

STATA code explanation

Author: Monica Cueto Tapia

Position: Impact Evaluation Consultant - IDB Invest

The Stata script provided plays a crucial role in preparing and refining datasets associated with GreenPyme projects. Initially, it initializes the dataset by removing existing data and imports the Excel file "Greenpyme_db_2018.xlsm" from a specified location, saving it as "Greenpyme_db_0.dta." This process involves dropping unnecessary variables and filtering projects based on approval status. Moreover, it standardizes variables by applying predefined labels, converting yes/no questions into 1s and 0s, and transforming string variables into numeric ones, ensuring data consistency. Additionally, the script standardizes sector names, introduces a new variable for miscellaneous sectors, and performs replacements and manipulations based on specific conditions within the dataset. Overall, the script streamlines the preprocessing and cleaning of the GreenPyme project dataset, covering data importation, variable manipulation, standardization, and categorization to prepare for subsequent analysis.

The provided Stata code undertakes various data transformations and calculations to optimize dataset preparation. It addresses non-numeric values by replacing them with missing values and converting selected columns into numeric format. The code conducts value replacements based on predefined conditions, generates new variables, and calculates conversion factors for energy consumption. Furthermore, it handles missing values and inconsistencies, computes potential economic savings, and adjusts variables to ensure proper formatting for further analysis. This code is an integral part of the data cleaning and analysis workflow, focusing on variable manipulation, new variable generation, and maintaining data integrity, which is particularly valuable for projects involving energy consumption and cost optimization. Moreover, the Stata script encompasses functionalities such as generating descriptive tables for GreenPyme projects, conducting propensity score matching (PSM) analysis, and modeling adoption probabilities. It systematically executes steps for data processing, analysis, and modeling, providing insights into treatment effects and control group balance in the analyzed dataset, thereby offering a comprehensive approach to understanding the GreenPyme project's impact and outcomes.