Pearson Correlation Coefficient: Step-by-Step Calculation

Step 1: Define Two Continuous Variables

Given two datasets:

$$X = [43, 21, 25, 42, 57, 59, 79, 75, 87, 81]$$

$$Y = [99, 65, 79, 75, 87, 81, 109, 105, 125, 119]$$

Step 2: Compute Pearson Correlation Coefficient Using SciPy

The Pearson Correlation Coefficient is calculated using the formula:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

In Python, this can be computed using the `scipy.stats.pearsonr` function:

$$r, p\_value = \texttt{pearsonr}(X, Y)$$

The result is:

$$r \approx 0.96, \quad p\_value \approx 0.0001$$

Step 3: Compute the Numerator (Covariance Between X and Y)

The covariance is given by:

$$\text{Cov}(X, Y) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

First, compute the means:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

For X:

$$\bar{x} = \frac{43 + 21 + 25 + 42 + 57 + 59 + 79 + 75 + 87 + 81}{10} = 56.9$$

For Y:

$$\bar{y} = \frac{99 + 65 + 79 + 75 + 87 + 81 + 109 + 105 + 125 + 119}{10} = 94.4$$

Now calculate the product of deviations for each data point:

$$(x_1 - \bar{x})(y_1 - \bar{y}) = (43 - 56.9)(99 - 94.4) = -13.9 \cdot 4.6 = -63.94$$

$$(x_2 - \bar{x})(y_2 - \bar{y}) = (21 - 56.9)(65 - 94.4) = -35.9 \cdot -29.4 = 1055.46$$

Compute this for all points, then sum:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 3771.48$$

Step 4: Compute the Denominator

The denominator is the product of the standard deviations of X and Y:

$$\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}$$

For X, compute squared deviations:

$$(43 - 56.9)^2 = 193.21, \quad (21 - 56.9)^2 = 1288.81, \quad \ldots$$

Sum the squared deviations:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 5031.33$$

For Y, repeat the calculation:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = 2036.16$$

The denominator is:

$$\sqrt{5031.33 \cdot 2036.16} \approx 3197.24$$

Step 5: Compute the Pearson Correlation Coefficient

Substitute the values:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$r = \frac{3771.48}{3197.24} \approx 0.96$$

Step 6: Visualize the Relationship

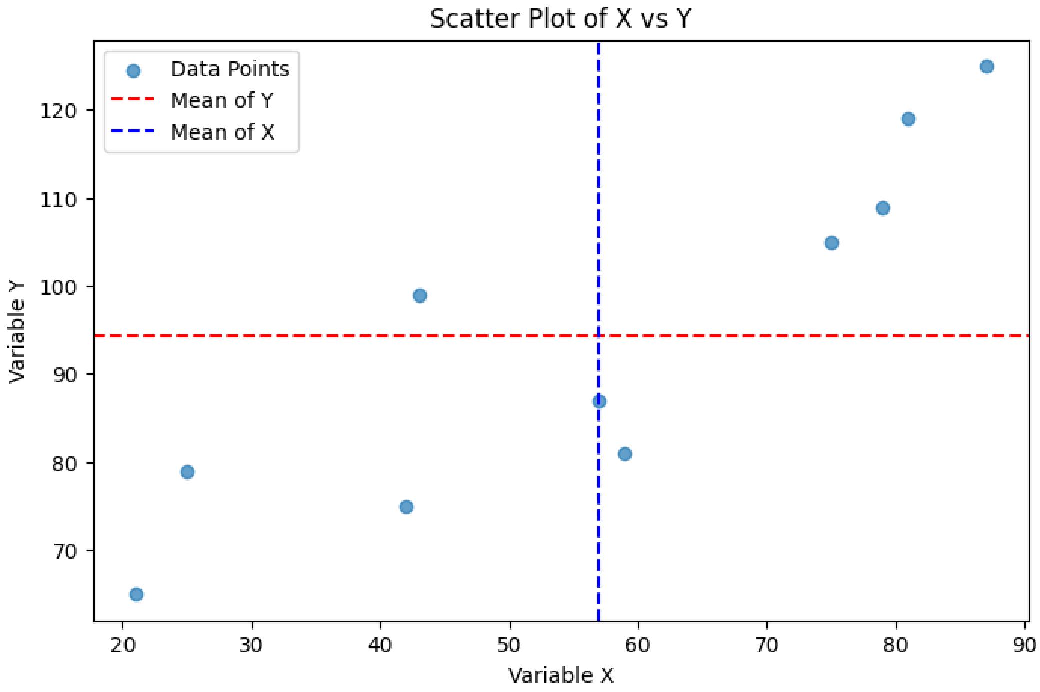
The relationship between X and Y can be visualized using a scatter plot, with horizontal and vertical lines representing the means ($\bar{x}$ and $\bar{y}$).

The Python code for visualization is:

```
plt.figure(figsize=(8, 5))
plt.scatter(x, y, alpha=0.7, label="Data Points")
plt.xlabel("Variable X")
plt.ylabel("Variable Y")
plt.title("Scatter Plot of X vs Y")
plt.axhline(y=np.mean(y), color='red', linestyle='--', label='Mean of Y')
plt.axvline(x=np.mean(x), color='blue', linestyle='--', label='Mean of X')
plt.legend()
plt.show()
```

Compile Solution

```
Pearson Correlation Coefficient Results
Correlation Coefficient (r): 0.89
P-Value: 0.0006
Manually Calculated Correlation Coefficient (r): 0.89
```


Scatter Plot of X vs Y
Data Points
Mean of Y
Mean of X
Variable Y
Variable X