A/ ASSIGNMENT RECAP

- Write **a 3500-word report** constructing a dataset to analyze determinants of **income inequality** for a specific year, including the **impact of globalization**, by reviewing **the literature**, cleaning and analyzing data, estimating **econometric models**, and **proposing policy recommendations**.

Suggested Structure:

I. **Part 1: Introduction & Literature Review**
  A. **Introduction (Suggested 150 words)**
  B. **Literature review and research questions (Suggested 450 words)**
II. **Part 2: Data cleaning and data description**
  A. **Summary Statistics (Suggested 200 words)**
  B. **Data Visualization(Suggested 150 words)**
  C. **Missing values and outliers(Suggested 150 words)**
  D. **Variable transformation(Suggested 150 words)**
III. **Part 3: Model Specification & Analysis**
  A. **Model Specification**
      1. **Dependent variable (Suggested 250 words)**
      2. **Independent variable (Suggested around 150 words)**
      3. **Population Model (Suggested around 150 words)**
  B. **Part 3.2 Estimation & Interpretation**
      1. **Results and Coefficients of Model 1 (Suggested around 150 words**
      2. **Hypothesis and T-test of Model 1(Suggested around 150 words)**
      3. **Potential multicollinearity in Model 1 (Suggested around 150 words)**
      4. **Heteroskedasticity in Model 1 (Suggested around 150 words)**
      5. **New model proposal (Model 2) (Suggested around 150 words)**
      6. **Extended Model (Model 3) (Suggested around 150 words)**
      7. **Potential endogenous variables in Model 1 (Suggested 100 words)**
IV. **Part 4: Conclusion and Policy Implications**
  A. **Findings Summary (Suggested around 150)**
  B. **Policy Recommendations (Suggested 150 words)**
  C. **Limitations and Suggestion for Improvements (Suggested 150 words)**
V. **Part 5: Industrial Talk. (Suggested 250 words)**

B. KEYWORD EXPLANATIONS

1. Regression Analysis

Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables by estimating a regression model, typically represented as an equation.

2. Multicollinearity

Multicollinearity occurs when two or more independent variables in a regression model are highly correlated, making it difficult to isolate the individual effects of each variable.

3. Heteroskedasticity

Heteroskedasticity is a statistical term describing the presence of non-constant variance of errors in a regression model, which can lead to biased standard errors and affect the reliability of statistical tests.

4. Coefficient

The estimated parameter values from a regression model quantify the effect of each independent variable on the dependent variable.

5. Log Form and Quadratic Form

Log form involves transforming a variable by taking its natural logarithm. Quadratic form involves squaring a variable. These transformations are used to change the functional relationship between variables in regression analysis

6. Goodness-of-fit

Goodness-of-fit - Statistics like R-squared indicate how well a regression model fits and explains the variation in the dependent variable based on the predictors.

7. Dependent Variable and Independent Variable:

The dependent variable is the outcome or response variable in a statistical analysis, representing what you aim to explain or predict. Independent variables are the predictors or factors used to explain or predict the changes in the dependent variable.

8. Regression Function Form:

The regression function form is the equation that represents the relationship between the dependent and independent variables in a regression analysis.

9. Significance level

The probability threshold used to determine statistical significance, most commonly 0.01, 0.05 or 0.10.

10. Binary Variable

A binary variable, also known as a dummy variable, takes on only two values, typically 0 and 1, to represent the presence or absence of a particular characteristic.

11. Endogenous variables

An endogenous variable, in the context of econometrics and statistical modeling, is a variable within a regression model that is determined by or dependent on other variables within the same model. In simpler terms, an endogenous variable is influenced by the factors or variables being studied within the model itself.

12. Regression

A statistical method to study the relationship between variables, like how education affects income.

13. p-value

The probability that a result occurred randomly. Small p-values indicate statistical significance.

14. R-squared

A statistic that indicates how well the model fits the data. Higher values mean more variance is explained.