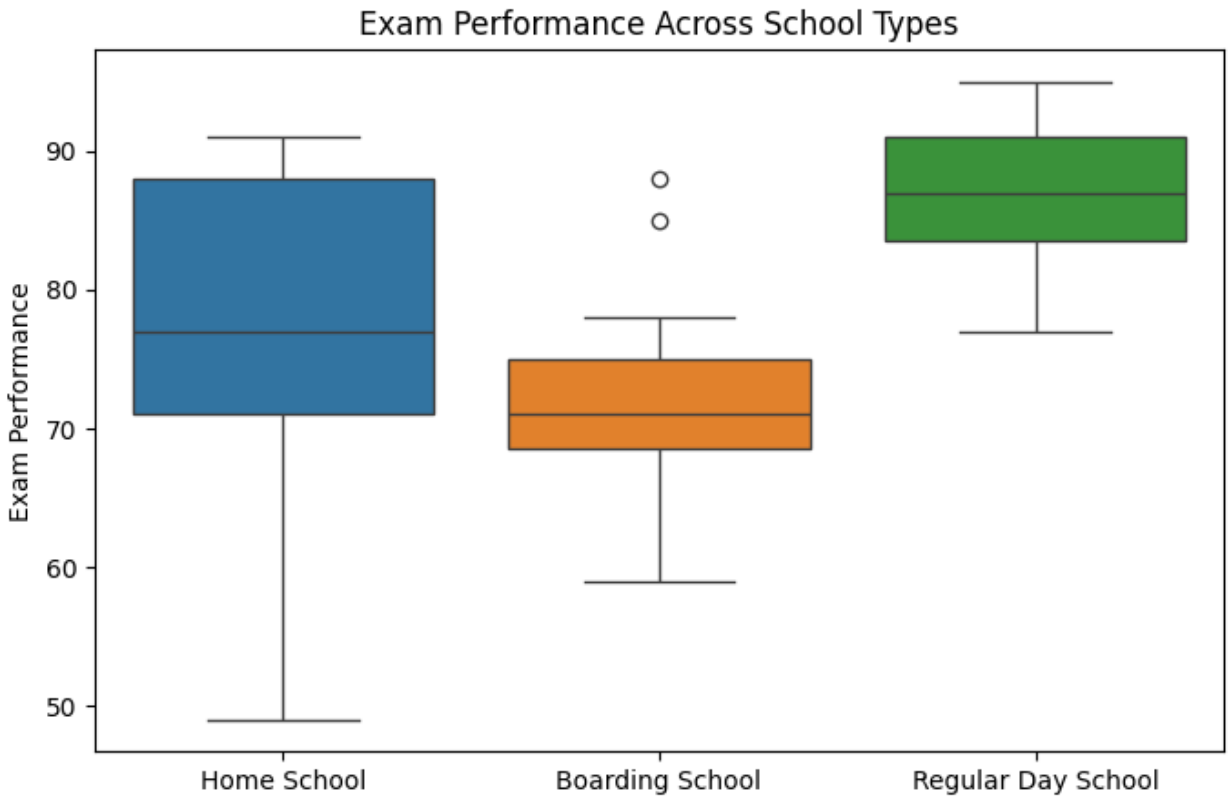
ANOVA- In Class Activity

Roll No: 2021102016

```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Data extracted from the image
home_school = np.array([89, 75, 49, 84, 88, 77, 73, 67, 69, 88, 91])
boarding_school = np.array([85, 78, 59, 88, 71, 69, 72, 72, 68, 59,
70])
regular_day_school = np.array([91, 88, 84, 91, 95, 93, 85, 87, 83, 80,
77])

# 1. Check for Normality (Shapiro-Wilk Test)
shapiro_home = stats.shapiro(home_school)
shapiro_boarding = stats.shapiro(boarding_school)
shapiro_regular = stats.shapiro(regular_day_school)

print("### Normality Check (Shapiro-Wilk Test)")
print(f"Home School: p-value = {shapiro_home.pvalue:.4f}")
print(f"Boarding School: p-value = {shapiro_boarding.pvalue:.4f}")
print(f"Regular Day School: p-value = {shapiro_regular.pvalue:.4f}\n")

# 2. Check Homogeneity of Variance (Levene’s Test)
levene_test = stats.levene(home_school, boarding_school,
regular_day_school)
print("### Homogeneity of Variance (Levene's Test)")
print(f"Levene's Test p-value = {levene_test.pvalue:.4f}\n")

# 3. Perform One-Way ANOVA
anova_test = stats.f_oneway(home_school, boarding_school,
regular_day_school)
print("### ANOVA Test Results")
print(f"ANOVA p-value = {anova_test.pvalue:.4f}\n")

# 4. Compute Effect Size (Only if ANOVA is significant)
if anova_test.pvalue < 0.05:
    SS_between = anova_test.statistic * (len(home_school) - 1)  # Sum
of squares between
    SS_total = np.var(np.concatenate([home_school, boarding_school,
regular_day_school]), ddof=1) * (len(home_school) * 3 - 1)
    eta_squared = SS_between / SS_total
```

```
    print(f"### Effect Size (Eta Squared) for Significant ANOVA")
    print(f"Effect Size = {eta_squared:.4f}\n")

# 5. Visualization (Box Plot)
plt.figure(figsize=(8, 5))
sns.boxplot(data=[home_school, boarding_school, regular_day_school])
plt.xticks([0, 1, 2], ["Home School", "Boarding School", "Regular Day
School"])
plt.ylabel("Exam Performance")
plt.title("Exam Performance Across School Types")
plt.show()

# 6. Perform Post-hoc Test (if ANOVA is significant)
if anova_test.pvalue < 0.05:
    # Combine data and create group labels
    all_scores = np.concatenate([home_school, boarding_school,
regular_day_school])
    groups = (["Home School"] * len(home_school)) + (["Boarding
School"] * len(boarding_school)) + (["Regular Day School"] *
len(regular_day_school))

    # Perform Tukey's HSD test
    tukey_result = pairwise_tukeyhsd(all_scores, groups, alpha=0.05)

    print("### Tukey's HSD Test Results")
    print(tukey_result)
```

```
### Normality Check (Shapiro-Wilk Test)
Home School: p-value = 0.1752
Boarding School: p-value = 0.3620
Regular Day School: p-value = 0.9488

### Homogeneity of Variance (Levene's Test)
Levene's Test p-value = 0.1225

### ANOVA Test Results
ANOVA p-value = 0.0037

### Effect Size (Eta Squared) for Significant ANOVA
Effect Size = 0.0171
```



Tukey's HSD Test Results

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Boarding School	Home School	5.3636	0.3971	-4.6775	15.4047	False
Boarding School	Regular Day School	14.8182	0.0029	4.7771	24.8593	True
Home School	Regular Day School	9.4545	0.0681	-0.5866	19.4957	False

Analysis of Exam Performance Across School Types

1. Normality Check (Shapiro-Wilk Test)

The Shapiro-Wilk test checks whether the data in each group follows a normal distribution.

- **Home School:** (p = 0.1752) (Normally distributed)
- **Boarding School:** (p = 0.3620) (Normally distributed)
- **Regular Day School:** (p = 0.9488) (Normally distributed)

Interpretation: Since all p-values are **greater than 0.05**, we fail to reject the null hypothesis, meaning that the data is normally distributed for all groups.

---

2. Homogeneity of Variances (Levene's Test)

Levene's test checks whether the variances of the three groups are equal.

- **Levene's Test p-value:** (p = 0.1225)

Interpretation: Since (p > 0.05), we fail to reject the null hypothesis, meaning the variances are **equal** across all groups. This satisfies the assumption for ANOVA.

---

3. ANOVA Test Results

ANOVA tests whether there is a significant difference in mean exam scores among the three school types.

- **ANOVA p-value:** (p = 0.0037)

Interpretation: Since (p < 0.05), we reject the null hypothesis, meaning **at least one school type has a significantly different mean performance**.

---

4. Effect Size (Eta Squared)

Effect size (Eta Squared) measures the strength of the relationship between school type and exam performance.

- **Effect Size (Eta Squared):** (\eta^2 = 0.0171)

Interpretation:

- (n^2 < 0.01) → Small effect
- (0.01 <= n^2 < 0.06) → **Medium effect** (Our effect size is in this range)
- (n^2 >= 0.06) → Large effect

Conclusion: The difference in exam scores between school types has a **moderate effect size**.

---

5. Tukey's HSD Test (Post-hoc Analysis)

Since ANOVA was **significant**, we perform Tukey's test to identify which groups differ significantly.

Comparison	Mean Difference	p-value	Significant?
Boarding vs. Home School	5.36	0.3971	No
Boarding vs. Regular Day	14.82	0.0029	Yes
Home vs. Regular Day	9.45	0.0681	No

Interpretation:

- **Boarding vs. Regular Day School:** Significant difference (**p = 0.0029**).
- **Home vs. Boarding & Home vs. Regular Day:** No significant difference.

Findings

All groups are normally distributed.
Variances are equal (Levene's test).
ANOVA shows significant differences in exam performance.
Effect size indicates a moderate difference.
Post-hoc analysis:

- **Regular Day School** performs **significantly better** than **Boarding School**.
- **No significant difference** between Home School and the other groups.

Conclusion

- Students from **Regular Day Schools** performed **significantly better** than those from **Boarding Schools**.
- No significant difference was found between **Home School** and the other groups.
- The effect size suggests **moderate differences** in performance based on school type.