Statistical Analysis of WHO Life Expectancy Dataset Using Scatter Plots, Covariance & Chebyshev's Inequality

1. Introduction

The World Health Organization (WHO) provides access to large datasets that help analyze global health and development indicators. One such dataset includes key variables like life expectancy, GDP, and schooling across multiple countries over time. These indicators are essential to understanding the health, education, and economic well-being of nations.

In this project, we selected three continuous numerical variables — **Life Expectancy**, **GDP (Gross Domestic Product)**, and **Schooling (average years of education)** — to explore how they are related. We performed this analysis using Google Sheets, a powerful and accessible data tool that allowed us to visualize patterns, calculate statistical values, and apply formulas such as covariance and Chebyshev's inequality.

The main objectives of this project are:

- To use **scatterplots** to visually understand relationships between variables

- To calculate **covariance** and understand whether two variables increase or decrease together

- To compute the **mean and standard deviation** of each variable

- To apply **Chebyshev's inequality** to determine the spread and typical range of values, even when the data is not normally distributed

The analysis was done using Google Sheets.

2. Scatter Plots & Interpretation
## Chart 1: GDP vs Life Expectancy

GDP vs Life Expectancy
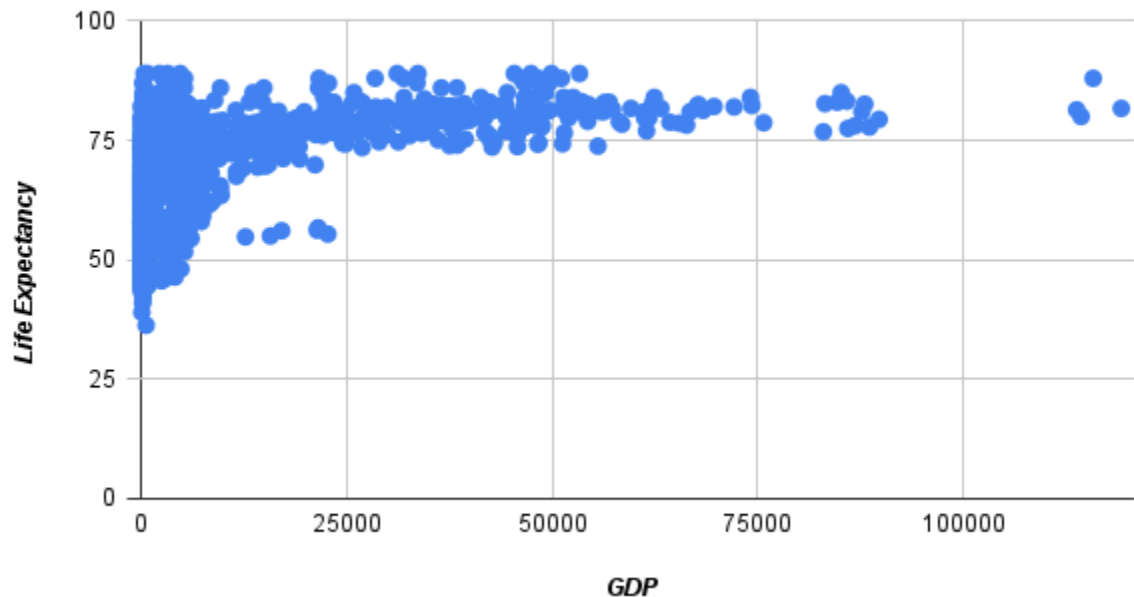
The scatter plot demonstrates that there is a clear positive relationship between GDP and life expectancy. As GDP increases, life expectancy tends to rise as well. At lower GDP levels, life expectancy values are widely spread, with some countries experiencing high life expectancy and others much lower. In contrast, at higher GDP levels, most countries cluster around a high life expectancy, typically between 75 and 85 years, and the variation becomes much smaller. This pattern suggests diminishing returns: after reaching a certain GDP threshold, further increases in GDP lead to only slight improvements in life expectancy. Additionally, a few outliers can be observed, where some countries have unusually high or low life expectancy compared to their GDP.

a. Chart 2: Schooling vs Life Expectancy

Schooling vs Life Expectancy
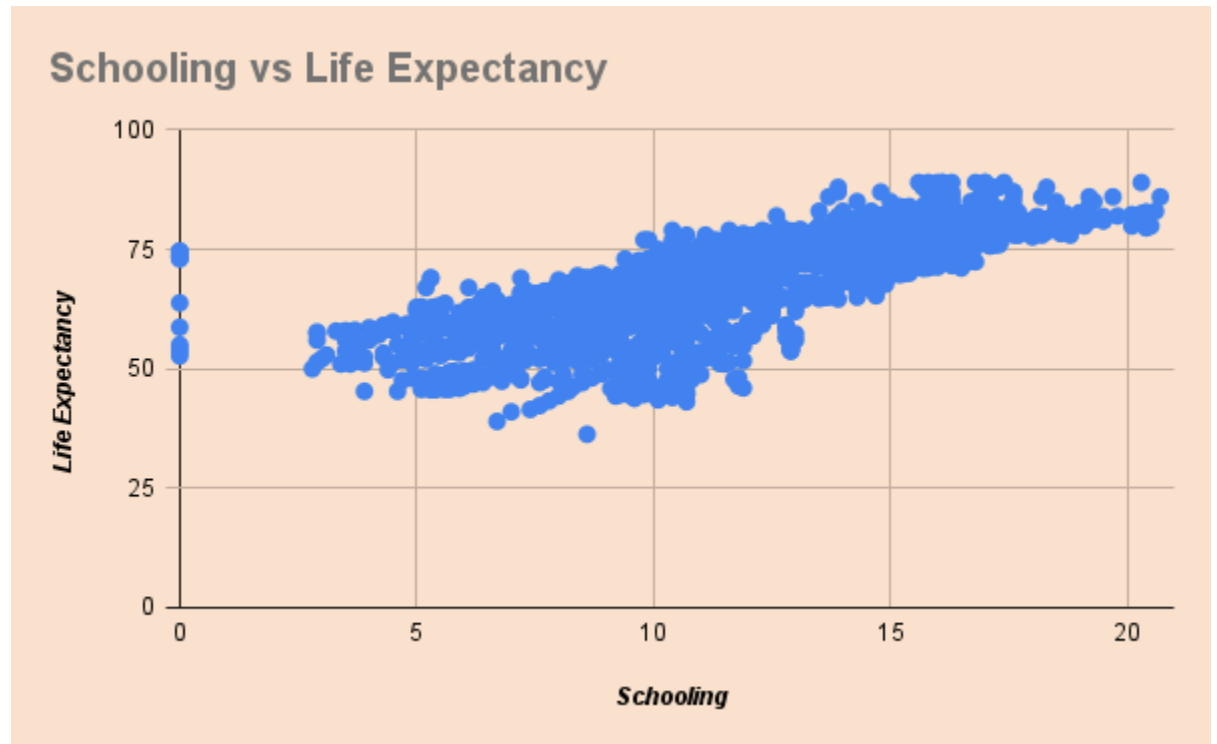
The scatter plot illustrates a clear positive relationship between years of schooling and life expectancy. As the average years of schooling increase, life expectancy also tends to rise. At lower levels of schooling, life expectancy values are more widely dispersed, with some countries exhibiting much lower life expectancy than others. As schooling increases, the data points cluster more tightly, and most countries achieve higher life expectancy, typically ranging from 70 to 85 years. This trend suggests that greater educational attainment is associated with improved health outcomes and longevity. A few outliers are present, indicating that some countries have unusually high or low life expectancy relative to their average years of schooling.

b. Chart 3: Schooling vs GDP
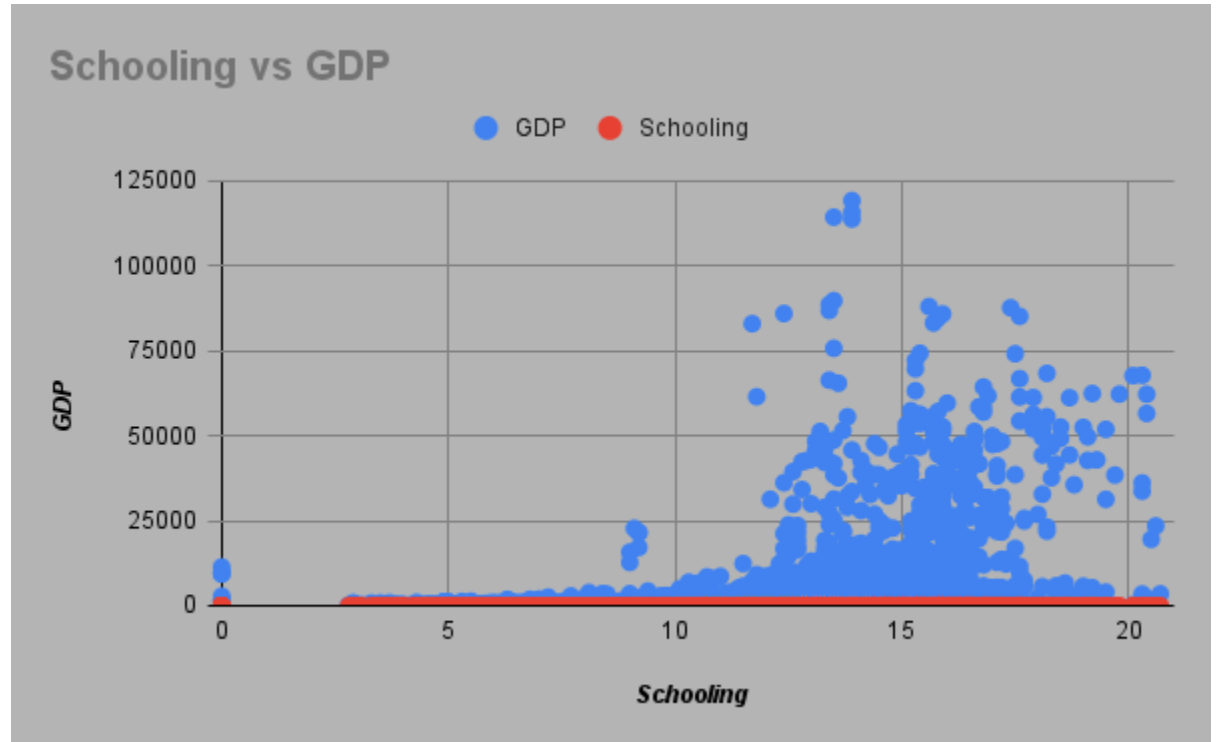
The scatter plot displays the relationship between years of schooling and GDP. As the average years of schooling increase, GDP values also tend to rise, indicating a positive association between education and economic prosperity. At lower levels of schooling, GDP values are generally low and tightly clustered, suggesting limited economic output in countries with less education. As schooling increases, the spread of GDP values becomes much wider, with some countries achieving very high GDP, while others remain lower. This pattern highlights that while higher education is often linked to greater economic output, other factors also influence GDP, leading to significant variability at higher levels of schooling. A few outliers can be observed, where some countries have unusually high or low GDP relative to their average years of schooling

3. Convariance and Interpretation

Variable Pair	Covariance Value
GDP & Life Expectancy	63546.99
Schooling & Life Expectancy	23.55
Schooling & GDP	21749.81

The covariance table shows how pairs of variables—GDP, Life Expectancy, and Schooling—move together in your dataset:

- GDP & Life Expectancy: The large positive covariance (63546.99) indicates that countries with higher GDP generally have higher life expectancy. This suggests a strong positive relationship between economic prosperity and longevity.
- Schooling & Life Expectancy: The small positive covariance (23.55) suggests a weak but positive relationship; as average years of schooling increase, life expectancy tends to rise slightly.
- Schooling & GDP: The positive covariance (21749.81) shows that countries with more years of schooling typically have higher GDP, indicating a positive association between education and economic output.

In summary, all pairs show positive covariance, meaning these variables tend to increase together. The strength of the relationship is strongest between GDP and Life Expectancy, and between Schooling and GDP, while the link between Schooling and Life Expectancy is weaker.

4. Mean and Standard Deviation

Variable	Mean	Standard Deviation
GDP	7483.16	14267.30
Life Expectancy	69.22	9.52
Schooling	11.99	3.36

The table summarizes the mean and standard deviation for GDP, life expectancy, and schooling in the dataset. GDP has the highest standard deviation, indicating it is the most variable among the three variables, while schooling has the lowest standard deviation, showing it is the most consistent. The average life expectancy is about 69.22 years, with most values clustered within approximately 9.5 years of the mean. The mean years of schooling is 11.99, suggesting that, on average, people receive about 12 years of education. Overall, these statistics provide a concise overview of the central tendency and variability for each variable, highlighting that GDP values are widely dispersed, while schooling levels are relatively stable across the dataset.

5. Chebyshev's Inequality (k = 2)

Variable	Mean	Standard Deviation	Lower Bound	Upper Bound
GDP	7483.16	14267.30	-21051.45	36017.77
Life Expectancy	69.22	9.52	50.18	88.27
Schooling	11.99	3.36	5.28	18.71

According to Chebyshev's inequality with k=2k = 2k=2, we can say that at least **75%** of the data values lie within two standard deviations from the mean for each variable.

- For **Life Expectancy**, 75% of the values lie between **50.18 and 88.27** years, showing most countries fall within a typical life span range.

- For **Schooling**, 75% of the values lie between **5.28 and 18.71** years, which reflects the general range of average education levels across countries.

- For **GDP**, the formula gave a lower bound of **−21051.45**, which is not valid in real-world context since GDP cannot be negative. Therefore, we consider the practical lower bound as **0**, and interpret that at least 75% of countries have GDP values between **0 and 36017.77**.

This analysis helps us understand the variability and distribution range of these key development indicators, even without assuming a normal distribution.

6. Conclusion

This analysis helps us understand that higher schooling is generally associated with better life expectancy and higher GDP. However, income inequality across countries causes wide variability in GDP, which is reflected in the high standard deviation and Chebyshev's bounds. This exercise demonstrated the usefulness of scatterplots, covariance, and Chebyshev's Inequality in real-world datasets.