**A close-up of a logo

Description automatically generated**

**PROBABILITY AND STATISTICS LAB**

**PROJECT REPORT**

**Analyzing the correlation Relationship Between Income Inequality and Economic Growth**

Group Members:

1. Dhanishka-22BAI1207
2. Shivam -22BAI1274
3. **Ayush-22BAI1275**
4. Karthikeya-22BAI1301
5. Khoushal-22BAI1315
6. Akash R-22BAI1336

**INTRODUCTION:**

The relationship between income inequality and economic growth is a longstanding and complex topic in economics. Researchers have debated whether high levels of income inequality hinder or promote economic growth, and what factors influence this relationship. Income inequality can affect economic growth in various ways. On one hand, people argue that some inequality incentivizes productivity, investment, and entrepreneurship, which can drive growth. On the other hand, high inequality may lead to reduced social cohesion, lower levels of health and education for disadvantaged populations, and increased political instability, all of which can negatively impact economic growth.

This relationship is often influenced by factors such as the stage of economic development, social policies, and regional economic conditions, making it difficult to establish a one-size-fits-all conclusion. This report explores the complex relationship between income inequality and economic growth and aims to provide a comprehensive analysis of these dynamics and their implications.

**ABSTRACT:**

This project presents an R-based analysis exploring the correlation between income inequality and economic growth. Using publicly available datasets, we conduct statistical analyses to investigate how income inequality impacts economic performance across various regions and time periods. The project utilizes data manipulation, visualization, and correlation techniques in R to identify patterns and examine potential causal relationships. Key findings are illustrated through visualizations, providing a clear view of how income disparity might influence growth dynamics. This analysis aims to inform policy decisions and contributes to a broader understanding of socio-economic factors impacting economic development. The project underscores the versatility of R for economic data analysis and statistical modeling in social science research.

**LITERARY REVIEW:**

The paper **"Usefulness of Correlation Analysis"** [1]by Samithamby Senthilnathan (SSRN Electronic Journal, July 2019) emphasizes the valuable role of correlation analysis in understanding relationships between variables within social science research. It highlights how correlation coefficients, such as Pearson’s r, provide insights into the strength and direction of associations, making it easier to identify patterns and connections within complex datasets.

Senthilnathan illustrates how correlation analysis can uncover significant relationships between socioeconomic factors, behavioral trends, and other data in a straightforward, interpretable manner. The study underscores correlation as a foundational tool that aids researchers in forming hypotheses and guiding further exploration within diverse research contexts.

The paper **"Income Inequality as a Determinant of Economic Growth: A Cross-Country Analysis"[2]** by Suwoto et al. investigates how income inequality influences economic growth across countries. Using the Gini coefficient to measure inequality, the authors analyze its impact on growth while controlling for factors like education, unemployment, and savings rates.

The findings suggest that the relationship between inequality and growth is complex. In some countries, higher inequality correlates with slower growth, possibly due to limited access to resources and reduced social mobility. In other cases, inequality may spur economic growth by concentrating wealth and incentivizing investment. These results highlight that the effect of income inequality on growth can vary significantly depending on a country’s specific economic and social context.

The paper **"The Relationship between Income Inequalities and Economic Growth: New Evidence"** [3] explores the complex link between income inequality and economic growth. Previous studies have shown mixed results, with some suggesting inequality harms growth by increasing social instability, while others argue it can spur innovation and investment. The authors review theories like the Kuznets Curve but note that the relationship is not always consistent. Their new evidence indicates that inequality's effect on growth depends on factors like institutional quality and economic development stage. The paper highlights the need for a nuanced understanding of this relationship.

In **"Inequality and Economic Growth,"** Joseph Stiglitz [4]critiques the view that inequality is necessary for economic growth. He argues that excessive inequality hampers growth by limiting access to opportunities, reducing social mobility, and fostering instability. Stiglitz reviews literature showing that inequality can stifle demand and undermine long-term economic progress. He advocates for policies such as progressive taxation and better access to education and healthcare, suggesting that reducing inequality can lead to more inclusive and sustainable growth. His work highlights the negative effects of inequality on economic stability and growth.

The paper "New Tools for Predicting Economic Growth Using Machine Learning"[5] focuses on using machine learning techniques to predict economic growth by integrating these tools into economic forecasting. The authors explore various algorithms, including regression models and decision trees, that allow for the identification of key predictors of economic performance, enhancing predictive power beyond traditional methods. They also emphasize the importance of cross-validation to ensure robustness in model predictions. This work contributes to the growing field of applying machine learning to economics by improving the accuracy and flexibility of growth predictions​

**DATASET DESCRIPTION:**

**Content**

This dataset comprises historical information encompassing various indicators concerning Inequality in Income on a global scale. The dataset prominently features: ISO3, Country, Continent, Hemisphere, Human Development Groups, UNDP Developing Regions, HDI Rank (2021), and Inequality in Income from 2010 to 2021.

**Dataset Glossary (Column-wise)**

* **ISO3** - ISO3 for the Country/Territory
* **Country** - Name of the Country/Territory
* **Continent** - Name of the Continent
* **Hemisphere** - Name of the Hemisphere
* **Human Development Groups** - Human Development Groups
* **UNDP Developing Regions** - UNDP Developing Regions
* **HDI Rank (2021)** - Human Development Index Rank for 2021
* **Inequality in Income from 2010 to 2021** - Inequality in Income from year 2010 to 2021

**Data Dictionary**

* **UNDP Developing Regions**:
  + **SSA** - Sub-Saharan Africa
  + **LAC** - Latin America and the Caribbean
  + **EAP** - East Asia and the Pacific
  + **AS** - Arab States
  + **ECA** - Europe and Central Asia
  + **SA** - South Asia

Data preprocessing:

**Data Cleaning and Preparation**

* **Filtering Missing Values:**
  + The initial steps involve filtering out rows with missing HDI (Human Development Index) ranks and inequality data.
  + This ensures that the subsequent analysis is performed on a clean dataset, which is crucial for accurate results**.**
* **Handling Missing Values in Inequality Columns:**
  + The code utilizes drop\_na() to exclude rows with any missing values in columns related to inequality.
  + For specific numeric columns, it uses gsub() and as.numeric() to convert string values (with commas) to numeric format. This is critical for accurate statistical calculations

**Analysis of the Model**

**Overview**

The analysis focuses on the relationship between Human Development Index (HDI) rankings and income inequality across various countries from 2010 to 2021. The data is segmented by continent and human development groups, providing insights into socio-economic disparities and trends over time.

**Key Components of the Analysis**

1. **Average HDI Rank and Average Inequality in Income by Continent**
   * The average HDI ranks and inequality in income are calculated for different continents, highlighting disparities in human development.
   * **Findings**:
     + Continents such as Europe and Oceania typically exhibit low average HDI ranks and lower average inequality in income compared to Africa, America and parts of Asia.
     + This suggests a correlation between geographical location and socio-economic development.

**2. Human Development Groups and Average Inequality**

* + The analysis categorizes countries into human development groups (high, medium, low, very high) and calculates average income inequality for each group.
  + **Findings**:
    - Very high human development groups show significantly lower average income inequality compared to low and medium groups.
    - This suggests that policies aimed at improving human development can effectively reduce income inequality

**3. Income Inequality Trends (2010-2021)**

* + A line plot visualizes income inequality trends for the top 5 and bottom 5 countries based on HDI rank.
  + **Findings**:
    - Countries like Norway and Switzerland (top HDI) show lower income inequality, while countries like Burundi and Chad (bottom HDI) exhibit higher inequality.
    - This trend indicates that higher HDI is often associated with lower income inequality.

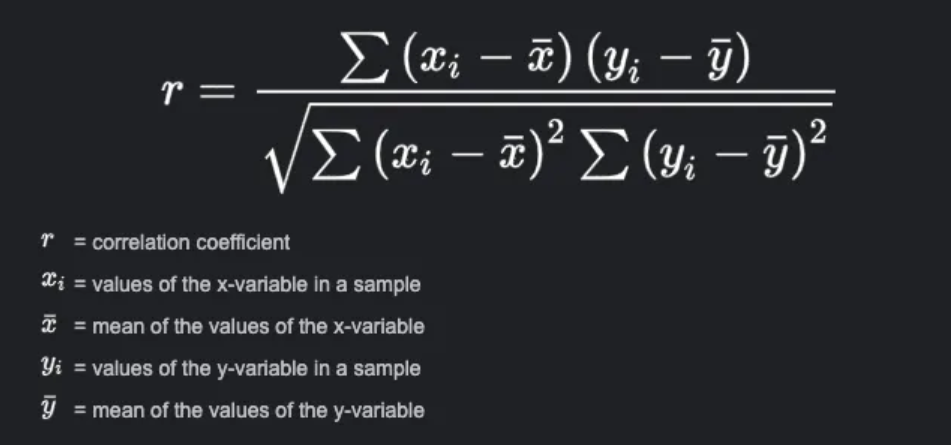
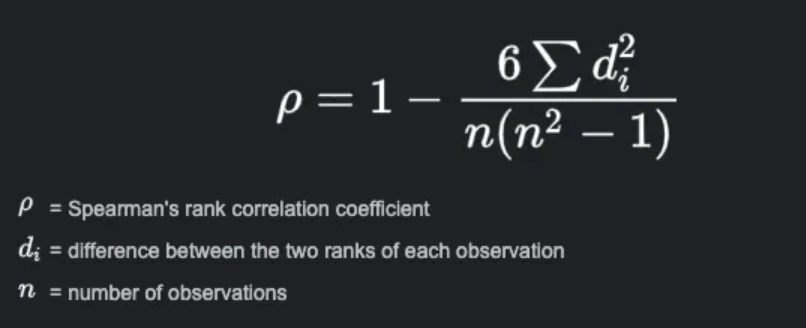
**4. Boxplot of Inequality in Income by Year**

* + Boxplots provide a visual summary of income inequality across different years, allowing for the identification of outliers and the distribution of income inequality.
  + **Findings**:
    - The boxplots reveal fluctuations in income inequality over the years, with some years showing significant outliers, particularly in lower HDI countries.

**5. Scatter Plots with Regression Line of HDI Rank vs. Income Inequality (2010-2021)**

* + Scatter plots illustrate the relationship between HDI rank in 2021 and income inequality for each year from 2010 to 2021.
  + The data points are scattered around the regression line. This suggests that the relationship might be more complex and influenced by other factors.
  + **Findings**:
    - As HDI rank increases, income inequality tends to increase.
    - This relationship is consistent across multiple years, reinforcing the idea that better human development correlates with more equitable income distribution.

**Statistical Models**

1. Descriptive Statistics Model
   * This model includes calculating averages, medians, and other summary statistics for HDI ranks across different continents and years. It helps in understanding overall trends and differences in human development.
2. Comparative Analysis Model
   * This model involves comparing income inequality metrics between the top 5 countries and the bottom 5 countries based on HDI rank. It highlights disparities in income inequality in relation to human development.
3. Boxplot Visualization Model
   * Utilizing boxplots to visualize the distribution of income inequality over different years. This model helps in identifying trends, medians, and outliers visually, providing insights into data distribution.
4. **Linear Regression Model**
   * A predictive model that explores the relationship between HDI rank and income inequality. It aims to understand how changes in HDI rank might predict variations in income inequality.
5. **Correlation Analysis Model**
   * This model assesses the strength and direction of the relationship between HDI rank and income inequality over the years. It is useful for identifying patterns and correlations in the data.
   * Pearson Correlation: 
   * Spearman Correlation: 

**Pearson Correlation:**

The Pearson correlation measures the linear relationship between the two variables (HDI Rank and Inequality in Income).Pearson correlation values range from 0 (no linear relationship) to +1 (perfect positive linear relationship) and -1 (perfect negative linear relationship).

**Spearman Correlation:**

The Spearman correlation measures the rank-order relationship between the two variables (HDI Rank and Inequality in Income). It is based on the relative ranking rather than actual values.Similar to Pearson, Spearman correlation values range from 0 (no monotonic relationship) to +1 (perfect positive monotonic relationship) and -1 (perfect negative monotonic relationship).

**OUTPUT AND CONCLUSION**

* + The analysis of HDI ranks and income inequality from 2010 to 2021 reveals significant disparities across continents, with higher HDI countries like Australia exhibiting lower income inequality compared to lower HDI countries (high HDI rank)) such as Burundi and Chad. Boxplot and Scatter plot illustrate fluctuations in income inequality over the years, indicating that increase in HDI rank causes income inequality to increase. Correlation analysis supports this trend, showing a moderate positive relationship between HDI rank and income inequality, suggesting that as HDI rank increases, inequality slightly increases. Overall, these findings emphasize the importance of enhancing human development to promote more equitable income distribution**.**
  + Spearman’s correlation is better for non-linear data because it assesses monotonic relationships based on ranks, making it robust for non-linear trends, non-uniform rates of change, and data with outliers. Pearson’s correlation is more suited for linear relationships and is sensitive to non-linear patterns and outliers, potentially giving misleading results if the relationship is non-linear.
  + **Spearman is more reliable here as our data is non-linear with many outliers**

**Overall Inference:**

* **Human development (measured by HDI Rank)** seems to be positively correlated with **income inequality**, and this relationship has become **stronger over the years**.
* While the correlation is moderate and positive, it is not perfect, which means that **other factors** might also be influencing income inequality, beyond just HDI Rank.

References:

[1] Senthilnathan, S. (2019). Usefulness of Correlation Analysis. SSRN Electronic Journal.

[2]Suwoto, T., & Zhai, Y. (2016). Income Inequality as a Determinant of Economic Growth: A Cross-Country Analysis. Georgia Institute of Technology

[3]Kowalska, M. (2021). The Relationship between Income Inequalities and Economic Growth: New Evidence. European Research Studies Journal.

[4] Stiglitz, J. E. Inequality and Economic Growth.

[5]Bang, J. T., Basuchoudhary, A., & Sen, T. New Tools for Predicting Economic Growth Using Machine Learning: A Guide for Theory and Policy. St. Ambrose University; Virginia Military Institute.