# Analysis of Synthetic Opioid Deaths Using Multiple Regression

## Main Goal

The main goal of this analysis is to build a multiple linear regression model to:

1. Understand Relationships: Examine how various drug-related deaths (e.g., heroin, cocaine, psychostimulants, and natural/semi-synthetic opioids) contribute to the number of synthetic opioid deaths.

2. Predict Synthetic Opioid Deaths: Use these predictors to estimate or forecast the number of synthetic opioid deaths based on trends in other drug-related deaths.

3. Identify Key Drivers: Determine which factors (predictors) are most significantly associated with synthetic opioid deaths, helping policymakers, researchers, and public health officials focus interventions on critical areas.

## Response Variable

Synthetic opioids, excl. methadone (T40.4): The target variable to model and predict the number of deaths involving synthetic opioids.

## Predictor Variables

Based on data completeness and relevance, the chosen predictors are:

1. Heroin (T40.1): Strongly related to opioid use; non-missing data should be sufficient.

2. Cocaine (T40.5): Frequently co-occurring with synthetic opioid deaths; data availability is decent.

3. Psychostimulants with abuse potential (T43.6): Could contribute due to similar usage patterns.

4. Natural & semi-synthetic opioids (T40.2): Includes prescription opioids, which may correlate with synthetic opioid trends.

## Excluded Variables

Variables with higher missing data or less direct relevance were excluded:

1. Methadone (T40.3): Limited direct correlation to synthetic opioids and possibly higher missing values.

2. Natural, semi-synthetic, & synthetic opioids, incl. methadone (T40.2-T40.4): This combined category overlaps with other predictors and may introduce multicollinearity.

## Conclusion

This analysis aims to provide insights into the opioid crisis by analyzing how different types of substance abuse are interconnected and contribute to synthetic opioid mortality rates. The findings will help inform interventions and strategies for addressing this pressing public health issue.  
  
**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.**