# Practical 5

Perform one-way ANOVA to compare means across multiple groups. Conduct post-hoc tests to identify significant differences

# ANOVA (Analysis of Variance)

ANOVA is used to test if there are statistically significant differences between the means of three or more groups. A **One-way ANOVA** tests for differences in the means of a single independent variable across multiple groups.

# Steps for Performing One-Way ANOVA:

1. **Formulate Hypotheses:**
   * **Null Hypothesis (H₀):** The means of all groups are equal.
   * **Alternative Hypothesis (H₁):** At least one group mean is different from the others.
2. **Perform the ANOVA Test:**
   * We'll use scipy.stats.f\_oneway for the One-Way ANOVA.
3. **Post-hoc Test (Tukey's Test):**
   * If the ANOVA test indicates significant differences, a post-hoc test like **Tukey's HSD** (Honestly Significant Difference) can be performed to identify which specific groups have significant differences.
   * We will use statsmodels.stats.multicomp.pairwise\_tukeyhsd for this.

# Python Code for One-Way ANOVA and Post-hoc Test

from statsmodels.stats import multicomp import matplotlib.pyplot as plt

import seaborn as sns

# Sample Data: Salaries in different departments data = {

'Department': ['Marketing', 'Marketing', 'Marketing', 'Sales', 'Sales', 'Sales', 'HR', 'HR', 'HR'],

'Salary': [50000, 52000, 53000, 55000, 57000, 60000, 45000, 47000, 48000]

}

# Creating DataFrame df = pd.DataFrame(data)

# Perform One-Way ANOVA

marketing\_salaries = df[df['Department'] == 'Marketing']['Salary'] sales\_salaries = df[df['Department'] == 'Sales']['Salary'] hr\_salaries = df[df['Department'] == 'HR']['Salary']

f\_statistic, p\_value = stats.f\_oneway(marketing\_salaries, sales\_salaries, hr\_salaries)

print(f"F-statistic: {f\_statistic}") print(f"P-value: {p\_value}")

# Hypothesis Interpretation alpha = 0.05

if p\_value < alpha:

print("Reject the null hypothesis. There is a significant difference in the salaries across departments.")

else:

print("Fail to reject the null hypothesis. There is no significant difference in the salaries across departments.")

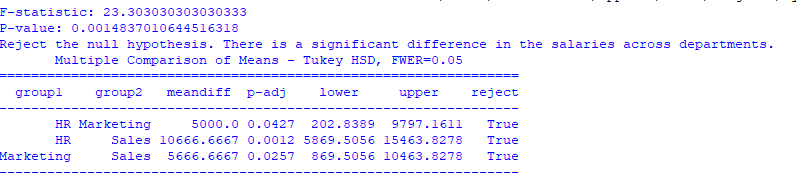
# Post-hoc test (Tukey HSD)

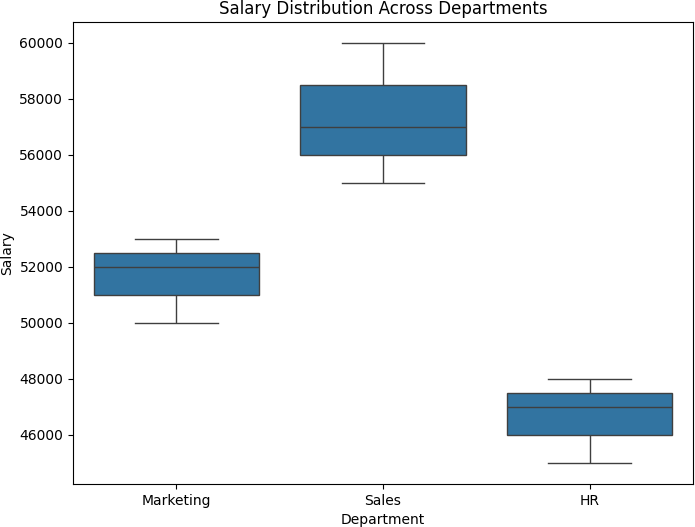
comp = multicomp.pairwise\_tukeyhsd(df['Salary'], df['Department'], alpha=0.05)

print(comp)

# Visualization (Boxplot to show the salary distribution across departments) plt.figure(figsize=(8,6))

sns.boxplot(x='Department', y='Salary', data=df) plt.title('Salary Distribution Across Departments') plt.show()





# Conclusion:

* **One-Way ANOVA** helps test if there's a significant difference in means between three or more groups.
* If the test is significant, a **post-hoc test** (like **Tukey's HSD**) identifies which specific pairs of groups have significant differences in their means.