

SPEC REU R Resources: Applied Introduction to T-Tests, Correlation, and OLS regression – Group Work

Alix Ziff, Gaea Morales, Zachary Johnson, Claudia Salas Gimenez and Ben Graham

January 2025

In this groupwork, you will analyze real-world data to explore the relationship between GDP per capita and the Human Development Index (HDI), addressing the question: Does greater economic wealth lead to better social and human development outcomes? You will apply concepts from previous modules, including cleaning and reshaping data, creating visualizations, and performing regression analysis by fitting a linear model to interpret potential relationships between variables. Let's get started!

Initial Setup

Begin by setting up your working directory and loading the necessary libraries, along with the `HDI_GDP_pc.csv` dataset.

To better understand the HDI variables in the `HDI_GDP_pc.csv` data, refer to the [UNDP Human Development Index website](#), where you can explore the latest HDI dataset and its measures. Additionally, the GDP per capita data was sourced from the World Development Indicators. For more information, consult the [World Development Indicators data catalog](#). Please note that this dataset has been modified for this groupwork, so variable names and formats may differ slightly from those in the original UNDP and World Development Indicators datasets.

Does Economic Wealth Improve Human Development?

A key question in human development research is whether greater economic wealth leads to better social and human outcomes. Wealthier countries are generally expected to provide better education, healthcare, and overall quality of life, resulting in higher HDI scores. To analyze this relationship, we will use GDP per capita as the independent variable (a measure of economic wealth) and HDI as the dependent variable.

Exercise 1: Clean and Reshape the Data

Before analyzing the data, we need to reshape it into a country-year format and ensure that GDP per capita values are stored as numeric. When importing datasets, numeric values may sometimes be misread as text, which can cause errors during analysis. We will use the `as.numeric()` function to convert these columns before reshaping the data.

Exercise 2: Visualize Relationship Between HDI and GDP Per Capita

Now that the data is properly formatted, let's create a scatterplot to visualize the relationship between GDP per capita and HDI. Challenge yourself to customize your visualization to produce a publication-quality figure.

Exercise 3: Draw an Estimated Line of Best Fit

Next, let's add a line of best fit to your scatterplot. What does this line suggest about the relationship between GDP per capita and HDI?

Exercise 4: Interpret the Regression Results

Finally, to test this relationship, let's fit an Ordinary Least Squares (OLS) regression model. Then, interpret the summary statistics:

- Identify the slope of HDI and interpret it.
- Is the relationship between HDI rank and mean infant mortality significant? How do you know?