The Puzzle of Wealth: Understanding Per Capita GDP Pieces

By: Avery Brazelton, Isabela Barton, Navya Aenugu, Rishi Sarraff

Introduction

GDP is a crucial predictor, offering insights into a nation's financial well-being and stability. A robust economy signifies higher incomes, job opportunities, and improved living standards, fostering a positive cycle of growth and prosperity for the population. This prompts the important question: What factors truly impact per capita GDP? By determining these key factors we hope to gain greater insight into the most impactful ways to support a developing country's financial sucess.

Research Questions

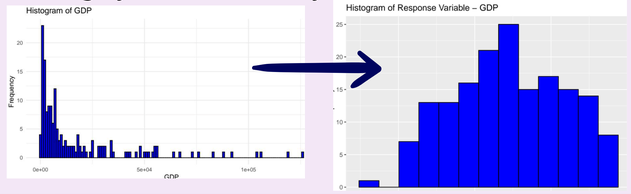
What factors impact a country's GDP?

1. Do countries with high fertility rates have lower GDPs on average?

2. Is there a relationship between the average age of a population and GDP?

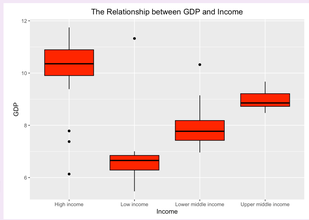
3. What is the most significant predictor of GDP?

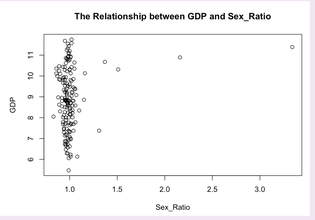
Because the response variable, GDP per capita, was very right skewed the normality assumption was violated. We took a log of GDP per capita and now the normality condition is satisfied because the histogram is roughly unimodal and symmetric.



EDA

Because the GDP means are significantly different for different levels of income and there is a very high correlation between life expectancy and GDP, it makes sense that these two are included in our final model. Additionally, the correlation between GDP and sex ratio is very low and was not found to be significant.



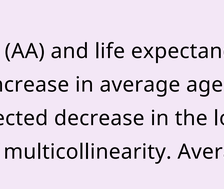


Data Summary

Variable Name	Measurement	
GDP	Gross Domestic Product (GDP) per capita is the total economic output divided by the population of the country	GDP
Life_Expectancy	Life expectancy is the average number of years a person is expected to live,	LE
Density	Population density is the average number of people living per square kilometer of land.	DE
Dependency	The dependency ratio is the ratio of the dependent population to the working-age population. The dependent population is people aged 0-14 and 65+. The working-age population is people ages 15-64.	DP
Fertility	The total fertility rate or TFR is the average number of children that a woman has in the country. Generally, replacement TFR is 2.1.	FR
Sex_Ratio	The gender ratio is the number of males to the number of females in the country.	SR
Debth	Debt as a percentage of GDP is the ratio of a country's total debt to its Gross Domestic Product/GDP.	DB
Income	Income group classification is determined by a country's Gross National Income (GNI) per capita. Lower Income: $1,135 or less of GNI per capita. Middle Income: between $1,136 and $13,845. High Income: $113,846 or more of GNI per capita.	IC
Growth	Population growth rate is the percentage change in a country's population from one year to the next.	GR
Size	The size of a country by population is the total number of people in the country.	SZ
Average Age	The mean age of the country's population	AA

Quantitative and Interactions

Multicollinearity Screening



With high average and individual VIFs, we need to variable screen for pairwise relationship

Add Qualitative Factors

Final: $\log(E(GDP)) = \beta_0 + \beta_1 \log(AA) + \beta_2 LE + \beta_3 \log(AA) * LE + \beta_4 LI + \beta_5 LMI + \beta_6 UMI$

Because Income was the only significant qualitative variable, the other qualitative variables were not added to the model and qualitative interactions were not tested.

- Low Income = 1 if Country is classified as Low Income, 0 otherwise
- Lower Middle Income = 1 if Country is classified as Lower Middle Income, 0 otherwise
- Upper Middle Income = 1 if Country is classified as Upper Middle Income, 0 otherwise
- Base Case: High Income = 0

Check Assumptions

Normality: histogram of residuals is roughly unimodal, symmetric

Mean Zero: no clear pattern that could be modeled into plots

Constant Variance: no fanning pattern in residual plots

Analysis

Variable

Step	Variable	Added/ Removed
1	Life_Expectancy	addition
2	Average.Age	addition
3	Sex_Ratio	addition
4	Dependency	addition
5	Fertility	addition

Initial Model: $\ln(GDP) = B0 + B1\log(AA) + B2(LE) + B3(SR) + B4(DP) + B5(FR)$

Final Model: $\ln(GDP) = B0 + B1\log(AA) + B2(LE) + B3(SR) + B4\log(AA*LE)$

We used the results of the stepwise selection to create our initial model. After confirming thee interaction between age and life expectancy is significant by comparing it to the model without the interation, we will add the qualitative terms

Summary Statistics of Final Model & Output:

$$\widehat{\log(GDP)} = 23.3857 - 5.3241\log(AA) - 0.1848LE + 0.0726\log AA * LE - 3.0819LI - 1.8835LMI - 0.9782UMI$$

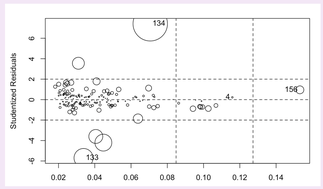
Adjusted R^2: 0.8684

F-Statistic: 175.8 on 6 and 153 DF

P-value:< 2.2e-16

Outliers

Five influential outliers had dffits significantly different from the average, so we conclude these are not truly representative of the population. The model without these values had higher adjusted r squared, higher p-values, lower residual standard error.



Box Cox Analysis

The Box-Cox Analysis confirmed that taking the log of GDP was the best transformation to apply because our gamma result was very close to 0.

Conclusion

An increase in life expectancy (LE) and the interaction between the log average age (AA) and life expectancy (LE) (logAA*LE) is associated with an expected increase in the log of GDP per capita. Conversely, an increase in average age (AA), lower-income (LI), middle-income (MI), and upper-middle-income (UMI) is associated with an expected decrease in the log of GDP per capita. Fertility is not a good predictor of GDP because of evidence of high levels of multicollinearity. Average age and income are much more significant.

Our model is adequate with an R2 of 0.8733, meaning 87.33% of the variability observed in GDP per capita is explained by the regression model. This is much better than the first model with an R2 of 0.6393 and the pre-residual analysis model with an R2 of 0.7396. Our final model provides the best fit with the best parameters. We determined that average age, life expectancy, and income group are significant indicators of the financial stability and economic success of a country.

Limitations

- Countries only included if present in all datasets used, selection bias against small / less developed countries
- Countries have different policies when it comes to how frequently they collect census data from their populations, so information may not be up to date for certain countries
- Data from 2020 may have been inaccurate or irregular due to effects of the COVID-19 pandemic

Future Steps

- Do a similar data analysis of the data collected in 2030 to correct for pandemic data
- Compare our current GDP findings with findings from data from 2010 to see which countries have GDPs that have improved or worsened over time. By viewing the trends, we can make even more accurate predictions
- Do further analysis on the relationship between a country's capital, labor, and total factor productivity.

Avery Brazelton, Isabela Barton, Navya Aenugu, Rishi Sarraff (NARB)

Professor Varanyak

STAT 3220: Introduction to Regression Analysis

November 29th, 2023

Works Cited

About Us. (2023). Focus Economics. https://www.focus-economics.com

Andrew Chatzky. (2021, May 11). Why Does the Census Matter? Council on Foreign Relations. https://www.cfr.org/backgrounder/why-does-census-matter

Country Indicators. (n.d.). [dataset]. https://www.focus-economics.com/country-indicator/

International Database. (2023, July 12). [dataset]. United States Census Bureau. https://www.census.gov/programs-surveys/international-programs/about/idb.html#:~:text=Methodology,one%20of%20our%20primary%20resources.

GDP per capita. (2021). [Map]. Accessed November 28, 2023. Our World in Data. https://ourworldindata.org/grapher/gdp-per-capita-worldbank

How to Forecast GDP in the Long Run: A Primer. (2022, October 20). The Conference Board. https://www.conference-board.org/topics/recession/how-to-forecast-GDP-in-the-long-run#:~:text=Longer%2Dterm%20forecasts%20of%20GDP,buildings%2C%20machinery%2C%20or%20software.

Kunal Sen. (2021, January). Why should I care about economic growth? UNU-WIDER. https://www.wider.unu.edu/publication/why-should-i-care-about-economic-growth#:~:text=Growth%20creates%20wealth%2C%20some%20of,and%20gain%20improved%20living%20standards.

Linda Jacobsen. (2023, February 10). How Accurate Was the 2020 Census—And Why Should You Care? Population Reference Bureau. https://www.prb.org/resources/how-accurate-was-the-2020-census-and-why-should-you-

care/#:~:text=The%20PES%20estimated%20a%20net,2010%20to%205.8%25%20in%202020.

Matthias Doepke. (2022, July 22). THE NEW ECONOMICS OF FERTILITY. International Monetary Fund. https://www.imf.org/en/Publications/fandd/issues/Series/Analytical-Series/new-economics-of-fertility-doepke-hannusch-kindermann-tertilt#:~:text=It%20suggests%20that%20as%20parents,across%20countries%20and%20over%20time.

Our Censuses. (2022, August 4). United States Census Bureau. https://www.census.gov/programs-surveys/censuses.html#:~:text=It%20helps%20the%20government%20decide,each%20State%20holds%20in%20Congress.&text=The%20U.S.%20census%20counts%20every%20resident%20in%20the%20United%20States.

Population and Economic Growth. (n.d.). Vaia. Retrieved October 19, 2023, from https://www.hellovaia.com/explanations/macroeconomics/economic-performance/population-and-economic-growth/#:~:text=For%20example%2C%20an%20increase%20in,the%20country%20as%20a%20result!

Top Questions About the Survey. (2023, June 27). United States Census Bureau. https://www.census.gov/programs-surveys/acs/about/top-questions-about-the-survey.html