Life Expectancy and Socio-economic Factors Modeling Using "Life Expectancy (WHO)" Dataset

This report outlines the process of simulating and modeling the relationship between life expectancy and socio-economic factors using the "Life Expectancy-WHO" dataset from Kaggle. The approach involves importing, cleaning, analyzing, and validating the dataset, and then running simulations to predict outcomes.

Step 1: Importing and Cleaning the Dataset

Importing the Dataset

1. Download the Dataset: Obtain the dataset from Kaggle via the [Life Expectancy-WHO Dataset](https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who).

2. Load into SPSS: Once downloaded, load the dataset into SPSS through `File -> Open -> Data`. The dataset is typically in CSV or Excel format.

Data Cleaning

- Missing Value Check: Use descriptive statistics to identify missing data.

- Handling Missing Values:

- Listwise Deletion: Remove cases with missing data if they are minimal.

- Imputation: For significant missing data, impute values using the mean or median for numerical variables.

- Anomaly Detection: Identify outliers or erroneous data, such as negative life expectancy values.

- Standardization: Ensure all categorical variables are consistently coded.

Step 2: Multivariate Linear Regression Analysis

Design of the Analysis

- Dependent Variable: Life expectancy

- Independent Variables: GDP, adult mortality, immunization rates, and other socio-economic factors.

Running the Regression

1. Steps in SPSS: Go to `Analyze -> Regression -> Linear`. Set life expectancy as the dependent variable and socio-economic factors as independent variables.

2. Assumption Checks:

- Linearity of Residuals

- Normality of Residuals

- Homoscedasticity of Residuals

Interpretation of Results

- Coefficients: Interpret the coefficients to understand the relationship between each socio-economic factor and life expectancy.

- Statistical Significance: Use p-values to determine which factors are statistically significant.

Step 3: Simulating Socio-Economic Changes

Running Simulations

1. Scenario Development: Create scenarios by varying factors, such as:

- A 10% increase in GDP

- A 5% decrease in adult mortality rate

- A 15% rise in immunization rates

2. Prediction: Using the regression model, predict the impact on life expectancy for each scenario.

Discussion of Impact

- Economic Factors: Analyze how improvements in GDP and immunization rates can lead to increased life expectancy.

- Public Health Policy: Discuss how reductions in adult mortality might influence public health policy.

Step 4: Model Validation and Sensitivity Analysis

Model Validation

- Compare Predicted vs Actual Values: Use descriptive statistics to compare the predicted and actual life expectancy.

- Accuracy Measures: Calculate metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) to assess prediction accuracy.

Sensitivity Analysis

- Impact of Factors: Conduct a sensitivity analysis to examine how changes in one socio-economic factor, while keeping others constant, affect life expectancy.

Outcomes and Implications

- Key Findings: Highlight the socio-economic factors that most significantly affect life expectancy, such as GDP and healthcare investments.

- Public Health Implications: Discuss how the results emphasize the importance of investing in health and economic development to improve life expectancy and public health outcomes.