**Question Two**

**1. Model Overview**

* **Dependent Variable (Y)**: Life expectancy at birth (in years) — the outcome or variable you are trying to predict.
* **Independent Variable (X)**: Adult mortality rate (per 1000 adults aged 15-60) — the predictor variable used to estimate life expectancy.

**2. ANOVA Table Breakdown**

| **Source** | **Sum of Squares** | **df** | **Mean Square** | **F** | **Sig.** |
| --- | --- | --- | --- | --- | --- |
| Regression | 128740.795 | 1 | 128740.795 | 2754.630 | .000 |
| Residual | 136749.965 | 2926 | 46.736 |  |  |
| Total | 265490.760 | 2927 |  |  |  |

**Regression:**

* **Sum of Squares (SS Regression)**: 128,740.795 represents the portion of the total variance in life expectancy that is explained by the adult mortality rate.
* **Degrees of Freedom (df)**: 1 (since there is only one predictor variable).
* **Mean Square**: This is the Sum of Squares divided by its degrees of freedom, which gives 128,740.795. It reflects the average variance explained by the predictor (adult mortality rate).
* **F-value**: 2754.630 — This is the ratio of the Mean Square of the Regression to the Mean Square of the Residuals. A high F-value indicates that the model explains a significant portion of the variance in life expectancy.
* **Sig. (p-value)**: 0.000 — This is the probability of observing such an extreme F-value if the null hypothesis (that adult mortality has no effect on life expectancy) were true. Since the p-value is less than 0.05, it suggests the predictor (adult mortality rate) has a statistically significant impact on life expectancy.

**Residual:**

* **Sum of Squares (SS Residual)**: 136,749.965 represents the variance in life expectancy not explained by the model (i.e., the unexplained variance).
* **Degrees of Freedom (df Residual)**: 2926 — This is the number of observations minus the number of parameters being estimated (n - 1 for one predictor).

**Total:**

* **Sum of Squares (SS Total)**: 265,490.760 is the total variance in life expectancy. This is the sum of the explained (regression) and unexplained (residual) variance.

**3. Interpretation**

* The **F-statistic** is very large (2754.630), and the **p-value (Sig.)** is extremely small (0.000). This indicates that the regression model is highly significant. In other words, the adult mortality rate (the predictor) significantly affects life expectancy at birth.
* The **Sum of Squares Regression (128,740.795)** is large relative to the **Residual Sum of Squares (136,749.965)**, meaning that a substantial portion of the variability in life expectancy is explained by the adult mortality rate.

**4. Conclusion**

The regression model indicates that adult mortality rate (per 1000 adults aged 15-60) has a significant and strong effect on life expectancy at birth. The model explains a large portion of the variance in life expectancy (as shown by the large regression sum of squares and high F-value), and the relationship between the variables is statistically significant (p < 0.05).

**Model Summary:**

* **Constant (Intercept)**: 78.018
  + This means that when the adult mortality rate is 0 (which is an unlikely but theoretical scenario), the life expectancy at birth would be approximately 78.018 years.

**Independent Variable:**

* **Adult Mortality Rate (per 1000 adults aged 15-60)**:
  + **B (Unstandardized Coefficient)**: -0.053
    - This suggests that for every unit increase in the adult mortality rate, life expectancy at birth decreases by 0.053 years.
  + **Standard Error**: 0.001
    - This is the measure of how much the estimate of the slope varies between samples.
  + **Beta (Standardized Coefficient)**: -0.696
    - This value shows the strength and direction of the relationship between adult mortality rate and life expectancy, with -0.696 indicating a strong negative correlation.
  + **t-statistic**: -52.485
    - This statistic tests whether the coefficient is significantly different from 0. A large negative t-statistic indicates that the relationship is highly significant.
  + **p-value (Sig.)**: 0.000
    - The p-value is less than 0.05, meaning the relationship between adult mortality rate and life expectancy is statistically significant.
  + **Tolerance & VIF**:
    - **Tolerance**: 1.000
    - **VIF (Variance Inflation Factor)**: 1.000
      * These values indicate there is no multicollinearity issue with the independent variable.

**Interpretation:**

* There is a **strong, negative relationship** between adult mortality rate and life expectancy at birth. As adult mortality increases, life expectancy decreases. The results are statistically significant, with a very low p-value (0.000), meaning we can confidently say that adult mortality rate is a significant predictor of life expectancy

**Model Summary Explanation:**

1. **R (Correlation Coefficient)**:
   * **R = 0.696**
   * This represents the strength and direction of the linear relationship between the independent variable (adult mortality rate) and the dependent variable (life expectancy). A value of 0.696 indicates a **moderately strong positive correlation** between the two.
2. **R-Square (Coefficient of Determination)**:
   * **R² = 0.485**
   * This indicates that **48.5%** of the variation in **life expectancy** can be explained by the **adult mortality rate**. In other words, nearly half of the variance in life expectancy is explained by changes in adult mortality rate, which is fairly substantial for a single-variable regression model.
3. **Adjusted R-Square**:
   * **Adjusted R² = 0.485**
   * This value adjusts the R-Square to account for the number of predictors in the model. Since this is a simple linear regression with one predictor, the R² and Adjusted R² are the same. The adjusted R² indicates the proportion of variance explained by the model, adjusted for the number of predictors.
4. **Standard Error of the Estimate**:
   * **Std. Error = 6.8364**
   * This is the standard deviation of the residuals (the prediction errors). It gives an idea of how far the predicted life expectancy values are from the actual values on average. A lower value indicates better model fit, and in this case, the typical error in predicting life expectancy is around 6.836 years.
5. **Durbin-Watson Statistic**:
   * **Durbin-Watson = 0.762**
   * This statistic tests for autocorrelation in the residuals of the regression model. Values close to 2 indicate no autocorrelation, values below 2 indicate positive autocorrelation, and values above 2 indicate negative autocorrelation.
   * **0.762** suggests some positive autocorrelation in the residuals, which means that consecutive error terms may be positively correlated, and this might affect the model’s assumptions of independence.

**Interpretation:**

* The model shows that adult mortality rate explains a significant portion (48.5%) of the variation in life expectancy. The relationship is moderately strong, and while the model is significant, there is evidence of autocorrelation (based on the Durbin-Watson value), which might suggest that there is room for improvement or additional predictors could be needed.

Iv). i explain the comparison of Hepatitis B, Infant Deaths, and Alcohol Consumption in relation to life expectancy, and validate the model, here's a structured approach i can follow:

**1.Model Validation: Predicted vs. Actual Life Expectancy**

After running the multiple linear regression in SPSS, the model provides predicted life expectancy values based on the independent variables: Hepatitis B, Infant Deaths, and Alcohol Consumption. To validate the model:

First, create predicted values.  
  
For every observation, SPSS will provide you with a predicted life expectancy following the completion of the regression analysis. This is predicated on the regression equation, which include the alcohol, infant mortality, and hepatitis B factors.

**Step 2: Evaluate Expected vs. Real**  
  
Compare these estimated values to the dataset's actual values for life expectancy.  
A scatterplot can be used to depict this comparison:  
Plot the Actual Life Expectancy (y-axis) and the Predicted Life Expectancy (x-axis) in SPSS.  
The points should ideally cluster around a 45-degree line, which represents the point at which anticipated and actual values agree.

**Sensitivity Analysis: Impact of Socioeconomic Factors**

1. **Scenario Analysis:** Create different scenarios by varying the values of the independent variables (e.g., increasing GDP, decreasing infant mortality, etc.).
2. **Predict Life Expectancy:** Use the trained model to predict life expectancy for each scenario.
3. **Analyse Changes:** Compare the predicted life expectancy values across different scenarios to understand how changes in socioeconomic factors affect the outcome.

**Comparing the Socioeconomic Factors**

Now, let’s compare how each of the factors influences life expectancy:

* **Hepatitis B**: A positive coefficient means that increasing vaccination rates has a beneficial effect on life expectancy. For instance, increasing vaccination by 10% could result in a relatively significant increase in life expectancy.
* **Infant Deaths**: A negative coefficient implies that reducing infant deaths would improve life expectancy. In a country with high infant mortality, policies aimed at reducing these deaths would have a meaningful impact on life expectancy.
* **Alcohol Consumption**: A negative coefficient means that as alcohol consumption increases, life expectancy decreases. Policies aimed at reducing alcohol consumption could result in improvements in life expectancy, though the effect may be smaller compared to other factors like infant mortality.

**Conclusion**

By comparing **Hepatitis B**, **Infant Deaths**, and **Alcohol Consumption**:

* **Infant Deaths** may have the largest immediate effect on life expectancy, as even small reductions in infant mortality can significantly increase life expectancy in countries with high infant death rates.
* **Hepatitis B vaccination** also has a strong positive impact, as improving vaccination rates can lead to better overall health outcomes and longer life expectancies.
* **Alcohol Consumption** plays a role, but its effect may be more gradual or less direct compared to the other factors.

**Model Summary Interpretation:**

* **R-squared (0.193)**: This indicates that 19.3% of the variability in life expectancy is explained by **Hepatitis B vaccination rates**, **Infant deaths**, and **Alcohol consumption**. While this is a modest fit, it still provides insights into how these factors impact life expectancy.
* **Coefficients**:
  + **Hepatitis B (0.064)**: A 1% increase in Hepatitis B vaccination coverage is associated with a 0.064-year increase in life expectancy.
  + **Infant deaths (-0.0092)**: For every additional infant death per 1,000 live births, life expectancy decreases by 0.0092 years.
  + **Alcohol (0.714)**: A 1-liter increase in alcohol consumption is associated with a 0.714-year increase in life expectancy (counterintuitive, potentially due to interaction with other factors or data issues).

II). i explain the comparison of Hepatitis B, Infant Deaths, and Alcohol Consumption in relation to life expectancy, and validate the model, here's a structured approach i can follow:

1.Model Validation: Predicted vs. Actual Life Expectancy

After running the multiple linear regression in SPSS, the model provides predicted life expectancy values based on the independent variables: Hepatitis B, Infant Deaths, and Alcohol Consumption. To validate the model:

First, create predicted values.

For every observation, SPSS will provide you with a predicted life expectancy following the completion of the regression analysis. This is predicated on the regression equation, which include the alcohol, infant mortality, and hepatitis B factors.

Step 2: Evaluate Expected vs. Real

Compare these estimated values to the dataset's actual values for life expectancy.

A scatterplot can be used to depict this comparison:

Plot the Actual Life Expectancy (y-axis) and the Predicted Life Expectancy (x-axis) in SPSS.

The points should ideally cluster around a 45-degree line, which represents the point at which anticipated and actual values agree.

Sensitivity Analysis: Impact of Socioeconomic Factors

1. Scenario Analysis: Create different scenarios by varying the values of the independent variables (e.g., increasing GDP, decreasing infant mortality, etc.).

2. Predict Life Expectancy: Use the trained model to predict life expectancy for each scenario.

3. Analyse Changes: Compare the predicted life expectancy values across different scenarios to understand how changes in socioeconomic factors affect the outcome.

Comparing the Socioeconomic Factors

Now, let’s compare how each of the factors influences life expectancy:

• Hepatitis B: A positive coefficient means that increasing vaccination rates has a beneficial effect on life expectancy. For instance, increasing vaccination by 10% could result in a relatively significant increase in life expectancy.

• Infant Deaths: A negative coefficient implies that reducing infant deaths would improve life expectancy. In a country with high infant mortality, policies aimed at reducing these deaths would have a meaningful impact on life expectancy.

• Alcohol Consumption: A negative coefficient means that as alcohol consumption increases, life expectancy decreases. Policies aimed at reducing alcohol consumption could result in improvements in life expectancy, though the effect may be smaller compared to other factors like infant mortality.

Conclusion

By comparing Hepatitis B, Infant Deaths, and Alcohol Consumption:

• Infant Deaths may have the largest immediate effect on life expectancy, as even small reductions in infant mortality can significantly increase life expectancy in countries with high infant death rates.

• Hepatitis B vaccination also has a strong positive impact, as improving vaccination rates can lead to better overall health outcomes and longer life expectancies.

• Alcohol Consumption plays a role, but its effect may be more gradual or less direct compared to the other factors.

Model Summary Interpretation:

• R-squared (0.193): This indicates that 19.3% of the variability in life expectancy is explained by Hepatitis B vaccination rates, Infant deaths, and Alcohol consumption. While this is a modest fit, it still provides insights into how these factors impact life expectancy.

• Coefficients:

o Hepatitis B (0.064): A 1% increase in Hepatitis B vaccination coverage is associated with a 0.064-year increase in life expectancy.

o Infant deaths (-0.0092): For every additional infant death per 1,000 live births, life expectancy decreases by 0.0092 years.

o Alcohol (0.714): A 1-liter increase in alcohol consumption is associated with a 0.714-year increase in life expectancy (counterintuitive, potentially due to interaction with other factors or data issues).