MA155 – Statistical Modeling (Report)

Name: _____ **Score:** ____ pts.

This project relates to statistical modeling with focus on using regression techniques to model and predict health expenditure in Africa.

Background and Purpose

There are many elements that influence both population health and health expenditures, such as income level, pollution related to the level of industrialization, environmental quality, etc. Offering quality health care services should be one of the most essential objectives of governments because they can lead to improved life expectancy, social and economic welfare, etc. The deterioration of environmental quality globally has a substantial effect on what we call "healthy living". Therefore, this project is purposed on the development of regression models to predict Health Expenditure taking into consideration the effect of environmental degradation in Africa using 2019 World Bank data (www.data.worldbank.org). This project assumes current health expenditure per capita (CHEPC) measured in international dollars at purchasing power parity as a proxy for the health variable (i.e., response variable) and carbon dioxide emissions (CO2) measured in metric tons per capita as a proxy for the environmental variable (i.e., predictor variable).

Your Task (Use Microsoft Excel throughout):

1. (xx points) Compute the mean, median, standard deviation, coefficient of variation, minimum, and maximum values of CHEPC and CO2 (Round all values to two decimal places).

Complete the table below!

Table 1: Mean, Median, Standard Deviation, Coefficient of Variation, Minimum, and Maximum of CHEPC and CO2

Measure	CHEPC (in International $)	CO2 (in Metric Tons per Capita)
Mean		
Median		
Standard Deviation		
Coefficient of Variation		
Minimum		
Maximum		

2. (xx points) Draw a scatter diagram showing the relationship between CHEPC (response/dependent variable) and CO2 (explanatory/independent variable). From the scatter diagram, briefly describe the relationship between CHEPC and CO2.

Place your answer here!

3. (x points) Compute the linear correlation coefficient (round the coefficient to 3 decimal places) and use it to describe the strength and direction of the relationship between CHEPC and CO2.

Place your answer here!

4. (xx points) Construct a scatter diagram showing the fitted linear regression line (line of best fit), fitted (estimated) model, and coefficient of determination.

Place your answer here!

5. (x points) From (4), write down the equation of the line of best fit and the coefficient of determination. Interpret the coefficient.

Place your answer here!

6. (10 points) Use the fitted linear regression model to determine the expected CHEPC (round the value to two decimal places) when CO2 is 3.850 (i.e., predict the value of CHEPC when CO2 is 3.850). Interpret the results.

Place your answer here!

7. (10 points) Assume the observed CHEPC for a given CO2 of 4.01 is 750.45. Calculate the residual of this observation. Briefly comment on the residual value. **Hint:** A residual is the difference between an observed value and a predicted value in regression analysis.

Place your answer here!