IMPACT OF EDUCATION LEVEL AND INCOME ON HOUSEHOLD SAVINGS: AN ECONOMETRIC APPROACH USING LINEAR REGRESSION

Objective:

The aim of this project was to examine how household income and education level influence annual savings using a multiple linear regression model. The goal was to produce a statistically validated report that could assist policy advisors, banks, and social researchers in understanding household saving behavior.

Client Requirements:

- Use real or publicly available economic data
- Perform a robust multiple linear regression using income and education level as key predictors
- Control for potential confounders (e.g., age, employment status)
- Test assumptions such as multicollinearity and heteroskedasticity
- Deliver a clear, concise report with regression tables and practical interpretation
- Format the report in APA style, with proper citations and visual support

Data Source:

Data was sourced from a national financial survey publicly available through the U.S. Bureau of Labor Statistics. It contained over 10,000 records with fields such as:

- Household annual income
- Years of formal education
- Age of head of household
- Marital status
- Employment status
- Annual household savings

Methodology:

1. **Data Cleaning and Preparation:**
    - Missing values were handled using pairwise deletion.
    - Outliers in income and savings were flagged and tested for influence using Cook's distance.
    - Education was treated as a continuous variable (years of education).
2. **Model Specification:**

$$\text{Savings}_i = \beta_0 + \beta_1 \cdot \text{Income}_i + \beta_2 \cdot \text{Education}_i + \beta_3 \cdot \text{Age}_i + \beta_4 \cdot \text{EmploymentStatus}_i + \epsilon_i$$

3. **Software Used:**
    - Analysis was performed using **Stata 17**.
    - Outputs were exported and refined in **MS Word** using **APA formatting**.
    - Visualizations were generated using **ggplot2 in R** for enhanced clarity.
4. **Diagnostics and Assumption Testing:**
    - **Multicollinearity:** Checked using VIF values (all < 2.5)
    - **Homoscedasticity:** Breusch-Pagan test showed no evidence of heteroskedasticity
    - **Normality of residuals:** Histogram and Q-Q plots confirmed approximate normality
    - **Linearity:** Partial residual plots supported linear relationships between predictors and outcome

Key Findings:

- **Income** had a positive and statistically significant effect on savings (**$p < 0.001$**), with each additional \$1,000 in annual income associated with an average \$145 increase in household savings.
- **Education** was also positively associated with savings (**$p < 0.05$**), indicating that individuals with higher educational attainment tend to save more.
- **Age** and **employment status** were significant control variables, confirming their impact on savings behavior.
- The **adjusted $R^2 = 0.48$**, indicating that nearly half the variation in savings could be explained by the model variables.

Deliverables:

- **Econometric Report:** 14-page APA-formatted report with regression tables, interpretations, and policy implications.
- **Charts:** 3 regression visualizations (scatterplot with fit line, residuals vs fitted, and coefficient plot).
- **Presentation Slides (Optional):** Condensed 5-slide summary for stakeholders.
- **Code File:** Annotated Stata .do file and R visualization script.

Impact & Client Feedback:

The client, a private economic consultancy working with a regional bank, reported that the analysis helped them redesign a savings incentive program targeting low-income households with limited formal education. The findings also informed their marketing strategy for savings accounts, aligning financial literacy with demographic targeting.