

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

Claudia Salas Gimenez and Ben Graham

January 2025

Welcome to the final assignment of the M7 Regression I module! In this homework, you'll apply your skills in data management and basic regression analysis to explore a research question of your choice. Unlike previous assignments, this task involves tackling an open-ended question, encouraging you to think critically about causal relationships and strengthen your analytical skills while experiencing the full process of conducting independent research—from identifying a question and cleaning data to analyzing and interpreting results.

Save your responses in your personal subfolder in the 412_413 shared Google Drive folder. The R script should be titled `HW_DM3_[YOUR_INITIALS]`. You can also save a copy of your R script to your own computer for future reference.

Let's dive in!

Overview

The goal of this assignment is to identify a causal relationship between two variables, find an appropriate dataset, run an OLS regression to analyze the relationship, and interpret your findings. This open-ended exercise allows you to explore a topic that interests you.

You can approach this assignment in one of two ways:

1. You can identify a causal relationship between two variables you want to analyze (e.g., “Does higher education expenditure increase literacy rates?”), then find data to test your hypothesis.
2. Use an existing dataset to formulate a research question based on the variables it includes.

1. Choose a Research Question

Your research question should specify how one variable is expected to influence another. Be specific and consider the expected direction of the relationship (e.g., “As education spending increases, literacy rates are expected to improve.”).

2. Locate and Prepare a Dataset

Once you've selected your research question, locate a dataset that includes the necessary variables. Prepare the data for analysis using data management techniques you've learned in earlier modules.

While you're welcome to reuse datasets from earlier assignments, challenge yourself to explore new data sources. Here are some great resources:

- The [World Development Indicators](#) is a comprehensive database with a wide range of global data on topics such as health, education, and economic development. It's a great starting point for exploring global trends and relationships between socioeconomic indicators, and we've use it in previous assignment, so you might be familiarize with how to clean the data.

- The [Correlates of War](#) provides historical data on interstate and intrastate conflicts, alliances, and territorial changes. These datasets are a rich resource for exploring research questions in international relations.
- Another option is to use replication data. Replication datasets are an essential part of transparent and reproducible research as they are collections of data that accompany published research papers, allowing others to verify, replicate, or build upon the authors' findings. The [Harvard Dataverse](#) is an open-access digital repository designed to store and share research datasets, with one of its key features being the availability of replication datasets. Using replication data not only helps you understand research projects and their variables but also highlights opportunities for further investigation.

3. Analyze the Data Interpret Your Findings

Run an OLS regression to analyze the relationship between your variables, and interpret the results. When interpreting your findings, consider the following:

- What does the slope of the independent variable tells you about the relationship?
- How much variation in the dependent variable is explained by your model?
- Is the relationship statistically significant? How do you know?

By completing this assignment, you'll gain basic hands-on experience in designing, conducting, and interpreting independent research. This process will not only strengthen your data analysis skills but also deepen your understanding of how to apply statistical tools to real-world questions.