

Factors Affecting GDP per Capita in 2022

A DATA-DRIVEN APPROACH TO ECONOMIC INSIGHTS

INTRODUCTION

Economic growth is central to understanding global development, with GDP per capita widely used as a key measure (Source: IMF). This study examines the impact of factors on the GDP per capita across 50 countries in order to determine if they are significant contributors to GDP per capita and thereby are useful predictors of a country’s economic well-being.

RESEARCH QUESTIONS

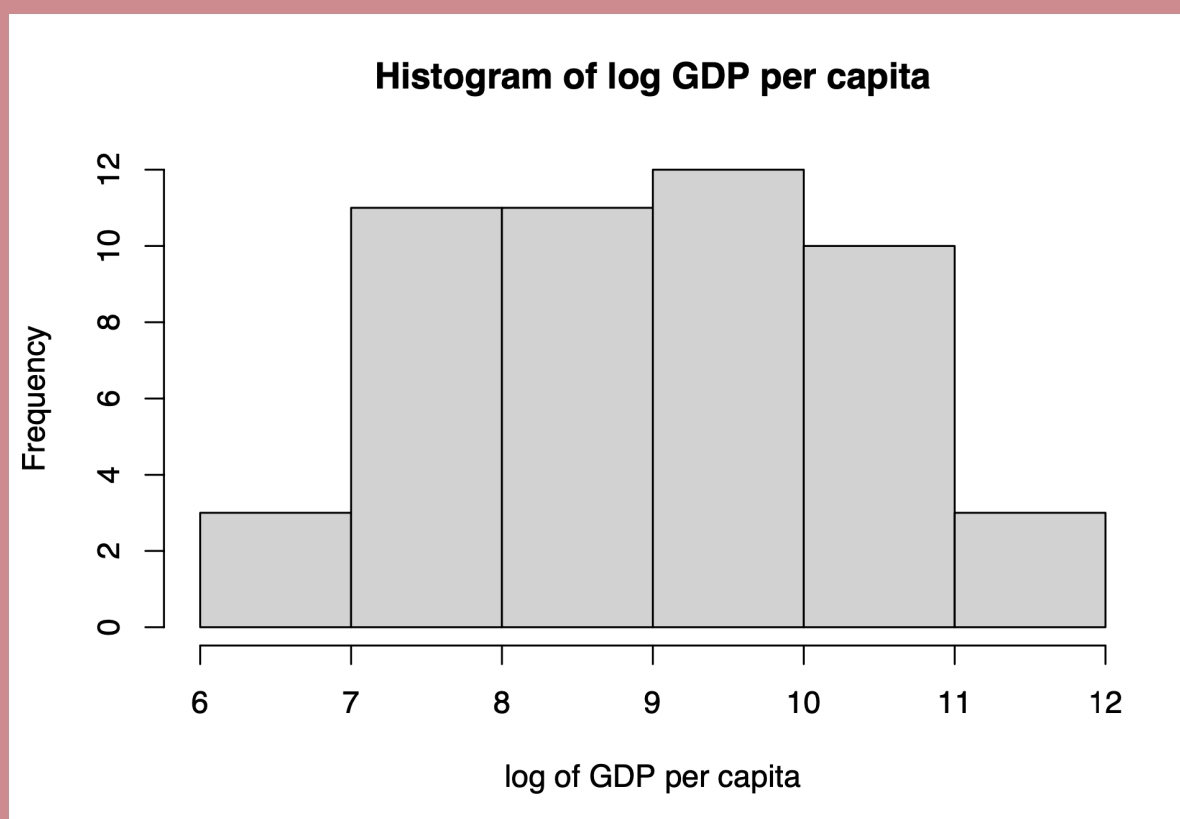
- 1. Do countries with a higher **percentage of population in the prime working age** have higher GDP per capita on average
- 2. Do countries with a higher **total fertility rate** have lower GDP per capita on average?
- 3. Do countries with a higher **unemployment rate** have a lower GDP per capita on average?

OUR DATA

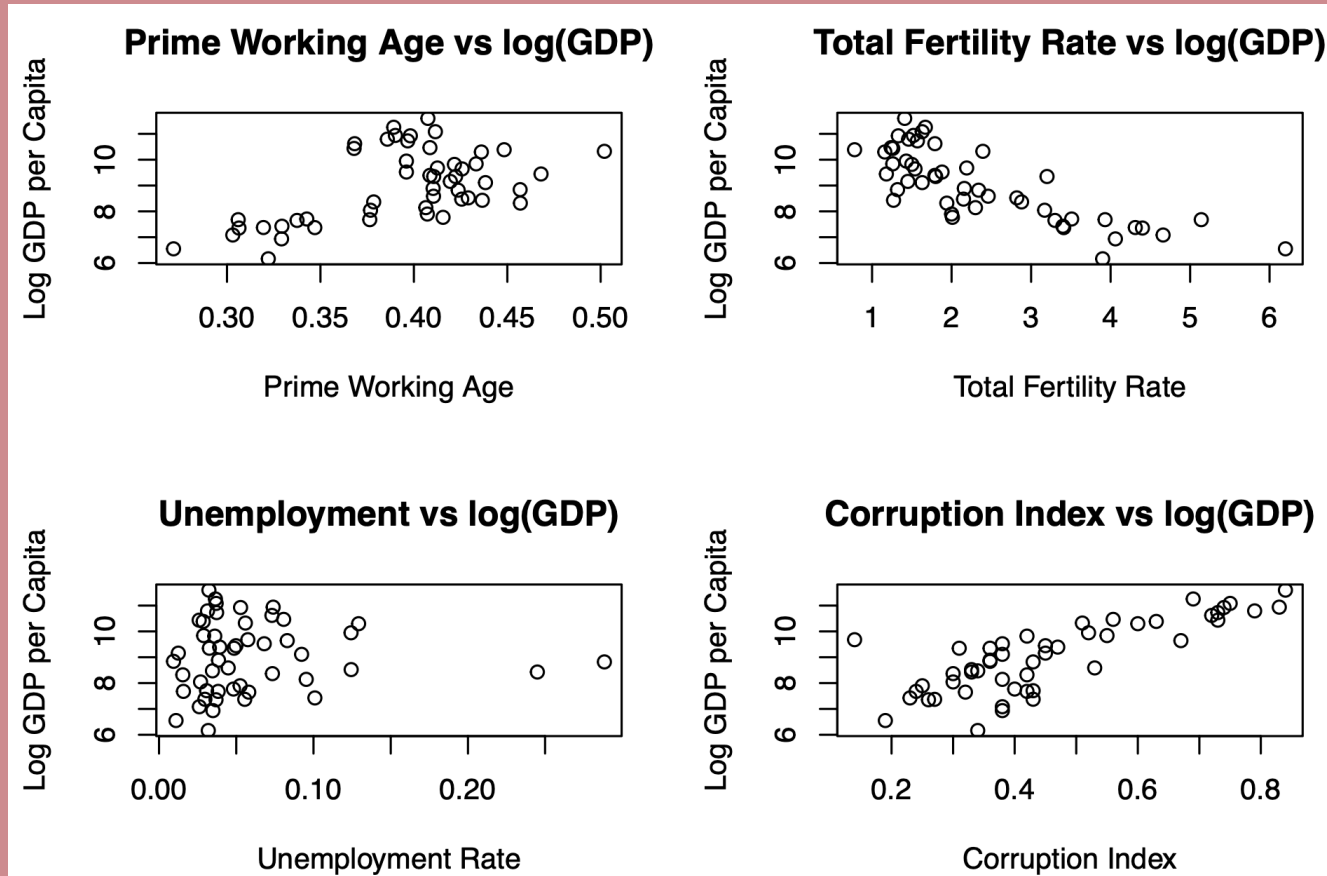
- Population of Interest:
 - SRS to choose **50 UN-Recognized Countries**
- Data:
 - Response Variable: GDP per Capita in 2022
 - Quantitative: Total Fertility Rate, Unemployment Rate, Gini Coefficient, Corruption Index, Prime Working Age %
 - Qualitative: Government Type, Development, Region

GRAPHICAL SUMMARIES

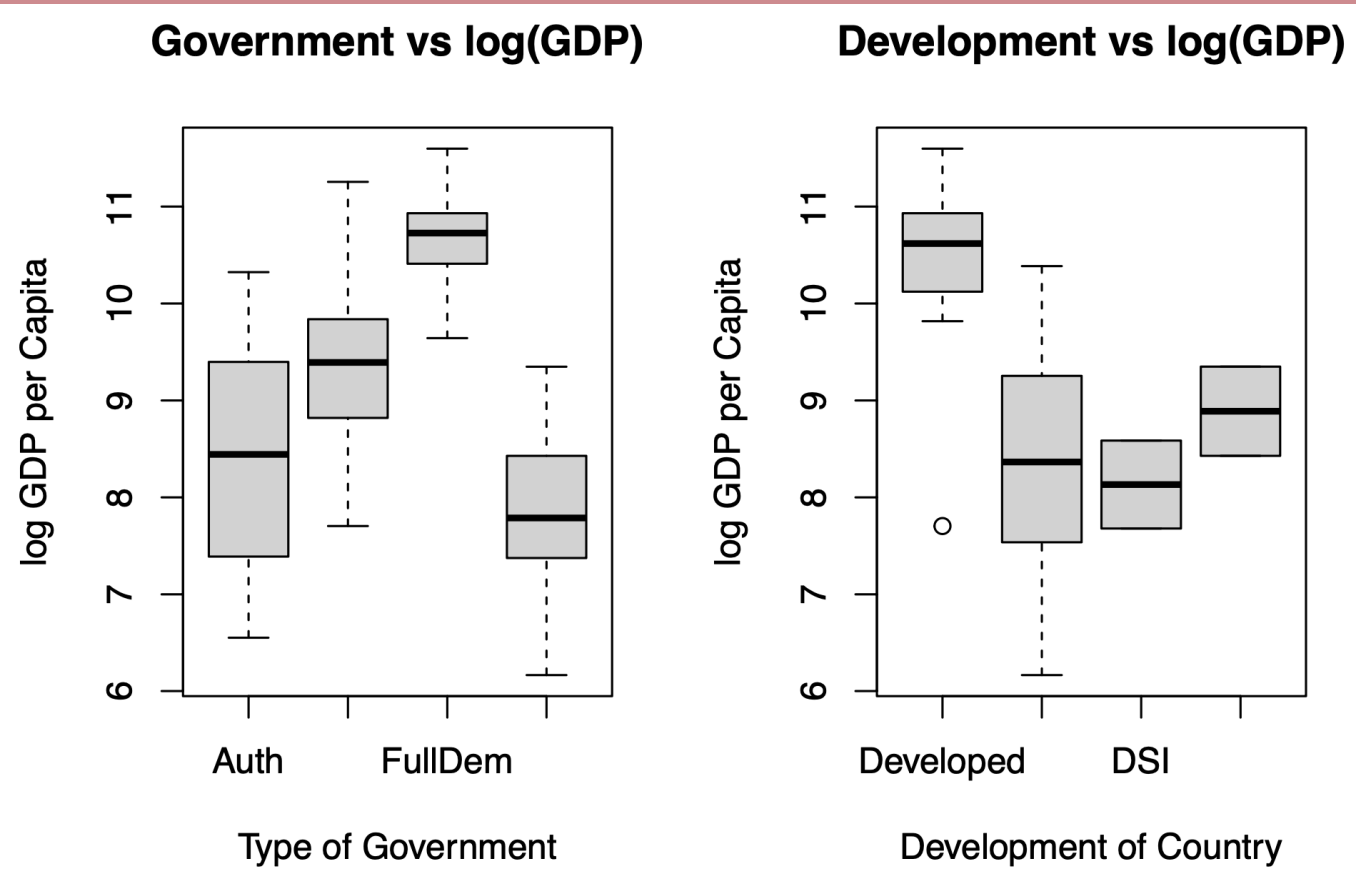
Histogram of the log of our response variable:



Scatter Plot of Quantitative Variables:



Box Plot of Qualitative Variables:



MODEL BUILDING

Stage One: Quantitative Variables

Initial: $GDP = \beta_0 + \beta_1(\text{Unemployment}) + \beta_2(\text{TFR}) + \beta_3(\text{Gini}) + \beta_4(\text{CI}) + \beta_5(\text{PWA}) + \beta_6(\text{Gini} \cdot \text{CI})$
Final: $GDP = \beta_0 + \beta_1(\text{TFR}) + \beta_2(\text{CI})$

Check for Multicollinearity:

- Correlation between TFR and CI (-0.5766) did not show concern of strong pairwise relationship
- Mean VIF = 1.498 < 3; all individual VIF < 10

Stage Two: Qualitative Variables

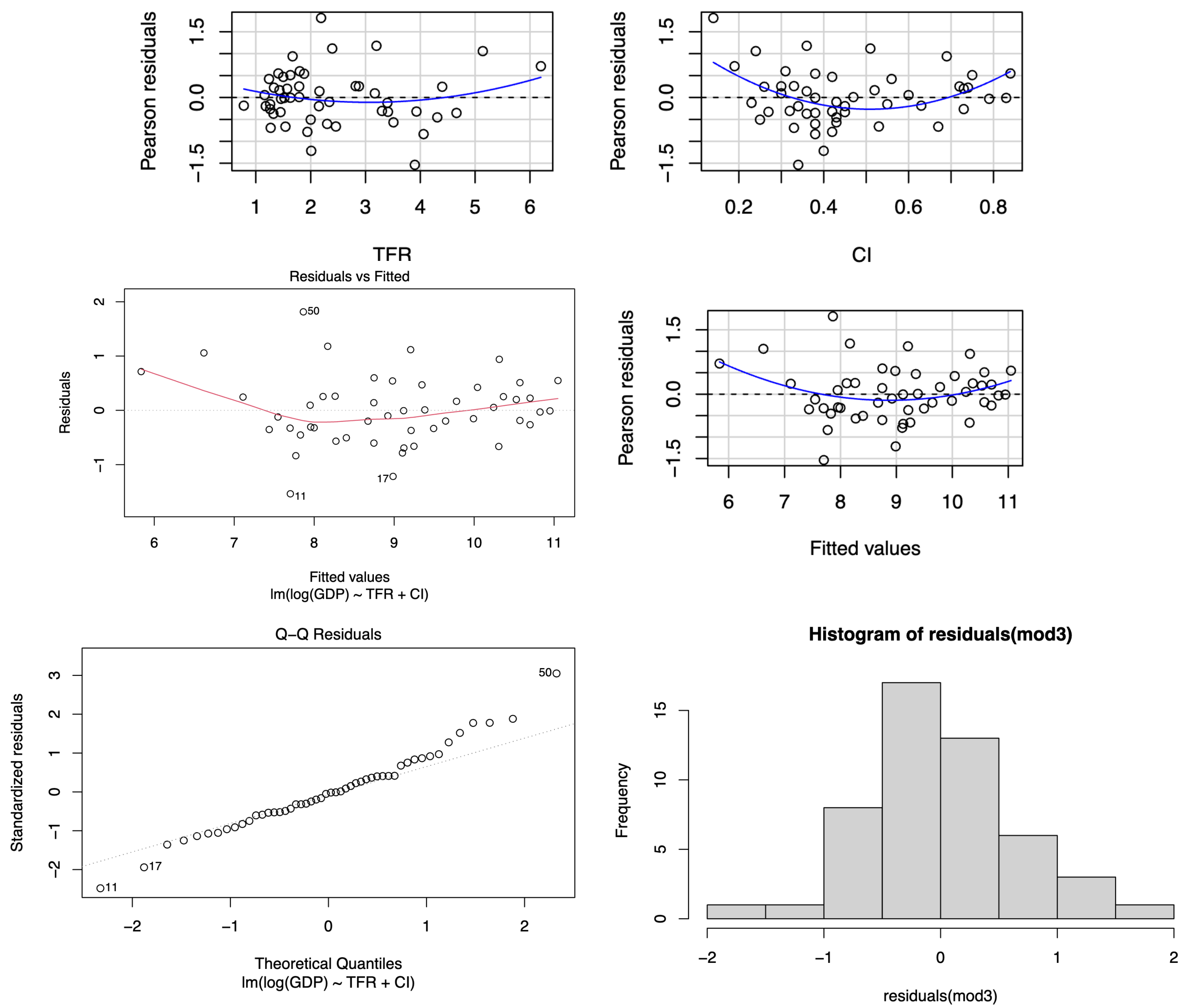
Initial: $GDP = \beta_0 + \beta_1(\text{TFR}) + \beta_2(\text{CI}) + \beta_3(\text{Govt}) + \beta_4(\text{Dev}) + \beta_5(\text{Region})$
Final: $\beta_0 + \beta_1(\text{TFR}) + \beta_2(\text{CI})$

Stage Three: Interactions

*no significant interactions found

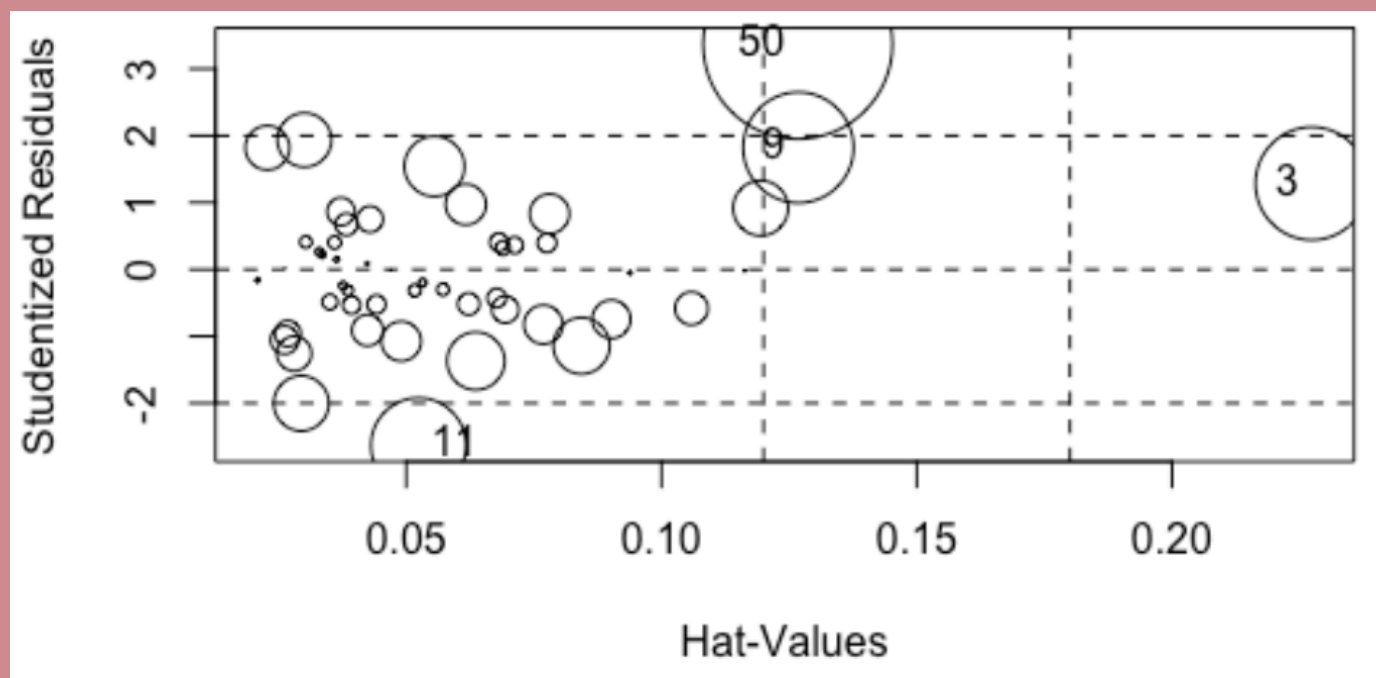
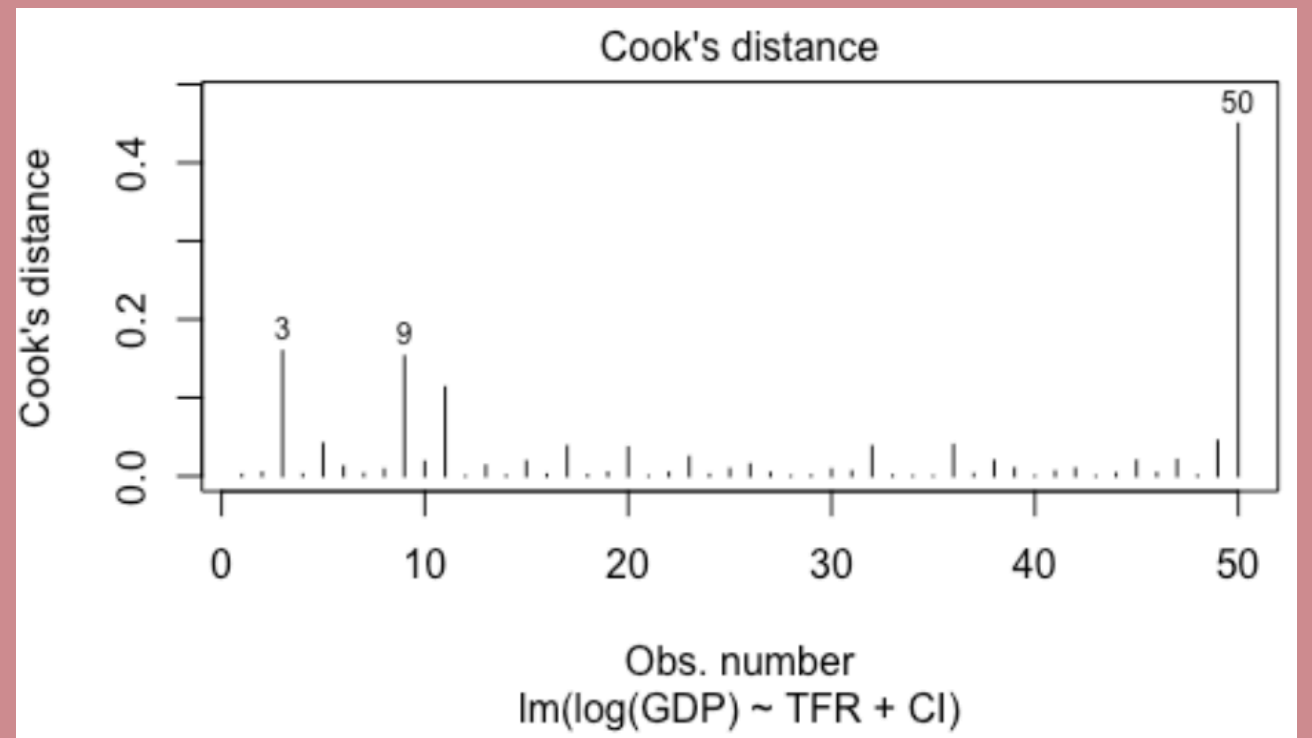
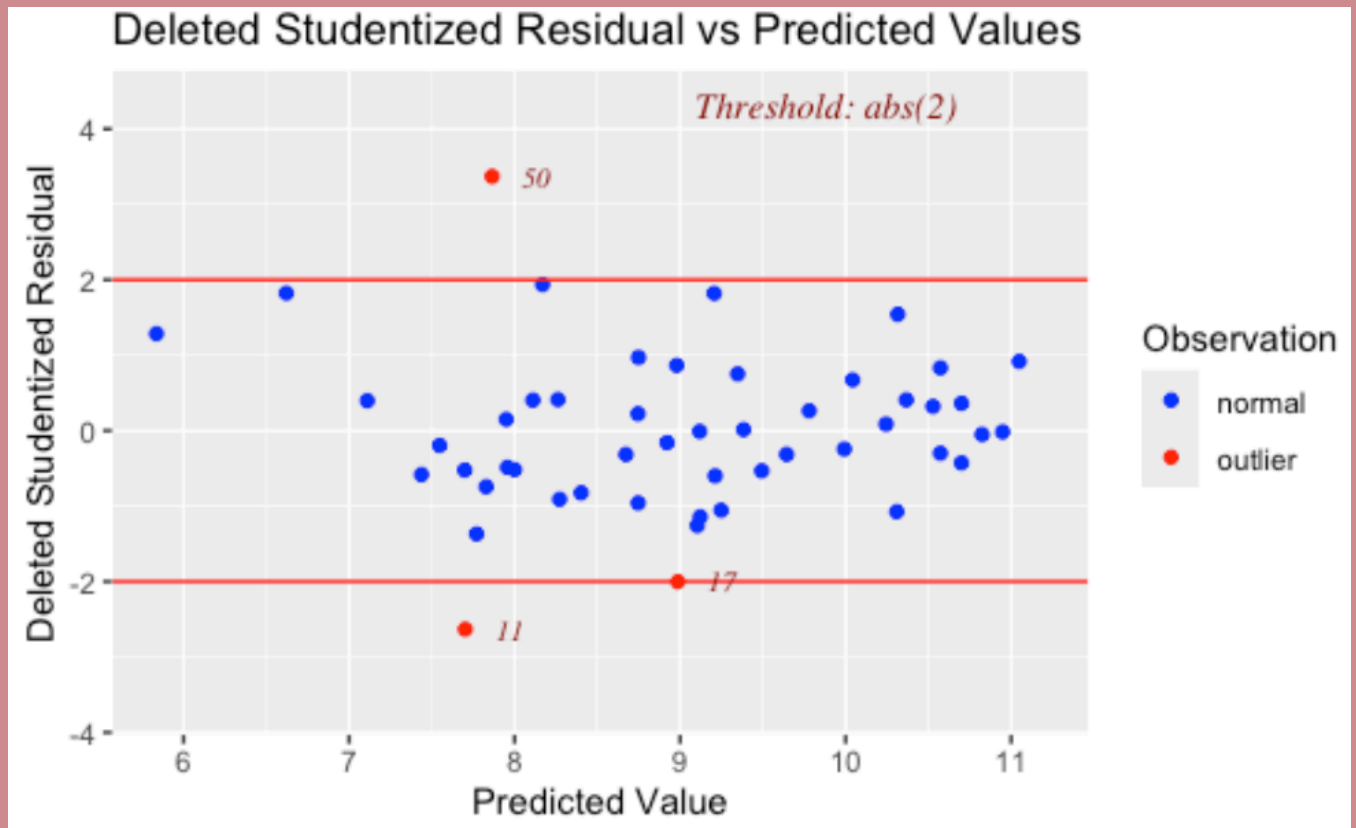
CHECKING MODEL ASSUMPTIONS

- Mean Zero: no clear quadratic trends (NOT VIOLATED!)
- Constant Variance: no “fanning out” pattern (NOT VIOLATED!)
- Normality: (NOT VIOLATED!)
 - QQ plot mostly follow a linear trend
 - histogram is normally distributed
- Independence: not a time series data (NOT VIOLATED!)



OUTLIERS/INFLUENTIAL POINTS

- Outliers through
 - Observations 3, 9, 11, 50
- Influential Points
 - Observations 11, 50
- Removed: 3, 9, 11, 50



FINAL PREDICTION MODEL

$$\text{hat}\{\log(\text{GDP})\} = 8.242 - 0.573 \cdot \text{TFR} + 4.524 \cdot \text{CI}$$

Residual Standard Error	0.5081 on 43 degrees of freedom	CONFIDENCE INTERVAL		
			2.5%	97.5%
Multiple R-Squared	0.8515	(Intercept)	7.429	9.054
Adjusted R-Squared	0.8446	TFR	-0.7535	-0.3918
P-Value	< 2.2e-16	CI	3.460	5.588

INTERPRETATIONS

The baseline expected log GDP per capita is 8.242, which is independent of the studied variables.

- A unit increase in the TFR will result in a 0.573 decrease in log(GDP), if all other variables are held constant.
- A unit increase in the CI will result in a 4.524 increase in log(GDP), if all other variables are held constant.

The model accounts for 84.446% of the variation in the data.

A p-value of less than an alpha value of 0.05, suggests the model is adequate at predicting GDP Per Capita.

At a 95% confidence level, while holding other explanatory variables constant

- A unit increase in TFR is associated with a decrease in log(GDP) ranging from 0.753 to 0.392.
- A unit increase in CI is associated with an increase in log(GDP) ranging from 3.460 to 5.588.

Future Research

In the future, it we could collect additional variables like education levels or infrastructure quality as they are also significant predictors of GDP. We could also refine the data collection methods to address missing data issues and expand the dataset to include more countries.

