**1. Importing the Dataset into SPSS and Cleaning the Data**

* **Imported the Dataset**:
  + Download the dataset from Kaggle: Life Expectancy Dataset.
  + Open SPSS, go to File > Open > Data, and select the CSV file.
* **Cleaned the Data**:
  + Identified and handle missing values by using Transform > Replace Missing Values or excluding them.
  + Check for inconsistencies or outliers using Analyze > Descriptive Statistics > Explore.
  + After cleaning, saved the dataset by going to File > Save As.

**2. Multiple Linear Regression Analysis**

* **Performed the Regression**:
  + Clicked to Analyze > Regression > Linear.
  + Set Life Expectancy as the dependent variable.
  + Choosedsocio-economic factors like GDP, Adult Mortality, and Immunization Rates as independent variables.
  + Clicked on Statistics to include necessary statistics like R² and Coefficients, then clicked OK to run the regression.

**Simulation of Socio-Economic Factors**

* **Simulate Changes**:
  + Used Data > Compute Variable to create new variables reflecting changes in GDP, Adult Mortality, etc.
  + Did re-run the regression analysis with these new variables to observe changes in Life Expectancy.

#### 3. **Generate New Variables for Simulation**

* **Compute New Variables**:
  + Go to Transform > Compute Variable.
  + For example, to simulate a 10% increase in GDP:
    - Name the target variable (e.g., GDP\_Sim).
    - Enter the formula: GDP \* 1.10 (for a 10% increase).
  + Repeat this process for other factors you wish to simulate.
* **Example Scenarios**:
  + **Scenario 1**: Increase GDP by 10%:
    - New variable: GDP\_Sim = GDP \* 1.10.
  + **Scenario 2**: Decrease Adult Mortality by 5%:
    - New variable: Adult\_Mortality\_Sim = Adult Mortality \* 0.95.
  + **Scenario 3**: Increase Immunization Rates by 20% for polio ,Measles,HepatitisB,Diphtheria,BMI for example
    - New variable: polio\_sim = polio \* 1.20.
    - For all the immunization cases.

#### 4. **Run the Regression Analysis with Simulated Data**

* **Perform Regression**:
  + Went to Analyze > Regression > Linear.
  + Usedthe simulated variables (e.g., GDP\_Sim, Adult\_Mortality\_Sim) as independent variables.
  + Set Life Expectancy as the dependent variable.
* **Interpret the Results**:
  + Compared the regression coefficients from the simulated data with those from the original data.
  + Analyzed how the predicted life expectancy changes under each scenario.

**Comparing the Descriptive Statistics**

**Original Dataset Descriptive Statistics**

* **Life Expectancy**: Mean = 70.057, Std. Dev = 8.5803
* **GDP**: Mean = 6752.40, Std. Dev = 12746.68
* **Adult Mortality**: Mean = 158.82, Std. Dev = 120.65
* **Hepatitis B**: Mean = 80.78, Std. Dev = 25.03
* **Measles**: Mean = 1976.00, Std. Dev = 9454.01
* **Polio**: Mean = 84.60, Std. Dev = 21.93
* **Diphtheria**: Mean = 85.01, Std. Dev = 21.29

**Simulated Dataset Descriptive Statistics**

* **Life Expectancy**: Mean = 70.057, Std. Dev = 8.5803 (same as the original)
* **GDP\_sim**: Mean = 7427.64, Std. Dev = 14021.35 (higher than the original, indicating more economic variability)
* **AdultMorality\_sim**: Mean = 150.88, Std. Dev = 114.62 (lower mean and lower standard deviation compared to the original)
* **Polio\_sim**: Mean = 101.52, Std. Dev = 26.32 (higher mean and more variability compared to the original)
* **Measles\_sim**: Mean = 2371.20, Std. Dev = 11344.81 (higher mean and standard deviation, indicating higher counts and more variability in the simulated dataset)
* **HepatitisB\_sim**: Mean = 76.74, Std. Dev = 23.77 (slightly lower mean and less variability)
* **Diphtheria\_sim**: Mean = 102.01, Std. Dev = 25.55 (higher mean and more variability)

**Key Insights from Comparison:**

1. **Life Expectancy** remains the same between the original and simulated datasets, meaning it was not affected by the simulation.
2. **GDP** is higher and more variable in the simulated dataset, showing increased economic differences.
3. **Adult Mortality** shows a decrease in both the mean and variability in the simulated data, which might reflect better healthcare conditions in the simulation.
4. **Polio** and **Measles** show higher mean values and greater variability in the simulated dataset, indicating potentially more extreme simulated health outcomes.
5. **Hepatitis B** has a slightly lower mean and less variability in the simulated dataset.
6. **Diphtheria** shows a higher mean and more variability, suggesting more significant simulated outbreaks or variance in vaccination coverage.

### ****Key Factors and Observations: for the correlations****

#### **1. GDP (Gross Domestic Product):**

* **Original Correlation:** r=0.440r = 0.440r=0.440, showing a positive moderate correlation.
* **Simulated Correlation (GDP\_sim):** r=0.440r = 0.440r=0.440, remains the same.
* **Interpretation:** Both before and after the simulation, an increase in GDP positively correlates with life expectancy, suggesting that higher economic wealth generally improves life expectancy.

#### **2. Adult Mortality:**

* **Original Correlation:** r=−0.714r = -0.714r=−0.714, a strong negative correlation with life expectancy.
* **Simulated Correlation (AdultMortality\_sim):** r=−0.714r = -0.714r=−0.714, unchanged after simulation.
* **Interpretation:** Adult mortality has a strong negative impact on life expectancy, meaning that higher mortality rates significantly reduce life expectancy. The simulation shows no change in this relationship.

#### **3. Polio Immunization (Polio):**

* **Original Correlation:** r=0.361r = 0.361r=0.361, positive moderate correlation.
* **Simulated Correlation (Polio\_sim):** r=0.361r = 0.361r=0.361, remains unchanged.
* **Interpretation:** Polio immunization positively impacts life expectancy, with a moderate influence. This remains consistent after simulation.

#### **4. Measles:**

* **Original Correlation:** r=−0.083r = -0.083r=−0.083, showing a weak negative correlation.
* **Simulated Correlation (Measles\_sim):** r=−0.083r = -0.083r=−0.083, unchanged.
* **Interpretation:** Measles shows a weak negative correlation with life expectancy, indicating that higher measles cases might slightly reduce life expectancy. The simulation does not affect this correlation.

#### **5. Hepatitis B:**

* **Original Correlation:** r=0.240r = 0.240r=0.240, a weak positive correlation.
* **Simulated Correlation (HepatitisB\_sim):** r=0.240r = 0.240r=0.240, no change.
* **Interpretation:** Hepatitis B immunization positively correlates with life expectancy, though the effect is weak. This relationship remains stable post-simulation.

#### **6. Diphtheria:**

* **Original Correlation:** r=0.360r = 0.360r=0.360, moderate positive correlation.
* **Simulated Correlation (Diphtheria\_sim):** r=0.360r = 0.360r=0.360, remains unchanged.
* **Interpretation:** Diphtheria immunization has a moderately positive effect on life expectancy. The correlation remains the same after simulation.

The simulation validates these relationships and can guide future public health policies aimed at improving life expectancy by targeting these socio-economic variables.

In the two **Model Summary** tables you shared, we can see the same results for both the original socio-economic factors and the simulated ones. Let’s break down the comparison:

**Key Metrics:**

1. **R (Correlation Coefficient):**
   * **Both Models:** R=0.784R = 0.784R=0.784
   * **Interpretation:** In both cases (original and simulated), there is a strong correlation between the independent variables (socio-economic factors) and the dependent variable (likely life expectancy), with an RRR value of 0.784.
2. **R Square (Coefficient of Determination):**
   * **Both Models:** R2=0.615R^2 = 0.615R2=0.615
   * **Interpretation:** In both models, approximately 61.5% of the variation in the dependent variable (life expectancy) can be explained by the independent variables (GDP, AdultMortality, Polio, Measles, HepatitisB, and Diphtheria). This suggests a good fit for the model in both the original and simulated cases.
3. **Adjusted R Square:**
   * **Both Models:** Adjusted R2=0.614\text{Adjusted } R^2 = 0.614Adjusted R2=0.614
   * **Interpretation:** The adjusted R2R^2R2, which accounts for the number of predictors, is nearly identical to R2R^2R2, indicating that the inclusion of the independent variables is appropriate and improves the model's accuracy.
4. **Standard Error of the Estimate:**
   * **Both Models:** SE=5.3311\text{SE} = 5.3311SE=5.3311
   * **Interpretation:** The standard error, which gives an idea of the accuracy of predictions, is the same for both the original and simulated models, indicating that the simulation did not affect the accuracy of the estimates
5. **Change Statistics:**
   * **Both Models:** The R2R^2R2 Change, FFF Change, dfdfdf, and Significance values are the same:
     + R2R^2R2 Change = 0.615
     + FFF Change = 533.260
     + df1=6df\_1 = 6df1​=6, df2=2002df\_2 = 2002df2​=2002
     + Significance ppp-value = 0.000 (indicating that the model is statistically significant).
   * **Interpretation:** The statistical significance and fit of the model are identical for both the original and simulated data, showing no change in model performance after the simulation.

**Comparison Summary:**

* **No Change in Model Performance:** Both the original and simulated models have identical statistics, indicating that the simulation did not affect the relationship between the independent socio-economic factors (GDP, AdultMortality, Polio, etc.) and the dependent variable (life expectancy).
* **Good Model Fit:** The high RRR and R2R^2R2 values, along with a statistically significant FFF-change, indicate that both models are good at predicting the outcome.

This further validates the consistency between the original and simulated data, as both models produce identical results.

### ****Summary of Changes After Simulation:****

From the data, it is clear that after simulating the socio-economic factors, the correlations remain **unchanged** across all factors. This suggests that the simulation process, in this case, did not alter the fundamental relationships between life expectancy and socio-economic factors like GDP, adult mortality, and immunization rates.

### ****Policy Implications:****

* **Increase in GDP**: The positive correlation indicates that improving economic conditions, especially GDP, would likely enhance life expectancy.
* **Decrease in Adult Mortality**: Reducing adult mortality remains a significant factor in increasing life expectancy.
* **Immunization Programs**: Strong correlations with diseases like polio and diphtheria indicate the importance of maintaining and improving immunization coverage to improve public health outcomes.