A black and white logo

Description automatically generated

**Analysing the Relationship Between Education and Economic Output**

*Business Statistics*

**Angel E. Lanto**

Master’s in Business Analytics

**Objective**

Investigate the relationship between the level of education and its economic output at country

level using simple linear regression analysis.

**Research Question**

How does the average years of schooling in a country correlate with its GDP per capita?

**Hypothesis:**

Ho: The level of education does not have a significant influence on the economic output at country level.

Ha: The level of education has a significant influence on the economic output at country level.

**Data Source**

World Bank's World Development Indicators (WDI) database:

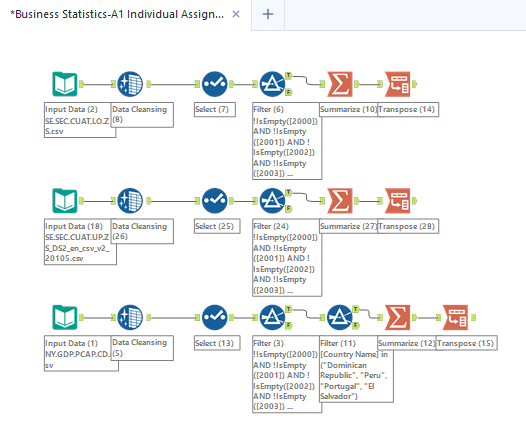
https://data.worldbank.org/indicator

1. **Data Collection**

This quantitative study used a correlational design to analyze the influence of average schooling on a country’s GDP per capita.

Since World Bank does not provide a direct indicator for "average years of schooling," the researchers obtained and used the data on educational attainment—the percentage of the population ages 25+ that completed lower (SE.SEC.CUAT.LO.ZS) and upper secondary education (SE.SEC.CUAT.UP.ZS)—and GDP per capita in USD (NY.GDP.PCAP.CD). Using educational attainment rates as a proxy for "average years of schooling" is effective because it reflects key educational milestones that correlate with total schooling years and human capital quality. These metrics are widely available, align with economic outcomes, and capture the impact of education on GDP per capita, making them a practical substitute when direct data is lacking. Although a simple linear regression would suffice, an additional education level was included to estimate average years of schooling more accurately.

1. **Data Preparation**



*Figure 1. Data Cleaning in Alteryx*

The researcher used Alteryx to prepare the data by cleaning the datasets: replacing nulls with blanks for string fields, removing unwanted characters, and trimming whitespaces. Furthermore, the years 2000–2020 was selected for analysis, filtered out rows with blanks, and calculated yearly averages for all countries with complete data (e.g., Dominican Republic, Peru, Portugal, and El Salvador). Finally, the data was transposed from rows to columns, and organized it in an Excel table for accurate data types and statistical analysis readiness.

1. **Exploratory Data Analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *SE.SEC.CUAT.LOZ.ZS* | | *SE.SEC.CUAT.UP.ZS* | | *NY.GDP.PCAP.CD* | |
| Mean | 47.83690475 | Mean | 35.82018011 | Mean | 8262.267895 |
| Standard Error | 1.434782576 | Standard Error | 1.212506475 | Standard Error | 427.3919849 |
| Median | 47.48249912 | Median | 35.42400055 | Median | 9023.680596 |
| Mode | #N/A | Mode | #N/A | Mode | #N/A |
| Standard Deviation | 6.574999758 | Standard Deviation | 5.556402703 | Standard Deviation | 1958.556122 |
| Sample Variance | 43.23062182 | Sample Variance | 30.873611 | Sample Variance | 3835942.083 |
| Kurtosis | -1.164188314 | Kurtosis | -1.128809882 | Kurtosis | -0.656743563 |
| Skewness | -0.064413471 | Skewness | 0.115536621 | Skewness | -0.820765773 |
| Range | 21.36749887 | Range | 18.23946113 | Range | 6112.392161 |
| Minimum | 36.96500063 | Minimum | 27.57400017 | Minimum | 4572.844374 |
| Maximum | 58.3324995 | Maximum | 45.813`4613 | Maximum | 10685.23654 |
| Sum | 1004.575 | Sum | 752.2237824 | Sum | 173507.6258 |
| Count | 21 | Count | 21 | Count | 21 |
|  | 1 |  | 3 |  | 5 |

*Table 1. Descriptive Statistics*

The mean values of secondary education attainment (lower and upper) show that the lower secondary education rate is generally higher than the upper secondary rate, indicating more people complete lower secondary education compared to upper secondary.

The GDP per capita has a much higher variance and standard deviation compared to education indicators, suggesting greater fluctuations in economic output relative to educational attainment rates.

Both educational attainment variables have skewness values near zero, indicating a roughly symmetric distribution. However, the GDP per capita has a negative skew, suggesting a tail towards higher values, which might indicate that a few countries have significantly higher GDP per capita than the majority.

Negative kurtosis in all variables indicates a distribution with lighter tails than a normal distribution, implying fewer extreme values.

*Figure 2. Significant Influence of Educational Attainment (at least lower and secondary) on GDP per capita (in USD)Significant Influence of Educational Attainment (at least lower and upper secondary) on GDP per capita (in USD)*

The scatter plot suggests a positive relationship between educational attainment (both lower and upper secondary completion rates) and GDP per capita over time. As the percentage of individuals who completed lower and upper secondary education increases, GDP per capita also tends to rise, indicating that countries with higher educational attainment may have higher economic outputs.

Furthermore, it shows an upward trend for GDP per capita and educational attainment over the years, which may imply that as time progresses, both GDP and education levels are increasing, potentially due to global improvements in education and economic development.

1. **Linear Regression Analysis**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.961277825 |  |  |  |  |  |  |  |
| R Square | 0.924055057 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.91561673 |  |  |  |  |  |  |  |
| Standard Error | 568.9370245 |  |  |  |  |  |  |  |
| Observations | 21 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 2 | 70892433.58 | 35446216.79 | 109.5069026 | 8.40407E-11 |  |  |  |
| Residual | 18 | 5826408.08 | 323689.3378 |  |  |  |  |  |
| Total | 20 | 76718841.66 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -10710.79122 | 1648.548183 | -6.497105351 | 4.13833E-06 | -14174.26243 | -7247.320009 | -14174.26243 | -7247.320009 |
| SE.SEC.CUAT.LOZ.ZS | 1376.006528 | 255.6966957 | 5.38140129 | 4.09752E-05 | 838.8077045 | 1913.205352 | 838.8077045 | 1913.205352 |
| SE.SEC.CUAT.UP.ZS | -1307.945241 | 302.5708903 | -4.322772887 | 0.000409877 | -1943.623093 | -672.2673888 | -1943.623093 | -672.2673888 |

*Table 2. Summary Output of the Linear Regression*

**Regression Equation**

Based on the regression output:

where:

* represents GDP per capita.
* is the percentage of the population with at least lower secondary education.
* is the percentage of the population with at least upper secondary education.

**Interpretation of the Slope and Intercept :**

* **Intercept :** This represents the estimated GDP per capita when both educational attainment indicators (lower and upper secondary completion rates) are zero. In practice, it has little real-world interpretive value, as these educational attainment values are unlikely to be zero in a real scenario.
* **Slope for Lower Secondary Education :** For each 1% increase in the population that completed lower secondary education, GDP per capita is expected to increase by approximately 1376.01 USD, assuming other variables remain constant. This positive coefficient suggests that lower secondary education has a beneficial effect on GDP per capita.
* **Slope for Upper Secondary Education** For each 1% increase in the population that completed upper secondary education, GDP per capita is expected to decrease by approximately 1307.95 USD, assuming other variables remain constant. This negative coefficient could imply that increasing upper secondary attainment without further educational progress or skill development may not directly translate to higher economic output, though this result may need further investigation for context.

**Coefficient of Determination (R-squared)**

* **R-squared = 0.924**: This value indicates that approximately 92.4% of the variation in GDP per capita can be explained by the educational attainment variables (lower and upper secondary completion rates). This high R-squared value suggests a strong relationship between these educational factors and economic output.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| RESIDUAL OUTPUT | | | |  | PROBABILITY OUTPUT | |
| *Observation* | *Predicted NY.GDP.PCAP.CD* | *Residuals* | *Standard Residuals* |  | *Percentile* | *NY.GDP.PCAP.CD* |
| 1 | 4088.008664 | 484.8357105 | 0.89827466 |  | 2.380952381 | 4572.844374 |
| 2 | 5132.087405 | -465.7996528 | -0.863005788 |  | 7.142857143 | 4666.287752 |
| 3 | 5785.664577 | -752.4342813 | -1.394065316 |  | 11.9047619 | 5033.230295 |
| 4 | 6140.719095 | -510.4940832 | -0.945812961 |  | 16.66666667 | 5630.225012 |
| 5 | 6816.336359 | -516.2822887 | -0.956537002 |  | 21.42857143 | 6300.05407 |
| 6 | 7027.604271 | -72.47828355 | -0.134283437 |  | 26.19047619 | 6955.125987 |
| 7 | 7314.636816 | 108.925893 | 0.201811392 |  | 30.95238095 | 7423.562709 |
| 8 | 8151.830249 | 316.6920034 | 0.586748037 |  | 35.71428571 | 8468.522252 |
| 9 | 8184.077042 | 1103.7804 | 2.045018429 |  | 40.47619048 | 8802.006238 |
| 10 | 8004.199512 | 797.8067256 | 1.478128672 |  | 45.23809524 | 9007.669059 |
| 11 | 8292.78461 | 730.895986 | 1.354160448 |  | 50 | 9023.680596 |
| 12 | 9578.158492 | -26.03761341 | -0.048240936 |  | 54.76190476 | 9139.989225 |
| 13 | 8887.50534 | 252.4838847 | 0.467787068 |  | 59.52380952 | 9287.857442 |
| 14 | 9351.057935 | 168.1523038 | 0.31154255 |  | 64.28571429 | 9304.3039 |
| 15 | 9537.472673 | 185.2074164 | 0.343141244 |  | 69.04761905 | 9519.210239 |
| 16 | 9883.283933 | -875.6148747 | -1.622286966 |  | 73.80952381 | 9552.120879 |
| 17 | 9959.416899 | -655.1129991 | -1.213754254 |  | 78.57142857 | 9722.68009 |
| 18 | 10544.03884 | -627.4674319 | -1.162534198 |  | 83.33333333 | 9854.527688 |
| 19 | 10591.78482 | 50.13522469 | 0.092887551 |  | 88.0952381 | 9916.571412 |
| 20 | 10603.34805 | 81.88848632 | 0.151718099 |  | 92.85714286 | 10641.92005 |
| 21 | 9633.610213 | 220.9174743 | 0.409302708 |  | 97.61904762 | 10685.23654 |

*Table 3. Residual Output*

The residual analysis with the provided plots suggests the following:

|  |  |
| --- | --- |
| *Figure 3. Line Fit Plot for Educational Attainment—the percentage of the population ages 25+ that completed lower (SE.SEC.CUAT.LO.ZS)* | *Figure 4. Line Fit Plot for Educational Attainment—the percentage of the population ages 25+ that completed lower (SE.SEC.CUAT.UP.ZS)* |

**Linearity:** The line fit plots indicate a linear relationship between educational attainment (SE.SEC.CUAT.LO.ZS and SE.SEC.CUAT.UP.ZS) and GDP per capita, so the linearity assumption is satisfied.

|  |  |
| --- | --- |
| *Figure 5. Residual Plot for Educational Attainment—the percentage of the population ages 25+ that completed lower (SE.SEC.CUAT.LO.ZS)* | *Figure 6. Residual Plot for Educational Attainment—the percentage of the population ages 25+ that completed lower (SE.SEC.CUAT.UP.ZS)* |

**Homoscedasticity and Independence:** Residual plots show varying spreads, suggesting heteroscedasticity. This might affect the reliability of standard errors. Meanwhile, there's no indication of sequential patterns in the residuals, so independence is likely met.

*Figure 7. Normal Probability Plot*

**Normality:** The normal probability plot shows deviations, indicating residuals may not be normally distributed.

Overall, the model meets linearity and independence assumptions, but heteroscedasticity and non-normality caution against over-reliance on significance tests. Adjustments like robust standard errors or transformation could improve model validity.

**Interpretation and Discussion**

This analysis explored how educational attainment correlates with GDP per capita, using the World Bank’s data on lower and upper secondary education completion as proxies for “average years of schooling.” While it captures general education trends, it excludes tertiary and vocational training, which may also impact GDP. The study is limited to data from 2000–2020 and countries with complete datasets, which may affect the generalizability of the results.

Overall, the analysis found that lower secondary education has a positive and significant impact on GDP per capita (coefficient = 1376.01, p = 0.00004). This suggests that each 1% increase in the population completing lower secondary education is associated with an increase of about 1376 USD in GDP per capita, indicating that foundational education supports economic growth. In contrast, the negative coefficient for upper secondary education (-1307.95, p = 0.0004) suggests that higher levels of secondary education alone may not significantly boost GDP without additional training or skills. The model explains 92.4% of the variance in GDP per capita (R-squared = 0.924), indicating a strong relationship between education and economic output.

**Strength and Direction of the Relationship**

The regression results indicate a strong and positive relationship between lower secondary education completion and GDP per capita, with a coefficient of 1376.01 and a highly significant p-value (p = 0.00004). This positive direction suggests that increasing basic education attainment correlates with higher economic output. Conversely, the relationship between upper secondary education completion and GDP per capita is negative (coefficient = -1307.95, p = 0.0004), implying that secondary education alone may not drive economic growth unless accompanied by further skills or educational levels.

**Limitations**

1. **No Direct Indicator:** The World Bank lacks a specific metric for "average years of schooling," requiring reliance on indirect measures.
2. **Incomplete Proxy**: Educational attainment (secondary completion rates) does not fully capture total schooling years or informal education, which may vary across countries.
3. **Educational Scope**: Only lower and upper secondary education were included, omitting higher education and vocational training impacts.
4. **Country Differences**: The model assumes a universal effect across countries, ignoring national differences in economic structures and education quality.
5. **Missing Influential Factors**: Key variables like health, infrastructure, and economic policies were not considered, which could also affect GDP.

**Potential Confounding Factors**

1. **Labor Market Conditions**: If an economy lacks high-skill jobs, higher education levels might not lead to GDP growth.
2. **Education Quality**: Completion rates do not account for quality, which affects how much education translates to economic productivity.

**Policy Implications**

1. **Strengthen Basic Education**: Investing in access to and quality of lower secondary education can support economic growth.
2. **Link Education to Market Needs**: Expand vocational training and skill development programs to ensure that higher education levels meet job market demands.

In summary, investing in foundational education and aligning secondary education with job markets could help drive economic growth in Dominican Republic, Peru, Portugal, and El Salvador. Future research could include other education levels and additional factors to better understand education’s economic impact.

**Appendix A.**

**Combination of Bar and Line Graph**

**Reference**

OpenAI. (2024). *ChatGPT* (November 13 Version). Retrieved from <https://chat.openai.com/>