R-squared = 0.2512

Adj R-squared = 0.2411

Total 118.937852 526 .22611759 Root MSE = .41425

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0460783	.0104972	4.39	0.000	.025456	.0667006
age2	-.0004829	.0001282	-3.77	0.000	-.0007347	-.0002311
educ	.0698266	.0087095	8.02	0.000	.0527163	.0869369
bachdeg	.0550045	.0589323	0.93	0.351	-.0607707	.1707796
married	-.0450407	.0382207	-1.18	0.239	-.120127	.0300457
union	.1190906	.0487785	2.44	0.015	.023263	.2149183
partime	-.0852492	.0475536	-1.79	0.074	-.1786705	.0081721
_cons	5.234853	.2394516	21.86	0.000	4.764439	5.705267

15 . est sto m2

16 . reg lwage age age2 educ bachdeg married union partime if female==0

Source SS df MS Number of obs = 613

F(7, 605) = 40.88 Model

55.6229257 7 7.94613225 Prob > F = 0.0000

Residual 117.604416 605 .194387464 R-squared = 0.3211

Adj R-squared = 0.3132 Total

173.227342 612 .283051212 Root MSE = .44089

lwage	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.0651364	.0108768	5.99	0.000	.0437756	.0864971
age2	-.0006906	.0001341	-5.15	0.000	-.0009541	-.0004272
educ	.0746225	.0081897	9.11	0.000	.0585389	.0907061
bachdeg	.1324412	.0590806	2.24	0.025	.0164132	.2484693
married	-.1179881	.0384985	-3.06	0.002	-.193595	-.0423812
union	.0717535	.0470772	1.52	0.128	-.0207011	.1642082
partime	.0813876	.0602215	1.35	0.177	-.0368809	.1996561
_cons	4.973466	.2311039	21.52	0.000	4.519603	5.427329

17

18 oaxaca lwage age age2 educ bachdeg married union partime, by (female)

Blinder-Oaxaca decomposition Number of obs = 1,140

Model = linear

Group 1: female = 0 N of obs 1 = 613

Group 2: female = 1 N of obs 2 = 527

lwage	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
overall						
group_1	7.328461	.0215725	339.71	0.000	7.28618	7.370743
group_2	7.152294	.0208182	343.56	0.000	7.111491	7.193097
difference	.1761674	.0299796	5.88	0.000	.1174086	.2349263
endowments	-.0108754	.01524	-0.71	0.475	-.0407453	.0189946
coefficients	.2034052	.0262659	7.74	0.000	.1519251	.2548854
interaction	-.0163625	.0090848	-1.80	0.072	-.0341684	.0014435
endowments						
age	-.0418898	.0333262	-1.26	0.209	-.107208	.0234284
age2	.0368865	.028685	1.29	0.198	-.0193349	.093108
educ	-.0147107	.01029	-1.43	0.153	-.0348787	.0054574

bachdeg	-.0003053	.0011863	-0.26	0.797	-.0026304	.0020198
married	.0005999	.0014259	0.42	0.674	-.0021948	.0033947
union	.0014525	.0027523	0.53	0.598	-.0039419	.0068468
partime	.0070915	.0043399	1.63	0.102	-.0014145	.0155974
coefficients						
age	.7559589	.5996742	1.26	0.207	-.4193809	1.931299
age2	-.3556041	.3176842	-1.12	0.263	-.9782537	.2670455
educ	.0634979	.158288	0.40	0.688	-.2467408	.3737366
bachdeg	.0111674	.0120925	0.92	0.356	-.0125336	.0348683
married	-.0332208	.0247562	-1.34	0.180	-.081742	.0153004
union	-.0079943	.0114747	-0.70	0.486	-.0304842	.0144956
partime	.0309875	.0145465	2.13	0.033	.0024769	.0594981
_cons	-.2613872	.3327853	-0.79	0.432	-.9136345	.3908601
interaction						
age	-.0173257	.0190593	-0.91	0.363	-.0546813	.0200299
age2	.0158708	.0183154	0.87	0.386	-.0200268	.0517683
educ	-.0010104	.0026129	-0.39	0.699	-.0061315	.0041108
bachdeg	-.0004298	.0016708	-0.26	0.797	-.0037046	.002845
married	.0009717	.002275	0.43	0.669	-.0034873	.0054306
union	-.0005773	.0013507	-0.43	0.669	-.0032247	.0020701
partime	-.0138617	.0072745	-1.91	0.057	-.0281194	.0003961

Project 3: Solow Growth Model

ECON 495 Section 1001

Spring 2022

Tyler Pohlmann, Bryce Ruggeroli, Jimmy Gamboa

Abstract:

The purpose of this project is to introduce the empirics of economic growth based on the Solow [3] model. In "A contribution to the Empirics of Economic Growth" (Quarterly Journal of Economics Growth, May 1992), Mankiw, Romer and Weil (MRW), [2] argue that "international differences in per capita income are best understood using an augmented Solow growth model." To replicate some of their empirical results, we will use their data set covering the period 1960-1985, for non-oil producing countries.

Introduction

The aim of this research is to show the international differences in per capita income are best understood using the main ideas of the Solow and augmented models. Paper 1, “Economic Growth in a Cross Section of Countries” by Robert J. Barro, observes that innovation, R&D spending, and technological investments are prerequisites for assuring competitiveness and progress, as well as long-term economic growth. The goal of this article is to see if an economy's innovation potential has an impact on long-term economic growth. The findings show that economic growth and innovation have a beneficial association. Paper 2, “A contribution to the Empirics of Economic Growth.” by Mankiw/Romer/Weil, investigates whether the Solow growth model is consistent with global diversity in living standards. The research also considers the Solow model's implications for standard of living convergence, or whether poor countries grow faster than rich countries. Paper 3, “A Contribution to the Theory of Economic Growth.” by Solow, discusses how all theories are based on assumptions that aren't entirely accurate. An essential assumption is one on which the findings are highly dependent, thus it is critical that crucial assumptions be realistic. In Solow's terms, the critical question of balance boils down to a comparison between the natural rate of growth, which is determined by the increase in the labor force in the absence of technological change, and the warranted rate of growth, which is determined by household and firm saving and investing habits. The majority of this study is devoted to a long-run growth model that accepts all of Solow's assumptions save the fixed proportions assumption.

Model

Table 1: Variables, Variable Definition, and Expected Signs

Dependent Variable	Variable Definition	Expected Sign
GDP85	GDP per working-person in 1985	Positive
Independent Variables		
IY	Investment as percentage of GDP	Positive
POPGR	population growth rate + advancement in knowledge + rate of capital depreciation	Negative
SCHOOL	Percentage of the working-age population in secondary school Average for 1960-85	Positive

The dependent variable in the model is GDP85. This variable represents the average GDP per working person in 1985. In addition to the dependent variable, there are three independent variables; IY, POPGR and SCHOOL. IY represents investment as a percentage of GDP which can also be described as the savings rate. We expect this variable to have a positive correlation with GDP85. POPGR is the summation of population growth, advance in knowledge, and capital depreciation. An increase of population is expected to have a negative correlation with GDP85 because as the population increases, the average GDP per person decreases. Finally, school represents the percentage of secondary educated people. This is expected to have a positive relationship with GDP85.

Descriptive Statistics

Table 2: Summary Statistics of Variables (Mean, Standard Deviation, Minimum, Maximum)

Variable	Obs	Mean	Std. dev.	Min	Max
gdp85	98	5309.765	5277.183	412	19723
popgr	98	2.20102	.8898621	.3	4.3
iy	98	17.67245	7.91833	4.1	36.9
school	98	5.396939	3.468992	.4	11.9

The table above displays the mean, standard deviation, minimum, and maximum and how they correspond to their variables. This figure explains the means and the amount of variability in our data. Regarding the standard deviation of variable gdp85 we can see it is more flat than bell shaped as there is a much higher amount of deviation compared to the other variables. Based on our data we can observe the minimum GDP of a person working in 1985 412 in Zaire between the countries while 19723 in Norway being the highest. The standard deviation is so high in this variable because it includes many underdeveloped countries compared to developed countries. Variable School is significant because it is known that countries that are behind in the GDP game dont have wide ease of access for education. That is why the SD is high at 3.4 even though most underdeveloped countries are under 1 because of access. Most of these number will be positively skewed in the conclusion of our project regarding GDP per working-person in 1985 because a larger portion of the world population is living in developed countries where the access to make money is easier.

Correlation Matrix

Table 3: Correlation Matrix

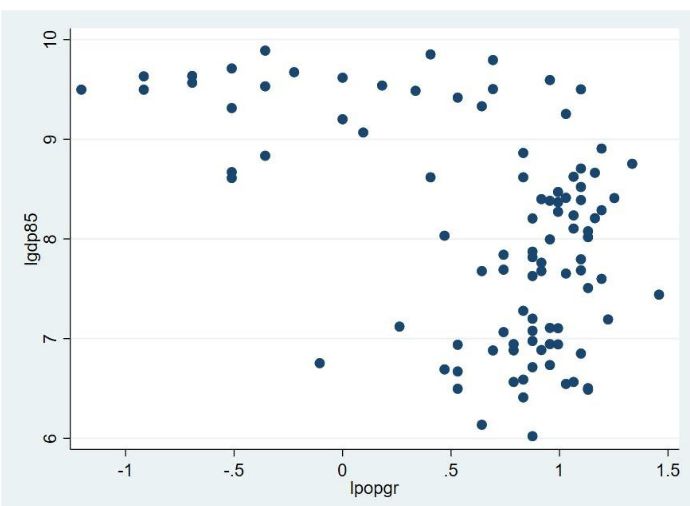
	Log(gdp85)

Log(popgr)	-0.4781
Log(iy)	0.7406
Log(school)	0.8123

The table above shows the correlation coefficients between the dependent and independent variables. The correlation coefficient between Log(POPGR) and Log(GDP85) is -0.4781. This shows a loose negative correlation between the two variables. This supports our prediction that population growth would have a negative effect on average GDP. The correlation coefficient between Log(GDP85) and Log(IY) is 0.7406. This shows a strong positive relationship between the variables which also supports our earlier hypothesis. The correlation coefficient between Log(SCHOOL) and Log(GDP85) is 0.8123. This shows the two variables have a strong positive relationship which is also consistent with the prediction in Table 1.

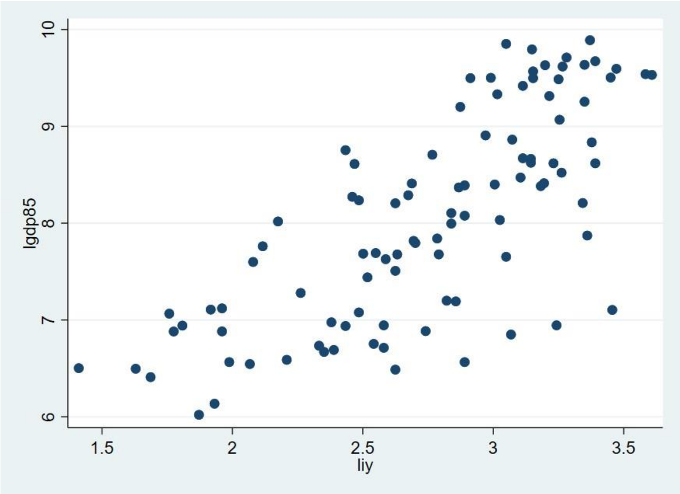
Scatter Plots

Scatter Plot 1: Log(GDP85) and Log(POPGR)



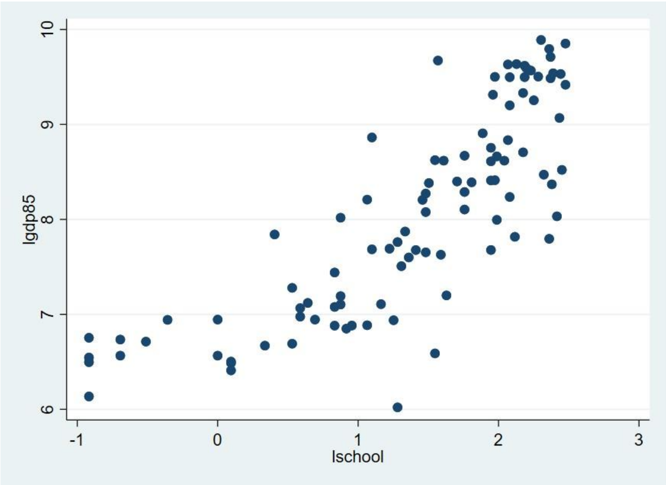
The scatter plot above shows a moderately weak negative relationship between Log(GDP85) and Log(POPGR). This is supported by the correlation matrix where the two variables have a correlation coefficient of -0.4781.

Scatter Plot 2: Log(GDP85) and Log(IY)



The scatter plot above shows a visual representation between the two variables Log(GDP85) and Log(IY). The two variables have a strong positive relationship with each other which is supported by their correlation coefficient of 0.7406 in the correlation matrix

Scatter Plot 3: Log(GDP85) and Log(School)



The scatter plot above shows a strong positive relationship between the variables Log(GDP85) and Log(SCHOOL). This is consistent with the findings of the correlation matrix when the two variables had a correlation coefficient of 0.8123.

Empirical Results

Table 3: Regression Results

	Solow	Augmented Solow
	lgdp85	lgdp85
liy	1.387^{***}	0.683^{***}
	(9.80)	(5.24)
lpopgr	-0.510^{***}	-0.432^{***}
	(-4.11)	(-4.71)
lschool		0.643^{***}
		(9.00)
_cons	4.562^{***}	5.568^{***}
	(10.65)	(16.64)
R-squared	0.6165	0.7940
F	76.37	120.79

N	98	98

The table above shows the regression results for both the basic Solow Model and the Augmented Solow model. In the basic Solow model, IY had a coefficient of 1.387 which was significant at all levels. POPGR had a coefficient of -0.510 which was also significant at all levels. The overall model had an R-squared of 0.6165 meaning that 61.65% of the variation can be explained by the variables in the model. In the augmented model, IY's coefficient is reduced quite significantly to 0.683. POPGR coefficient was reduced to -0.432. Finally, SCHOOL had a coefficient of 0.643. All variables were significant at all levels. The R-squared of the augmented model was 0.7940 meaning that 79.40% of the variation in the model can be explained by the variables. Based on the data in the table above, it seems that the SCHOOL variable plays a very large in explaining much of the variation in the data.

Convergence

Table 5: Test for Unconditional Convergence

	Log(GDP85)-Log(GDP60)
Log(GDP60)	.0943109
Constant	-0.2665776
N	98
R-squared	0.0363

Table 6: Test for Conditional Convergence

	(3)	(4)

	Solow	Augmented solow
liy	0.646***	0.521***
	(7.48)	(6.03)
lpopgr	-0.0886	-0.135*
	(-1.28)	(-2.06)
lschool		0.234***
		(3.96)
lgdp60	-0.149**	-0.300***
	(-2.82)	(-4.83)
_cons	-0.141	1.059*
	(-0.38)	(2.32)
N	98	98
R-sq	0.406	0.492

F	21.41	22.49

The convergence tables above will help us find if poorer countries grow faster than richer countries. In table 5, the coefficient for the income per capita variable, GDP60 is .0943. Because the coefficient is positive, we can assume that there is no evidence that poor countries grow faster than richer countries. Although when we include other variables, the results seem to change. When introducing conditions for the basic Solow model, the coefficient for the variable GDP60 changes to a negative coefficient. This means that based on the solow model with conditions, countries that are poorer do grow faster than countries with more money. In addition to this, the augmented solow model supports this claim even further. Not only is the coefficient negative, but it is -0.300 which can be compared to -.149 of the solow model. This means that with the condition of SCHOOL in place, countries actually converge faster. The coefficient under the solow model is significant at a p-value of 0.01 while the augmented solow model is significant at 0.001. This is consistent with expectations because the Solow model predicts convergence.

Conclusion

Throughout the report we have sought to determine whether poorer countries develop faster than rich countries. In order to find a solution, we utilized the Solow growth model in two forms that helped us control and determine what independent variables had the most impact on the dependent variable. In the end, we found that poor countries typically grow faster than richer countries. This result was derived from the changes in results when changing to a conditional convergence test from an unconditional one. It can then also be concluded that the Solow model was successful in its prediction that poor countries eventually converge into countries with higher GDPs.

References

-Robert J. Barro, “Economic Growth in a Cross Section of Countries” *The Quarterly Journal of Economics.* Vol. 106 May 1991: 407- 444.

-Mankiw, G., Romer, D., and Weil, D. “A contribution to the Empirics of Economic Growth.” *The Quarterly Journal of Economics.* Vol. 107, May 1992: 407-437. -

Solow, R. “A Contribution to the Theory of Economic Growth.” *The Quarterly of Economic Growth.* Vol. 29, 1956: 65-94.

Appendix

1. sum gdp85 popgr iy school

Variable Obs Mean Std. dev. Min Max

gdp85 98 5309.765 5277.183 412 19723

popgr 98 2.20102 .8898621 .3 4.3 iy 98

17.67245 7.91833 4.1 36.9 school 98

5.396939 3.468992 .4 11.9

2. . corr lgdp85 lpopgr liy lschool(obs=98) lgdp85 lpopgr liy lschool

lgdp85 1.0000 lpopgr -0.4781 1.0000

liy 0.7406 -0.3107 1.0000

lschool 0.8123 -0.2655 0.6327 1.0000

3.reg lgdp85 liy lpopgr

Source SS df MS Number of obs = 98 -F(2, 95) = 76.37

Model 69.6771988 2 34.8385994 Prob > F = 0.0000 Residual 43.3362069 95 .456170599

R-squared = 0.6165 -Adj R-squared = 0.6085

Total 113.013406 97 1.16508666 Root MSE = .6754

lgdp85 Coefficient Std. err. t P>|t| [95% conf. interval]

liy 1.387103 .1414873 9.80 0.000 1.106215 1.667991 lpopgr -.5097814 .1241232 -4.11

0.000 -.7561971 -.2633657 _cons 4.561814 .4281929 10.65 0.000 3.711744 5.411885

4. reg lgdp85 liy lpopgr lschool

Source SS df MS Number of obs = 98 -F(3, 94) = 120.79

Model 89.7355123 3 29.9118374 Prob > F = 0.0000 Residual 23.2778934 94 .247637164

R-squared = 0.7940 -Adj R-squared = 0.7875

Total 113.013406 97 1.16508666 Root MSE = .49763

5. reg gdpdif lgdp60

Source SS df MS Number of obs = 98 -F(1, 96) = 3.61

Model .701048806 1 .701048806 Prob > F = 0.0603 Residual 18.6277499 96 .194039061

R-squared = 0.0363 -Adj R-squared = 0.0262

Total 19.3287987 97 .199265966 Root MSE = .4405

gdpdif Coefficient Std. err. t P>|t| [95% conf. interval]

lgdp60 .0943109 .0496172 1.90 0.060 -.0041785 .1928003 _cons -.2665776 .3796046 -0.70 0.484 -1.020087 .4869316

6. eststo: reg gdpdif lgdp60 lpopgr liy

Source SS df MS Number of obs = 98 -F(3, 94) = 21.41

Model 7.84571452 3 2.61523817 Prob > F = 0.0000 Residual 11.4830841 94 .12216047

R-squared = 0.4059 -Adj R-squared = 0.3869

Total 19.3287987 97 .199265966 Root MSE = .34951

gdpdif Coefficient Std. err. t P>|t| [95% conf. interval]

lgdp60 -.1488957 .0527075 -2.82 0.006 -.2535476 -.0442437 lpopgr -.0885808 .0693268 -1.28 0.204 -.2262307 .0490691 liy .6460532 .0864117 7.48 0.000 .4744808 .8176257 _cons -.1408021 .3659401 -0.38 0.701 -.8673849 .5857806

7. eststo: reg gdpdif lgdp60 lpopgr liy lschool

Source SS df MS Number of obs = 98 -F(4, 93) = 22.49

Model 9.50346 4 2.375865 Prob > F = 0.0000 Residual 9.82533866 93 .105648803

R-squared = 0.4917 -Adj R-squared = 0.4698

Total 19.3287987 97 .199265966 Root MSE = .32504

gdpdif Coefficient Std. err. t P>|t| [95% conf. interval]

lgdp60 -.299685 .0620616 -4.83 0.000 -.4229272 -.1764429 lpopgr -.135011 .0655283 -2.06 0.042 -.2651373 -.0048847 liy .5209638 .0863419 6.03 0.000 .3495059 .6924218 lschool .2340072 .0590749 3.96 0.000 .1166963 .3513182 _cons 1.058585 .4555109 2.32 0.022 .1540305 1.963139

8.. esttab m1 m2, r2 scalar(F)

(1) (2) gdpdif

gdpdif

lgdp60 -0.149** -0.300*** (-

2.82) (-4.83)

liy 0.646*** 0.521***

(7.48) (6.03) project 3 fresults Sunday April 24 20:21:48

2022 Page 8 lpopgr -0.0886 -0.135*

(-1.28) (-2.06)

lschool 0.234***

(3.96)

_cons -0.141 1.059*

(-0.38) (2.32)

N 98 98

R-sq 0.406 0.492

F 21.41 22.49

Term Project: Life Expectancy

ECON 495 Section 1001

Spring 2022

Tyler Pohlmann

Abstract:

The purpose of this project is to discuss life expectancy among the countries of the world and the factors that drive it. The dataset comes from Kaggle with the information being collected by the World Health Organization. This paper will determine if and which variables influence life expectancy the most.

Introduction

This project will discuss the variables that determine a country's life expectancy. The data in question comes from a WHO dataset on Kaggle. The data set contains several variables including country name, year, life expectancy, Infant Deaths, Health Expenditure, Country BMI, Gross Domestic Product, and Schooling from the years 2000-2015. The goal of the project is to determine which if any variables influence life expectancy. In addition, the paper will detail the magnitude to which each variable affects life expectancy. Countries could potentially use this data to make policy decisions in order to better their constituents health. By the end of the paper, it should be clear on which variables influence life expectancy the most so that politicians of the countries can have a clear vision on policy.

Model

Table 1: Variables, Variable Definition, and Expected Signs

Dependent Variable	Variable Definition	Expected Sign
LifeExp	Country Life Expectancy	N/A
Independent Variables		
InfantDeath	Number of Infant Deaths	Negative
HealthExp	% of GDP spent on healthcare	Positive

BMI	Average BMI of Country	Negative
GDP	Gross Domestic Product per Capita	Positive
Schooling	Average years of schooling	Positive

The dependent variable in the model is life expectancy. The independent variables are: InfantDeath, HealthExp, BMI, GDP, and Schooling. InfantDeath refers to the number of infant deaths a country has each year. It is expected that InfantDeath has a negative correlation with life expectancy because the low age of infants should drive down the expected living age. HealthExp refers to a country's expenditure on health-related items as a percentage of their GDP. This is expected to have a positive correlation with life expectancy as a country with better healthcare should save more lives forcing life expectancy up. The next variable is BMI. BMI refers to a country's average Body Mass Index. A lower BMI is healthier, so we expect this variable to have a negative correlation with life expectancy. GDP refers to a country's gross domestic product per capita which is expected to have a positive correlation with life expectancy. Finally, Schooling refers to the average number of years a person in a certain country is schooled. It is expected that more schooling equates to a higher life expectancy.

Descriptive Statistics

Table 2: Summary Statistics of Variables (Mean, Standard Deviation, Minimum, Maximum)

Variable	Obs	Mean	Std. dev.	Min	Max

LifeExp	2482	69.38864	9.637862	36.3	89
InfantDeath	2487	31.24889	126.6175	0	1800
HealthExp	2487	872.128	2133.523	0	19479.91
BMI	2463	38.44458	19.97845	1.4	87.3
GDP	2487	7491.826	14276.59	1.68	119172.7
Schooling	2487	12.0817	3.399877	0	20.7

This section will detail the descriptive statistics (mean, standard deviation, minimum, maximum) for all the variables in the model. Life expectancy has a mean of 69.38864 with a standard deviation of 9.637862. This means the average life expectancy in the model is about 69 years old. Additionally, the lowest life expectancy in the model is 36.3 years and the highest is 89 years. The next variable, infant deaths, has an average of 31.24889 deaths in the model with a standard deviation of 126.6175. The country with the lowest infant deaths is 0 and the country with the highest has 1800 infant deaths. The average health expenditure in the dataset is 872.128 percent of a country's GDP with a standard deviation of 2133.523. The minimum a country spent on health-related expenses is 0 and the maximum was 19479.91% of their country's GDP. BMI had an average of 38.44458 in the model with a standard deviation of 19.97845. The minimum average BMI for a country is 1.4 with the maximum average being 87.3. The mean GDP per capita in the model was 7491.826 with a standard deviation of 14276.56. The minimum GDP per capita in the model was 1.68 and the maximum was 119172.7. The average amount of schooling in the model was 12.0817 years with a standard deviation of 3.399877. The minimum amount of schooling in the model is 0 and the maximum is 20.7 years.

Correlation Matrix

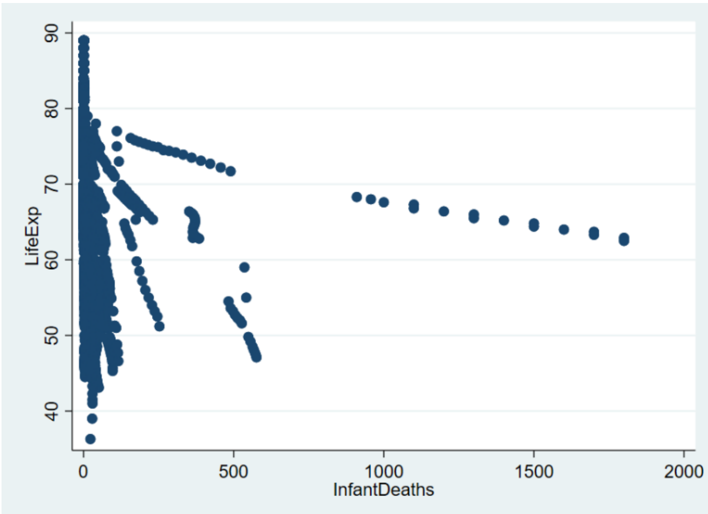
Table 3: Correlation Matrix

	LifeExp
InfantDeath	-0.1840
HealthExp	0.4069
BMI	0.5782
GDP	0.4599
Schooling	0.7545

The table above shows the correlation coefficients between the dependent and independent variables. The correlation coefficient between life expectancy and infant deaths is -0.1840 showing a negative weak correlation between the two variables. The correlation coefficient between life expectancy and health expenditure is 0.4069 showing a moderate positive correlation between health expenditure and life expectancy. The correlation coefficient between BMI and life expectancy is 0.5782 showing a strong positive relationship. GDP and life expectancy have a correlation coefficient of 0.4599 showing a moderate positive relationship. Finally, schooling and life expectancy have the strongest correlation coefficient of the model at 0.7545 showing a strong positive correlation.

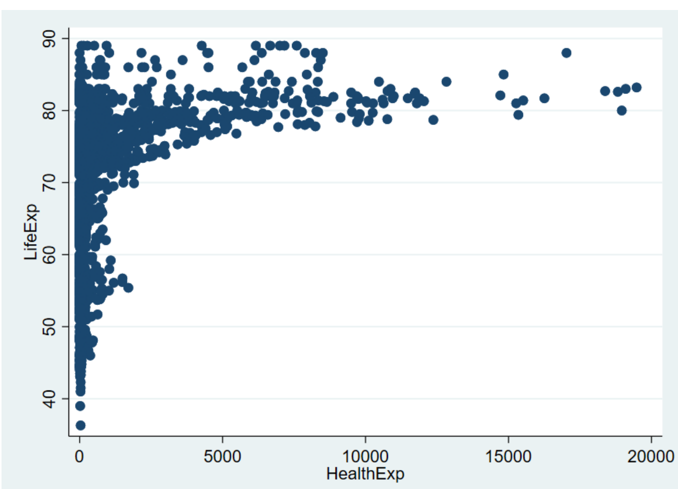
Scatter Plots

Scatter Plot 1: InfantDeath and LifeExp



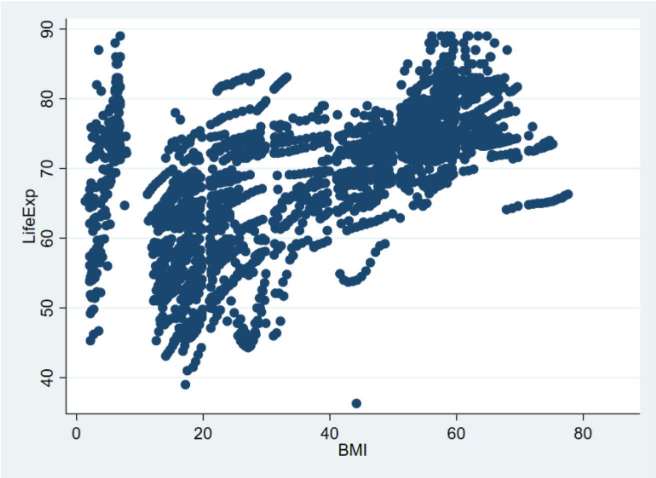
The scatter plot above shows the scatter plot relating infant deaths and life expectancy. The data points above show a weak negative correlation between the two variables. This also supports the correlation coefficient which supported a weak negative relationship between infant deaths and life expectancy. This was the expected sign when discussing the expected correlation in the first section.

Scatter Plot 2: HealthExp and LifeExp



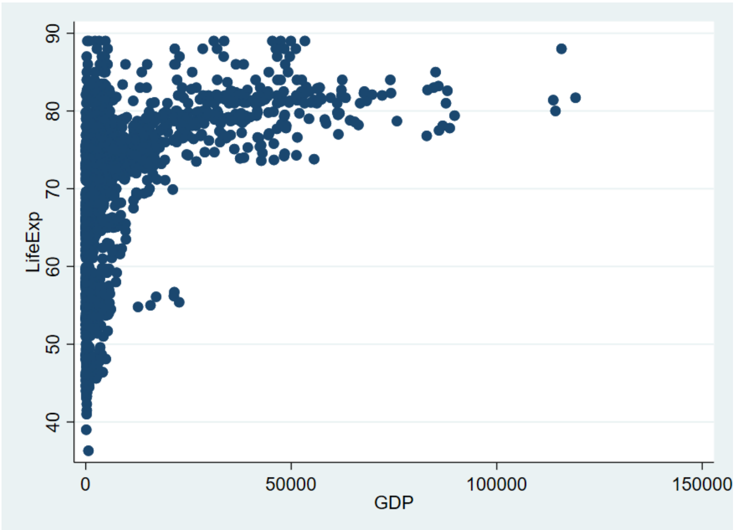
The scatter plot shown above depicts the relationship between health expenditure and life expectancy. The data points show a moderate positive relationship between health expenditure and life expenditure. This claim was supported by the correlation coefficient in the previous section, as well as the prediction discussed in the first section of the paper.

Scatter Plot 3: BMI and LifeExp



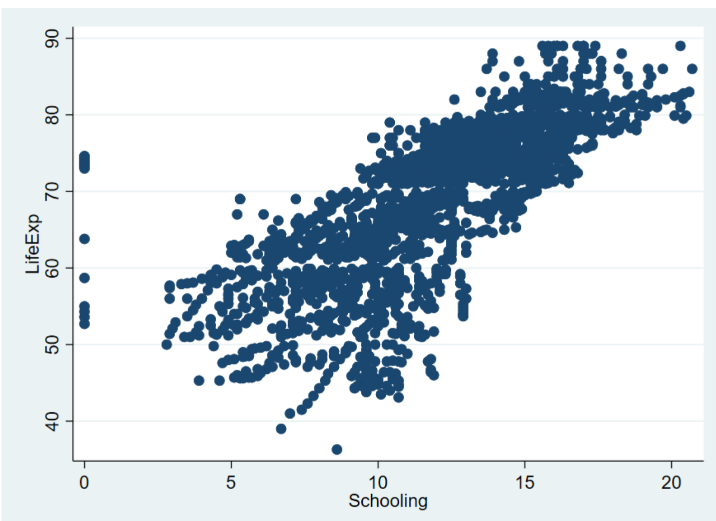
The scatter plot above shows a strong positive relationship between BMI and life expectancy. This claim was supported by the correlation coefficient in the previous section. However, the predicted sign of the correlation between these two variables in the first section was negative. Generally, a lower BMI is a healthy person. Above, we can see that people with higher BMI is correlated to live longer. This could be explained to be more of a wealth factor than a health factor. For example, someone with more money can buy more food and have a higher BMI. A person with more money may also be able to afford more healthcare causing them to have a longer life expectancy. This could be one reason that the predicted sign and actual correlation have a different sign.

Scatter Plot 4: GDP and LifeExp



The scatter plot above shows a moderate positive relationship between GDP and life expectancy. This was also supported by the correlation coefficient in the previous section as well as the prediction laid forth in the first section of the paper.

Scatter Plot 5: Schooling and LifeExp



This scatter plot depicts the relationship between life expectancy and schooling. The plot above shows a strong positive relationship between schooling and life expectancy. The correlation coefficient supports this claim and positive was also the expected sign in the first section of the project.

Empirical Results

Table 4: Regression Results

	LifeExp
InfantDeaths	-0.000227
	(-0.23)
HealthExp	-0.00000546
	(-0.04)
BMI	0.101***
	(13.49)
GDP	0.0000931***
	(4.71)
Schooling	1.658***
	(34.98)

_cons	44.77***
	(90.25)
R-squared	0.6164
F	787.86
N	2458

The table above shows the regression results from the model. InfantDeaths had a coeffiecient of -0.000227. This means every infant death had a decrease of 0.000227 years on life expectancy which was significant at no levels. HealthExp had a coefficient of -0.00000546 significant at no levels. This means that an increase of 1% of health expenditure decreased the life expectancy by 0.00000546 years. BMI had a coefficient of 0.101 significant at all levels. This means that for an increase in 1 BMI increased the life expectancy by 0.101. GDP had a coefficient of 0.0000931 significant at all levels. This means that a $1 increase per capita GDP raises the life expectancy by 0.0000931 years. Finally, schooling had a coefficient of 1.658 significant at all levels. This means that an increase of 1 year of school increased the life expectancy by 1.658 years. The R-squared of the model was 0.6164 meaning that 61.64% of the variation in the model can be explained by the variables.

Conclusion

The goal of this report was to determine which, if any, variables had an effect on the life expectancy of a country. Through the regression and correlation coefficients, it is clear that schooling has the most effect on life expectancy. Schooling and life expectancy had the highest correlation coefficient in the model, and the coefficient in the regression model was high and was significant at all levels. While some other variables like GDP and BMI were positive and significant at all levels, an increase in both those areas only made small changes in the overall life expectancy in the country. Based on the regression model, a country can increase their life expectancy by 1.658 years by increasing the average education level by one year. Politicians and lawmakers of countries looking to raise the life expectancy of their constituents should look to education funding in order to achieve their goals.

References

World Health Organization, Kumar Rajarshi, Kaggle. 2015. *Life Expectancy (WHO).*

Appendix

```
ECON 495 Section 1001  Wednesday May 11 22:04:05 2022   Page 1

                                                        (R)
                                    Statistics/Data analysis

                                      User: Tyler Pohlmann
                                   Project: Term Project

                                            ©
                                   17.0
                                   SE—Standard Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: 40-student lab perpetual
Serial number: 401706336008
  Licensed to: UNLV
               UNLV

Notes:
      1. Unicode is supported; see help unicode_advice.
      2. Maximum number of variables is set to 5,000; see help set_maxvar.

1 . *(8 variables, 2487 observations pasted into data editor)

2 . sum lifeexp infantdeaths healthexp bmi gdp schooling
```

Variable	Obs	Mean	Std. dev.	Min	Max
lifeexp	2,482	69.38864	9.637862	36.3	89
infantdeaths	2,487	31.24889	126.6175	0	1800
healthexp	2,487	872.128	2133.523	0	19479.91
bmi	2,463	38.44458	19.97845	1.4	87.3
gdp	2,487	7491.826	14276.59	1.68135	119172.7
schooling	2,487	12.0817	3.399877	0	20.7

```
3 . scatter lifeexp infantdeaths

4 . graph export "C:\Users\pohlmt1\Desktop\infant deaths scatter.png", as(png) na
> me("Graph")
file C:\Users\pohlmt1\Desktop\infant deaths scatter.png saved as PNG format

5 . scatter lifeexp healthexp

6 . graph export "C:\Users\pohlmt1\Desktop\heathexp scatter.png", as(png) name("G
> raph")
file C:\Users\pohlmt1\Desktop\heathexp scatter.png saved as PNG format

7 . scatter lifeexp bmi

8 . graph export "C:\Users\pohlmt1\Desktop\bmi scatter.png", as(png) name("Graph"
> )
file C:\Users\pohlmt1\Desktop\bmi scatter.png saved as PNG format
```

```
 9 . scatter lifeexp gdp\
   \ invalid name
   r(198);

10 . scatter lifeexp gdp

11 . graph export "C:\Users\pohlmt1\Desktop\gdp scatter.png", as(png) name("Graph"
   > )
   file C:\Users\pohlmt1\Desktop\gdp scatter.png saved as PNG format

12 . scatter lifeexp schooling

13 . graph export "C:\Users\pohlmt1\Desktop\schooling.png", as(png) name("Graph")
   file C:\Users\pohlmt1\Desktop\schooling.png saved as PNG format

14 . corr lifeexp infantdeaths healthexp bmi gdp schooling
   (obs=2,458)

                |  lifeexp infant~s health~p      bmi      gdp school~g
   -------------+------------------------------------------------------
        lifeexp |   1.0000
   infantdeaths |  -0.1840   1.0000
      healthexp |   0.4069  -0.0905   1.0000
            bmi |   0.5782  -0.2282   0.2520   1.0000
            gdp |   0.4599  -0.1080   0.8992   0.3041   1.0000
      schooling |   0.7545  -0.2070   0.4027   0.5716   0.4515   1.0000


15 . reg lifeexp infantdeaths healthexp bmi gdp schooling

         Source |       SS           df       MS      Number of obs   =     2,458
   -------------+----------------------------------   F(5, 2452)      =    787.86
          Model |  140470.095         5  28094.0189   Prob > F        =    0.0000
       Residual |  87434.5421     2,452  35.6584593   R-squared       =    0.6164
   -------------+----------------------------------   Adj R-squared   =    0.6156
          Total |  227904.637     2,457  92.7572799   Root MSE        =    5.9715

   ------------------------------------------------------------------------------
        lifeexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
   -------------+----------------------------------------------------------------
   infantdeaths |  -.0002271   .0009764    -0.23   0.816    -.0021416    .0016875
      healthexp |  -5.46e-06   .0001286    -0.04   0.966    -.0002577    .0002467
            bmi |   .1007213   .0074678    13.49   0.000     .0860775     .115365
            gdp |   .0000931   .0000198     4.71   0.000     .0000544    .0001319
      schooling |   1.657862   .0474006    34.98   0.000     1.564913    1.750812
          _cons |   44.76995   .4960926    90.25   0.000     43.79715    45.74275
   ------------------------------------------------------------------------------

16 . ssc install estout
   checking estout consistency and verifying not already installed...
   installing into C:\Users\pohlmt1\ado\plus\...
   installation complete.
```

```
17 . eststo: reg lifeexp infantdeaths healthexp bmi gdp schooling

      Source |       SS           df       MS      Number of obs   =     2,458
-------------+----------------------------------   F(5, 2452)      =    787.86
       Model |  140470.095         5  28094.0189   Prob > F        =    0.0000
    Residual |  87434.5421     2,452  35.6584593   R-squared       =    0.6164
-------------+----------------------------------   Adj R-squared   =    0.6156
       Total |  227904.637     2,457  92.7572799   Root MSE        =    5.9715

------------------------------------------------------------------------------
     lifeexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
infantdeaths |  -.0002271   .0009764    -0.23   0.816    -.0021416    .0016875
   healthexp |  -5.46e-06   .0001286    -0.04   0.966    -.0002577    .0002467
         bmi |   .1007213   .0074678    13.49   0.000     .0860775     .115365
         gdp |   .0000931   .0000198     4.71   0.000     .0000544    .0001319
   schooling |   1.657862   .0474006    34.98   0.000     1.564913    1.750812
       _cons |   44.76995   .4960926    90.25   0.000     43.79715    45.74275
------------------------------------------------------------------------------
   (est1 stored)

18 . esttab using lifeexp.rtf
   (output written to lifeexp.rtf)

19 .
```

Video Link

https://www.youtube.com/watch?v=7RRMQhqB98s