**Summary:**

This multivariate regression analysis uses deaths related to heroin, cocaine, psychostimulants, and natural/semi-synthetic opioids as explanatory variables to predict deaths involving synthetic opioids (excluding methadone), aiming to understand the relationships between these factors.

**Filtering Description and Rationale:**

* **Filtering Process**: The dataset is filtered to retain only rows corresponding to five specific indicators: heroin, cocaine, psychostimulants, natural/semi-synthetic opioids, and synthetic opioids (excluding methadone). After pivoting to a wide format with indicators as columns, rows with missing values are removed.
* **Rationale**: Filtering ensures the analysis focuses on relevant substances related to overdose deaths, while dropping rows with missing data avoids potential biases or inaccuracies introduced by incomplete records.

**Statistics Calculated with This Code:**

1. **Regression Coefficients (Beta)**: Measure the influence of each explanatory variable on the response variable.
2. **R-Squared**: Indicates how well the model explains the variance in the response variable.
3. **Residuals**: Differences between actual and predicted values, used to assess the model's accuracy.
4. **F-Statistic**: Evaluates the overall significance of the regression model.
5. **Standard Errors of Coefficients**: Quantify the variability in the estimated coefficients.
6. **T-Statistics**: Test the significance of each coefficient.
7. **P-Values**: Provide significance levels for each explanatory variable's contribution to the model.
8. **Residual Mean and Standard Deviation**: Summarize the error distribution in predictions.

**Final Regression Analysis Report**

**Step 1: Data Loading and Preprocessing**

* **Dataset Loaded Successfully: Data was loaded, and non-numeric entries in the "Data Value" column were coerced to NaN.**
* **Selected Indicators: Indicators used as predictors and response:**
  + **Predictors: 'Heroin (T40.1)', 'Cocaine (T40.5)', 'Psychostimulants with abuse potential (T43.6)', 'Natural & semi-synthetic opioids (T40.2)'**
  + **Response: 'Synthetic opioids, excl. methadone (T40.4)'**
* **Filtered Dataset:**
  + **Initial rows: 27,930**
  + **After pivoting: 4,367**
  + **After removing missing data: 2,162**
* **Shape of Final Data:**
  + **Predictors: 2,162 observations × 4 variables**
  + **Response: 2,162 observations**
* **Intercept Added: The predictors now include a constant column for the regression intercept.**

**Step 2: Regression Coefficients (β\betaβ)**

* **Intercept (β0\beta\_0β0​): 53.6544  
  (Baseline value of the response variable when all predictors are zero.)**
* **Predictors:**
  + **β1\beta\_1β1​ (Heroin): -0.1978 (Inverse relationship with the response.)**
  + **β2\beta\_2β2​ (Cocaine): 1.5846 (Positive relationship with the response.)**
  + **β3\beta\_3β3​ (Psychostimulants): 0.5209 (Positive relationship with the response.)**
  + **β4\beta\_4β4​ (Natural/Semi-Synthetic Opioids): -0.1387 (Inverse relationship with the response.)**

**Step 3: Predictions**

* **Sample Predictions (first 10):  
  Predicted values of synthetic opioid deaths:  
  [73.08, 78.68, 77.94, 74.12, 73.96, 78.15, 93.42, 105.06, 102.01, 95.62]**

**Step 4: Residuals**

* **Residuals: The difference between actual and predicted values of the response variable.**
  + **Sample Residuals (first 10):  
    [-62.08, -66.68, -64.94, -62.12, -62.96, -68.15, -81.42, -76.06, -65.01, -84.61]**
* **Residual Summary:**
  + **Mean Residual: -0.0000 (no significant bias in predictions).**
  + **Residual Standard Deviation: 168.36 (spread of residuals).**

**Step 5: Model Performance**

* **R-squared (R2R^2R2): 0.5927**
  + **Explains 59.27% of the variance in the response variable.**
* **F-Statistic: 784.8368**
  + **Indicates the model is statistically significant as a whole.**

**Step 6: Additional Metrics**

* **Standard Errors (SE): Reflect the variability in coefficients:**
  + **SE[0] (Intercept): 7.3546**
  + **SE[1] (Heroin): 0.0358**
  + **SE[2] (Cocaine): 0.0409**
  + **SE[3] (Psychostimulants): 0.0192**
  + **SE[4] (Natural/Semi-Synthetic Opioids): 0.0337**
* **T-Statistics: Test the significance of each coefficient:**
  + **t[0] = 7.2954**
  + **t[1] = -5.5180**
  + **t[2] = 38.7807**
  + **t[3] = 27.0848**
  + **t[4] = -4.1196**
* **P-Values: Probability that the coefficient is zero under the null hypothesis:**
  + **p[0]: 4.16×10−134.16 \times 10^{-13}4.16×10−13**
  + **p[1]: 3.84×10−83.84 \times 10^{-8}3.84×10−8**
  + **p[2]: 0.00.00.0 (Cocaine highly significant)**
  + **p[3]: 0.00.00.0 (Psychostimulants highly significant)**
  + **p[4]: 3.94×10−53.94 \times 10^{-5}3.94×10−5**

**Step 7: Regression Summary**

* **Regression Coefficients (Beta):  
  [53.6544, -0.1978, 1.5846, 0.5209, -0.1387]**
* **Standard Errors (SE):  
  [7.3546, 0.0358, 0.0409, 0.0192, 0.0337]**
* **T-Statistics:  
  [7.2954, -5.5180, 38.7807, 27.0848, -4.1196]**
* **P-Values:  
  [4.16e-13, 3.84e-08, 0.0, 0.0, 3.94e-05]**

**Step 8: Plotting Results**

* **Plot Saved:  
  The plot of actual vs. predicted synthetic opioid deaths was saved as predicted\_vs\_actual.png.**

**Final Summary**

* **Mean Residual: -0.0000**
* **Residual Standard Deviation: 168.36**
* **R-squared: 0.5927 (59.27% of variance explained)**
* **F-Statistic: 784.8368 (Overall model significance)**

**The model provides a moderately strong explanation of the relationship between the predictors and synthetic opioid deaths. Significant predictors include cocaine deaths and psychostimulant deaths.**