

GDP Predicting Regression Model

Author: Rohan Chawla

The WDI 2020 dataset, sourced from the World Bank, contains development indicators for all countries for 2020. This dataset includes economic metrics such as GDP per capita and demographics such as population. This analysis aims to predict **GDP** using two **independent** variables: **GDP per Capita** and **Total Population**.

A correlation analysis (*Table 1.1*) examined the strength and direction of relationships among the variables. The correlation coefficient of GDP with GDP per Capita and Total Population was found to be 0.14 and 0.58, respectively, indicating a weak and moderate positive relationship, suggesting that **GDP per Capita has less impact on GDP than Total population**.

The regression results (*Table 1.2*) indicate that GDP per Capita and Total Population contribute **significantly** to predicting GDP, as p-values are below 0.05, indicating their impact on GDP, **keeping other factors constant**. The adjusted R-squared value of 0.36 suggests that the two independent variables can explain approximately 36% of the GDP variability. The p-value of the F-statistic is less than 0.05, indicating that the overall regression model is statistically significant.

Graph 1.1 indicates a weak positive trend with a spread of data points, indicating that while GDP per Capita influences GDP, its effect is **not dominant**. The clustering of countries with low GDP per Capita values suggests that **economic performance varies significantly among nations with similar per capita income**. Graph 1.2 indicated a positive trend, suggesting that **GDP is closely tied to population size**. Countries with higher populations have broader GDP values. Regression equations on the plots confirm that Total Population has substantial coefficient than GDP per Capita, hence **a more decisive influence on GDP**.

The regression model fit assumption (*Graph 1.3*) analysis indicates that The Residuals vs. Fitted Values Plot shows a pattern, indicating potential non-linearity. This suggests that the relationship between GDP and the predictors may **not be linear**. The Residuals and Fitted Values Plot also indicates an increasing spread at **higher fitted values**, violating the assumption of constant variance. The Histogram shows a normally distributed bell curve. Hence, the **assumptions are unmet**, and the **model's predictions may be unreliable**.

In Conclusion, GDP per Capita and Total Population affect GDP, but Total Population has a more substantial impact. However, the model does not fully follow the expected patterns, meaning its predictions may be reliable.

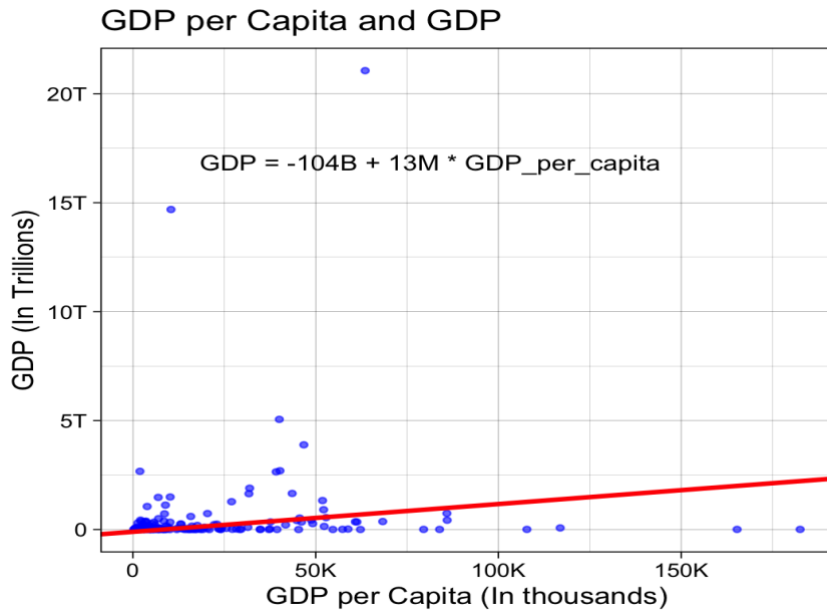
Table 1.1 Correlation Coefficients of the variables

	GDP	GDP per Capita	Total Population
GDP	1	0.14132025	0.57810213
GDP per Capita	0.1413203	1	-0.06622889
Total Population	0.5781021	-0.06622889	1

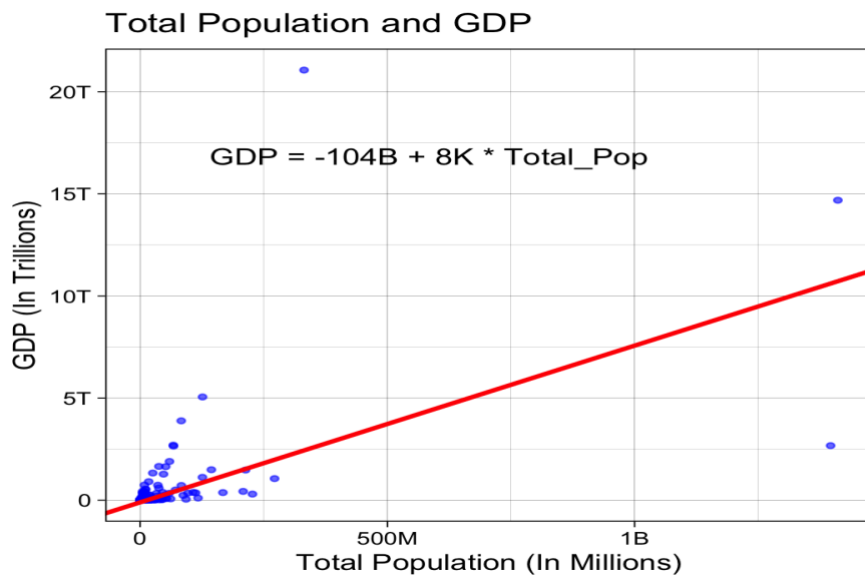
Table 1.2 Regression model summary

Metric	Estimate (Slope)	P-Value
Intercept	-1.042e+11 (-104.2 Billion)	0.41456
GDP per capita	1.275e+07 (12.75 Thousands)	0.00137
Total Pop	7.675e+03 (7.675 Million)	< 2e-16
F-Statistics	59.62	< 2.2e-16
Adjusted R square	0.3605	

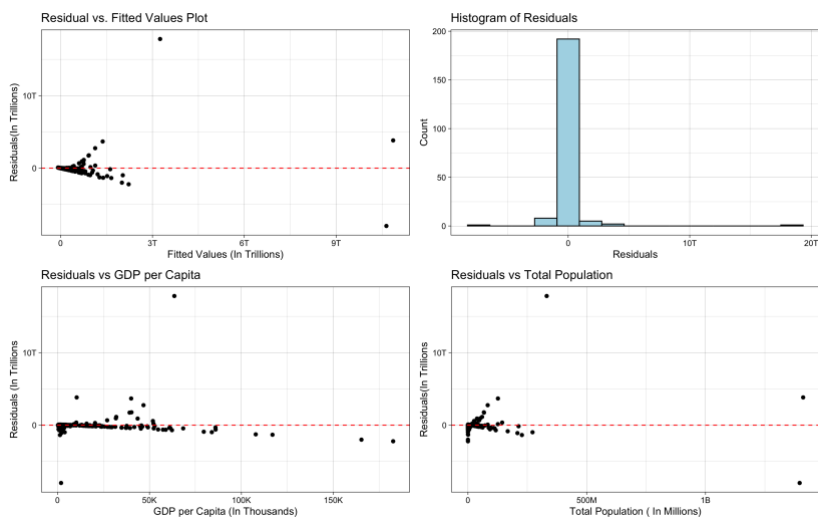
Graph 1.1 Scatter plot showing a linear relationship between GDP per capita (Thousands) and GDP (In Trillions)



Graph 1.2 Scatter plot showing a linear relationship between Total population (Millions) and GDP (Trillions)



Graph 1.3 Linear Regression Model Assumptions plot



Reference:

<https://chatgpt.com/#>