Assignment 4

Objective: Learn and apply statistical analysis techniques using Python, focusing on descriptive and inferential statistics, and the derivation of Ordinary Least Squares (OLS) models.

Dataset: Utilize the 'General Social Survey' (GSS) dataset, which contains data on social attitudes and behaviors collected over several decades. This dataset is relevant for social science and is accessible from the National Opinion Research Center (NORC) at the University of Chicago (GSS Data Explorer).

Tasks:

1. Load and Explore the Dataset:
    - Download the GSS dataset and import it into a Python environment using Pandas.
    - Perform initial data exploration – display the first few rows, understand the variables, and check for missing values.
2. Descriptive Statistics:
    - Calculate basic descriptive statistics (mean, median, standard deviation, etc.) for key variables.
    - Create simple visualizations (histograms, box plots) to understand the distribution of selected variables.
3. Inferential Statistics:
    - Perform a hypothesis test (e.g., t-test) to compare two groups within the dataset on a variable of interest.
    - Calculate confidence intervals for key statistics.
4. Correlation Analysis:
    - Analyze the correlation between two or more variables (e.g., income and education level).
    - Represent these relationships through scatter plots or correlation matrices.
5. Regression Analysis:
    - Derive an OLS regression model to examine the relationship between a dependent variable and one or more independent variables.
    - Interpret the coefficients of the model, and discuss the fit and predictive capabilities.
6. Model Diagnostics:
    - Perform diagnostic tests to check the validity of the OLS model assumptions (linearity, homoscedasticity, normality, etc.).
    - Address any issues found in the diagnostics through transformations or adjustments to the model.

Submission: A comprehensive report (Jupyter Notebook) including Python code, outputs, visualizations, and write down one long paragraph narrative explaining different steps of the analysis, findings, and conclusions.