One-Way ANOVA: Step-by-Step Implementation in Python

Step 1: Define Data for Multiple Groups

We are given the following datasets for three groups:

$$\text{Group 1: } [85, 90, 88, 75, 95]$$

$$\text{Group 2: } [70, 65, 60, 75, 80]$$

$$\text{Group 3: } [50, 55, 60, 65, 70]$$

Step 2: Combine Groups for Total Calculations

Combine all data points from the groups:

$$\text{All Data: } [85, 90, 88, 75, 95, 70, 65, 60, 75, 80, 50, 55, 60, 65, 70]$$

Calculate the grand mean ($\bar{X}$):

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Substitute the values:

$$\bar{X} = \frac{85 + 90 + \ldots + 70}{15} = 72.67$$

Step 3: Compute Group Means

The means for each group are:

$$\bar{X}_1 = \frac{85 + 90 + 88 + 75 + 95}{5} = 86.6$$

$$\bar{X}_2 = \frac{70 + 65 + 60 + 75 + 80}{5} = 70.0$$

$$\bar{X}_3 = \frac{50 + 55 + 60 + 65 + 70}{5} = 60.0$$

Step 4: Calculate Between-Group Variability (SSB)

The formula for the between-group variability is:

$$SSB = \sum_{k=1}^{g} n_k (\bar{X}_k - \bar{X})^2$$

Substitute the values for each group:

$$SSB = 5(86.6 - 72.67)^2 + 5(70.0 - 72.67)^2 + 5(60.0 - 72.67)^2$$

$$SSB = 5(13.93)^2 + 5(-2.67)^2 + 5(-12.67)^2$$

$$SSB = 5(194.03) + 5(7.13) + 5(160.43) = 970.75$$

Step 5: Calculate Within-Group Variability (SSW)

The formula for the within-group variability is:

$$SSW = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)^2$$

For Group 1:

$$SSW_1 = (85 - 86.6)^2 + (90 - 86.6)^2 + \ldots + (95 - 86.6)^2 = 132.8$$

For Group 2:

$$SSW_2 = (70 - 70.0)^2 + (65 - 70.0)^2 + \ldots + (80 - 70.0)^2 = 250.0$$

For Group 3:

$$SSW_3 = (50 - 60.0)^2 + (55 - 60.0)^2 + \ldots + (70 - 60.0)^2 = 500.0$$

Combine:

$$SSW = SSW_1 + SSW_2 + SSW_3 = 132.8 + 250.0 + 500.0 = 882.8$$

Step 6: Calculate Degrees of Freedom

- **Between-group degrees of freedom:**

$$\text{df}_{\text{between}} = g - 1 = 3 - 1 = 2$$

- **Within-group degrees of freedom:**

$$\text{df}_{\text{within}} = N - g = 15 - 3 = 12$$

Step 7: Calculate Mean Squares (MSB and MSW)

The formula for the mean square between (MSB) is:

$$MSB = \frac{SSB}{\text{df}_{\text{between}}} = \frac{970.75}{2} = 485.38$$

The formula for the mean square within (MSW) is:

$$MSW = \frac{SSW}{\text{df}_{\text{within}}} = \frac{882.8}{12} = 73.57$$

Step 8: Calculate the F-Statistic

The formula for the F-statistic is:

$$F = \frac{MSB}{MSW}$$

Substitute the values:

$$F = \frac{485.38}{73.57} \approx 6.60$$

Step 9: Perform ANOVA Using SciPy

Using Python's `scipy.stats.f_oneway` function, we can compute the F-statistic and p-value:

```
f_stat_scipy, p_value = f_oneway(group1, group2, group3)
```

The results are:

$$F_{\text{scipy}} \approx 6.60, \quad p_{\text{value}} \approx 0.01$$

Step 10: Interpret Results

Compare the p-value with $\alpha = 0.05$:

- If $p < \alpha$: Reject the null hypothesis. At least one group mean is significantly different.
- If $p \geq \alpha$: Fail to reject the null hypothesis. No significant difference between group means.

Since $p \approx 0.01 < 0.05$, we reject the null hypothesis.

Step 11: Visualize Group Data

The Python code to generate a boxplot is:

```
plt.figure(figsize=(8, 6))
data.boxplot(by='Group', column=['Values'], grid=False, notch=True)
plt.title('Group Comparisons')
plt.suptitle("")
plt.xlabel('Groups')
plt.ylabel('Values')
plt.show()
```

This visualization compares the distributions of the groups.

Compile Solution

```
One-Way ANOVA Results
Grand Mean: 72.20
SSB (Between-Group Variability): 1805.20
SSW (Within-Group Variability): 721.20
MSB: 902.60, MSW: 60.10
F-Statistic (Manual Calculation): 15.02
F-Statistic (SciPy): 15.02
P-Value (SciPy): 0.0005
Result: Reject the null hypothesis. At least one group mean is
significantly different.
```