A. ASSIGNMENT RECAP

- Write **a 3000-word report** analyze determinants of the **Human Development Index** (HDI) for a sample of countries, including reviewing **academic literature**, **running multiple regressions**, **interpreting results**, **testing hypotheses**, and proposing **policy recommendations**

Suggested Structure:

I. **Part 1: Overview and Data Description**
   A. **Overview of Topic (Suggested 150 words)**
   B. **Data description (Suggested 250 words)**
II. **Part 2: Initial estimation**
   A. **Linear Regression Model (Suggested 250 words)**
   B. **Model Estimation Using OLS (Suggested 250 words)**
III. **Part 3: Interpretation**
   A. **Interpret R-Squared (Suggested 250 words)**
   B. **F-test Interpretation (Suggested 250 words)**
   C. **T-test Interpretation (Suggested 250 words)**
   D. **Expectation & Actual Results Comparison (Suggested 250 words)**
   E. **Models Comparison (Suggested 250 words)**
IV. **Part 4: Further Estimation**
   A. **Dummy variables (Suggested 250 words)**
   B. **Interaction term (Suggested 250 words)**
   C. **Alternate model Estimation (Suggested 250 words)**
V. **Part 5: Conclusion**
   A. **Findings Summary (Suggested 150 words)**
   B. **Policies Proposal (Suggested 250 words)**

B. KEYWORD EXPLANATIONS

1. **Regression**

A statistical method used to estimate the relationship between a dependent variable and one or more independent variables based on observed data.

2. **Single Linear Regression**

A regression model with one independent variable used to estimate its linear effect on a continuous dependent variable. It takes the form:

$Y = \beta0 + \beta1X + \varepsilon$

Where Y is the dependent variable, X is the single independent variable, β0 is the intercept, β1 is the slope coefficient on X, and ε is the error term.

3. **Multiple Linear Regression**

A regression model with two or more independent variables used to estimate their linear effects on a continuous dependent variable. It takes the form:

$Y = \beta0 + \beta1X1 + \beta2X2 + ... + \betanXn + \varepsilon$

Where Y is the dependent variable, X1 to Xn are the multiple independent variables, β0 is the intercept, β1 to βn are the slope coefficients, and ε is the error term.

4. **Coefficient**

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

5. **Statistical significance**

A measure indicating whether a regression coefficient or test result is unlikely to have occurred by chance, determined by the p-value and significance level chosen.

6. **Goodness-of-fit**

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

7. **Hypothesis testing**

The use of sample data to determine whether to reject a hypothesis about a population parameter at a specified significance level based on statistical evidence.

8. **Mean**

A measure of central tendency calculated as the sum of all values divided by the number of values in a sample or population distribution.

9. Standard Regression Format

The conventional structure for presenting regression results including coefficient estimates, standard errors, and diagnostic statistics.

10. Descriptive statistics

A statement of no statistical significance or effect that is tested and either supported or rejected based on evidence from a sample.

11. Ordinary Least Squares (OLS)

A common method for estimating the coefficients in a linear regression model by minimizing the sum of squared residuals.

12. Adjusted R-squared

A modified version of R-squared that accounts for the number of predictors in the model. Used to assess goodness-of-fit.

13. F-test

A statistical test used to determine if the regression model as a whole has a statistically significant relationship with the dependent variable.

14. Significance level

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

15. Dummy Variable

A binary categorical variable coded as 1 or 0 used to represent a qualitative characteristic

16. Interaction term

A variable created by multiplying two predictors to estimate their combined effect and test moderation effects.

17. T-test

A statistical test used to determine if a regression coefficient is significantly different from zero based on its t-statistic.