MATH/STAT571B: Design of Experiments

Investigating Noise Levels in Campus Study Locations

Group 2

Humaira Nafisa Ahmed,

Paige Elizabeth Cooper,

Kai Shuen Neo,

R M Shahbab

University of Arizona

May 5, 2025

Table of Contents

Introduction

- **Purpose:** Evaluate noise levels in popular campus study locations to help students choose optimal study environments.
  - Noise affects study preferences. Some students prefer a quiet, calm environment, while others appreciate background noise.
  - The environment can significantly impact a student's ability to concentrate
- **Locations:** Main floor of the library, student union, engineering building common room
- **Factors studied:**
  - Time of the day: Morning (AM) vs. afternoon (PM).
  - Day of the week: Monday (beginning), Wednesday (middle), Friday (end)
- **Measurement methods:**
  - Objective rating: Average decibel readings via mobile app
  - Subjective rating: 0-10 noise levels
- **Research questions:**
  - Which study location is the most comfortable in terms of noise and distraction?
  - How does the expected noise of each location differ during the morning and afternoon?
  - Does the expected noise for each location change based on day of the week (beginning, middle, and end)?

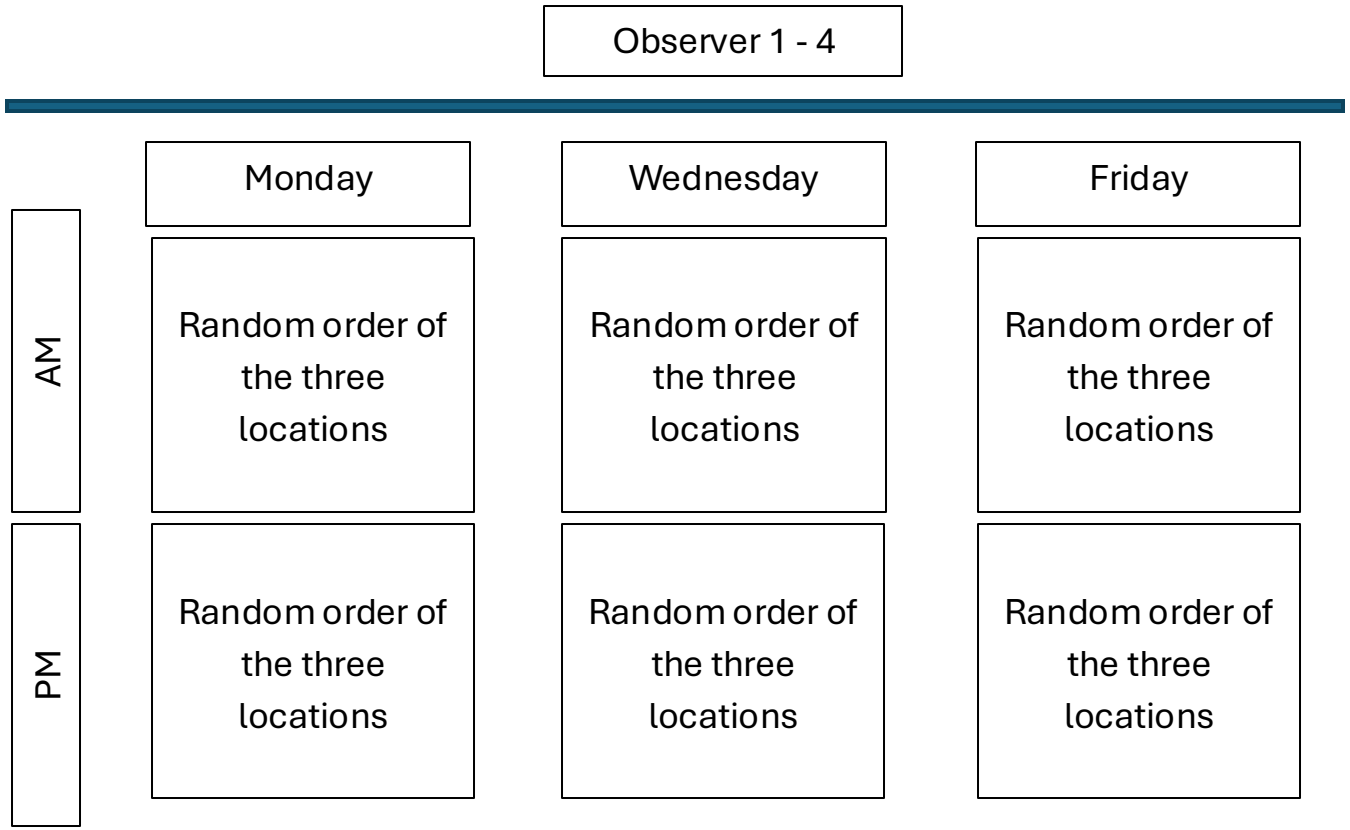
Motivation

- Helping students **choose the right study locations** based on actual data can enhance their study experience
- Noise levels can **vary by time of the day and day of the week**, but their impacts on study locations have not been documented for various locations on campus.
- Combining **subjective and objective ratings** ensures a comprehensive understanding of the comfort of the study environment
- Findings of the study on noise levels can assist the campus planners in **improving study spaces**

Proposed Design

- **Design: Full factorial design**
  - Three factors
    - Factor *a*: Location (3 levels)
    - Factor *b*: Day of week (3 levels)
    - Factor *c*: Time of day (2 levels)
    - Factor *d*: Blocks (4 observers)
  - Replicate, n = 2
  - Total number of observations,

    $a \times b \times c \times d \times e = 144$



Proposed Design

- Considers the main effects - location, day, and time, and their interactions, such as whether the location effect varies by day or time
- Helps identify which combinations of day, time, and location yield the highest or lowest levels of comfort or noise
- Allow us to formulate and test many hypotheses using ANOVA easily
- Including random effects for observers is essential as it accounts for their variability
- Protects against any bias present based on personal opinion or differences in mobile device sensitivity
- Results are generalized to the student population at the university

Proposed Analysis

The statistical model for the proposed design is as follows

$$Y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \delta_l + \epsilon_{ijklm}$$

- Y_{ijklm} *is the observed response for the m-th replicate in the combination of i-th Location, j-th Day of Week, k-th Time of Day, and l-th Observer.*
- μ *is the overall mean of the response.*
- α_i *is the fixed effect of Location (i = 1, 2, 3).*
- β_j *is the fixed effect of Day of Week (j = 1, 2, 3).*
- γ_k *is the fixed effect of Time of Day (k = 1, 2).*
- $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, *and* $(\alpha\beta\gamma)_{ijk}$ *represent the interaction effects.*
- δ_l *is the random effect of the Observer (Block) (l = 1, 2, 3, 4), assumed to be normally distributed as* $\delta_l \sim N(0, \sigma^2_\delta)$.
- ϵ_{ijklm} *is the random error, assumed to be normally distributed as* $\epsilon_{ijklm} \sim N(0, \sigma^2_\epsilon)$.

Data Collection

- Each observer would collect observations on
  - three days of the week (Monday, Wednesday, Friday),
  - at two times of the day (AM, PM),
  - at three locations (Student Union, Main Library, Common Room)
- Study locations would be randomized across all observers.
- Each observation has subjective and objective ratings
- Determine the response variable, Score
  - normalizing the objective and subjective ratings on the same scale. for any observations, x

$$Score_{subjective} = \frac{x}{Subjective\ rating\ max}$$

$$Score_{objective} = \frac{x}{Objective\ rating\ max}$$

  - averaging their normalized score to get a final composite score (weight = 0.5 for subjective and objective rating)

$$Score = Weight_{subjective} \times Score_{subjective} + Weight_{objective} \times Score_{objective}$$

Data Collection

✓ Sample data collection sheet for observer XX

Random order of location

Subjective ratings

Objective ratings

Response variable, Score = 3.24

Week 1

Monday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Student Union	11.32 am	5	64.33
	Common Room	11.15 am	1	45.1
	Main Library	11.52 am	2	56.3
PM	Student Union	12.31 pm	6	72.1
	Main Library	12.10 pm	2	55.25
	Common Room	12.45 pm	1	45.9

Wednesday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Common Room	11.00 am	2	47.6
	Main Library	11.34 am	3	55.95
	Student Union	11.14 am	7	74.65
PM	Main Library	12.20 pm	3	63.2
	Common Room	1.17 pm	1	44.4
	Student Union	1.00 pm	6	71.6

Friday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Common Room	10.53 am	1	42.9
	Student Union	11.55 am	4	63.6
	Main Library		3	53
PM	Common Room		1	42.33
	Student Union	3.33 pm	4	64
	Main Library	3.47 pm	3	58.15

Week 2

Monday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Main Library	11.14 am	3	56.35
	Student Union	11.21 am	5	68.7
	Common Room	11.42 am	1	44.3
PM	Common Room	1.50 pm	1	40.9
	Student Union	2.32 pm	5	67.3
	Main Library	2.26 pm	2	50.35

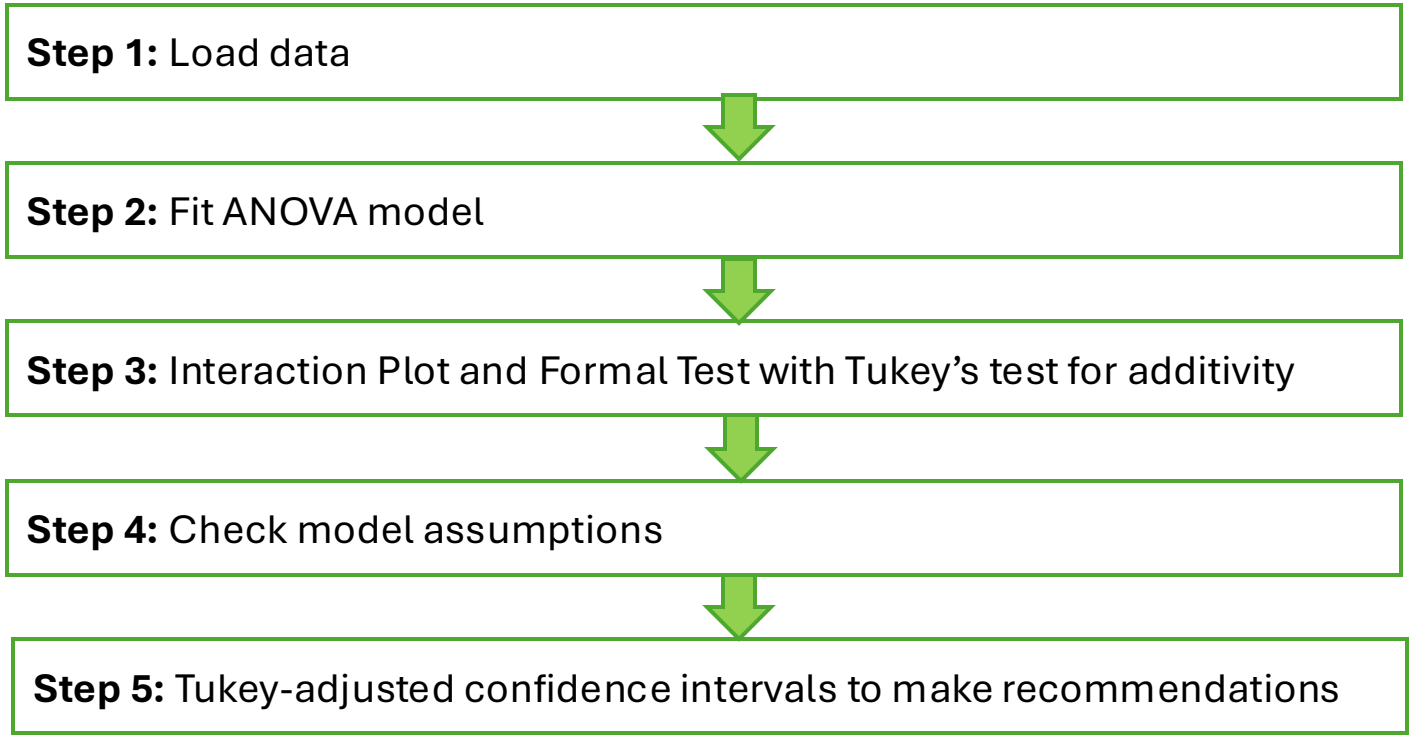
Wednesday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Common Room	10.54 am	1	42.5
	Main Library	11.33 am	4	56.45
	Student Union	11.12 am	6	70.8
PM	Student Union	2.55 pm	7	72.3
	Common Room	2.50 pm	3	53.3
	Main Library	12.05 pm	5	60.5

Friday

	Location Order	Time	Subjective Rating	Objective Rating (dB)
AM	Student Union	11.30 am	5	65.7
	Common Room	11.20 am	1	43.2
	Main Library	11.42 am	3	50.8
PM	Common Room	12.26 pm	3	51.8
	Main Library	12.09 pm	4	54
	Student Union	12.17 pm	6	67.6

Methodology



Step 1: Load data

- ✓ Import data
- ✓ Convert factors as categorical
- ✓ Check the data frame

```r
#----------------------------------------
# Load data & Convert factors as categorical
#----------------------------------------

# Data Import (may need to change based on where you saved the csv)
study <- read.csv("study.csv")

# Setting factors
study$Person <- as.factor(study$Person)
study$Day <- as.factor(study$Day)
study$Time <- as.factor(study$Time)
study$Replicate <- as.factor(study$Replicate)
study$Location <- as.factor(study$Location)

head(study)
```

```
##   X Person    Day Time Replicate      Location Score Rating   dB
## 1 1  Paige Monday   AM         1  Main Library  5.29      4 49.1
## 2 2  Paige Monday   AM         1   Common Room  3.66      2 40.7
## 3 3  Paige Monday   AM         1 Student Union  7.27      6 62.9
## 4 4  Paige Monday   PM         1 Student Union  8.04      7 66.4
## 5 5  Paige Monday   PM         1   Common Room  4.49      3 45.1
## 6 6  Paige Monday   PM         1  Main Library  6.18      5 54.3
```

Step 2: Fit ANOVA Model

✓ Fitting into ANOVA full factorial model

```
score_model <- lmer(Score ~ Day * Time * Location + (1 | Person), data = study)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                  Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
## Day                1.97   0.983     2   123   1.7601   0.17633
## Time              13.20  13.201     1   123  23.6353 3.479e-06 ***
## Location         384.19 192.096     2   123 343.9298 < 2.2e-16 ***
## Day:Time           0.37   0.184     2   123   0.3302   0.71940
## Day:Location       1.73   0.433     4   123   0.7756   0.54308
## Time:Location      3.93   1.966     2   123   3.5207   0.03259 *
## Day:Time:Location  0.94   0.235     4   123   0.4210   0.79327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation:

❑ Three-way interaction **Day: Time: Location** has a **P-value 0.79327 ≥ 0.05**, therefore, this interaction has no significant effect on the response with 0.05 significance level.

❑ Drop the three-way interaction term

❑ Fit a second-order model

Step 2: Fit ANOVA model

- ✓ Fitting into a second-order ANOVA full factorial model

score_model_2 <- lmer(Score ~ (Day + Time + Location)^2 + (1 | Person), data = study)

- ✓ Identify factors that have statistically significant effects on the response variable (score) in this second-order model

```
## Type III Analysis of Variance Table with Satterthwaite's method
##                Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
## Day              1.97   0.983     2   127   1.7928   0.17068
## Time            13.20  13.201     1   127  24.0744 2.784e-06 ***
## Location       384.19 192.096     2   127 350.3187 < 2.2e-16 ***
## Day:Time         0.37   0.184     2   127   0.3364   0.71501
## Day:Location     1.73   0.433     4   127   0.7900   0.53371
## Time:Location    3.93   1.966     2   127   3.5861   0.03055 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 2: Fit ANOVA model

Interpretation for the second-order model:

- ❑ Neither the factor 'Day' nor its interaction terms are significant
- ❑ Drop the factor 'Day'
- ❑ Fit a second-order but reduced model

```
score_model_red <- lmer(Score ~ Time*Location + (1 | Person), data = study)
```

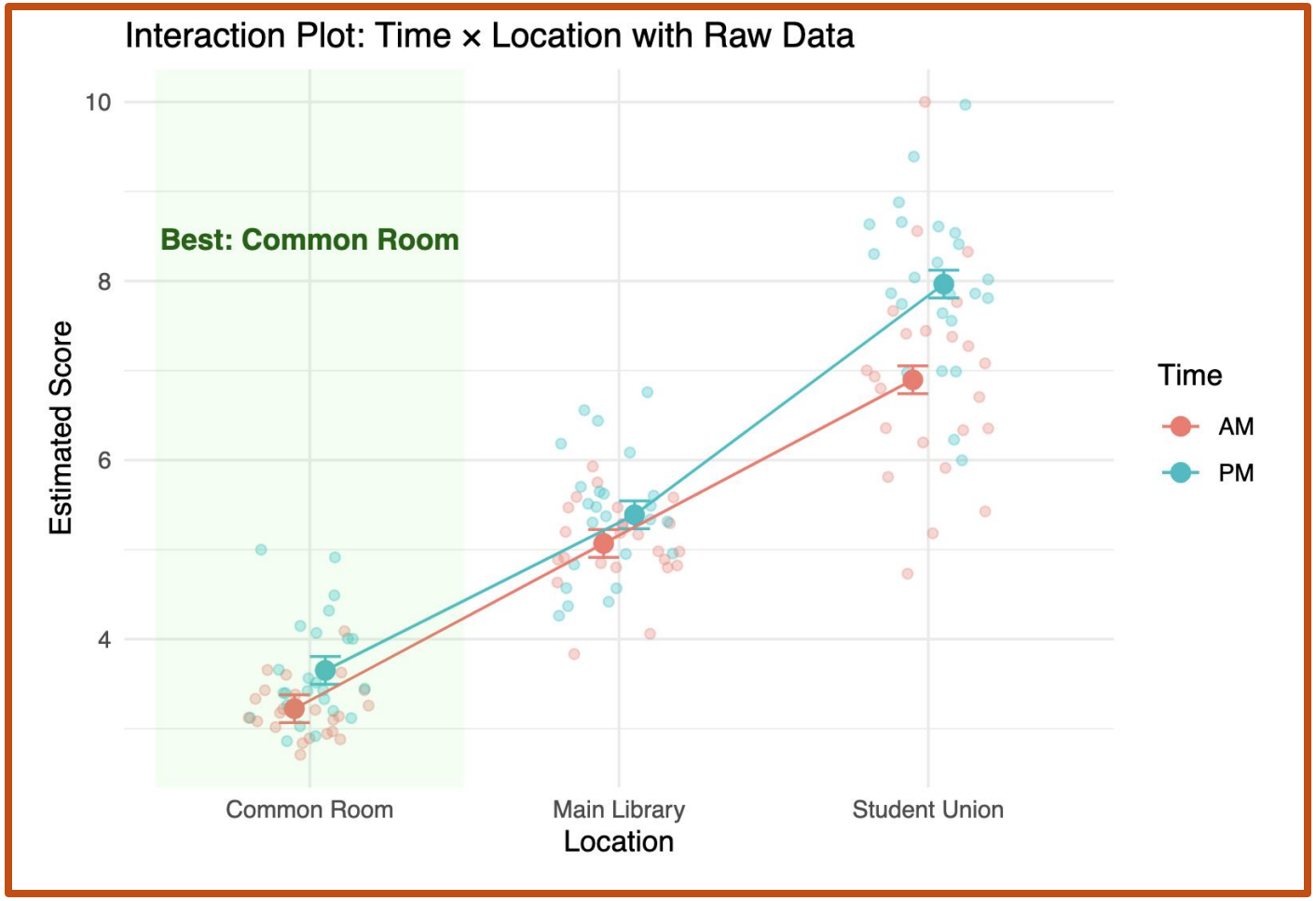
```
## Type III Analysis of Variance Table with Satterthwaite's method
##               Sum Sq Mean Sq NumDF DenDF  F value    Pr(>F)
## Time           13.20  13.201     1   135  24.1785 2.508e-06 ***
## Location      384.19 192.096     2   135 351.8344 < 2.2e-16 ***
## Time:Location   3.93   1.966     2   135   3.6016   0.02993 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation for the second-order reduced model:

- ❑ All factors, i.e. , **Time, Location, and their interactions**, are statistically significant in terms of having effects on the response with 0.05 significance level

Step 3: Interaction Plot

```r
#-----------------------------------------------------
# Interaction Plot
#-----------------------------------------------------

library(ggplot2)
library(emmeans)

# Estimated marginal means
emm <- emmeans(score_model_red, ~ Time * Location)
emm_df <- as.data.frame(emm)

# Plot
ggplot(emm_df, aes(x = Location, y = emmean, color = Time, group = Time)) +

# Raw data jittered (from full dataset)
geom_jitter(data = study, aes(x = Location, y = Score, color = Time),
            width = 0.2, alpha = 0.3, inherit.aes = FALSE) +

# Estimated means with error bars
geom_point(size = 3, position = position_dodge(0.2)) +
geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
              width = 0.2, position = position_dodge(0.2)) +
geom_line(position = position_dodge(0.2)) +

# Annotate the best location
annotate("rect", xmin = 0.5, xmax = 1.5, ymin = -Inf, ymax = Inf,
         alpha = 0.05, fill = "green") +
annotate("text", x = 1, y = max(emm_df$emmean) + 0.5,
         label = "Best: Common Room", color = "darkgreen", fontface = "bold") +

labs(title = "Interaction Plot: Time × Location with Raw Data",
       y = "Estimated Score", x = "Location") +
theme_minimal()
```



Interpretations:

- ❑ Two lines are not perfectly parallel, indicating an interaction between factors on the score.
- ❑ Common room consistently has the lowest score in the morning and evening; Best location to study.

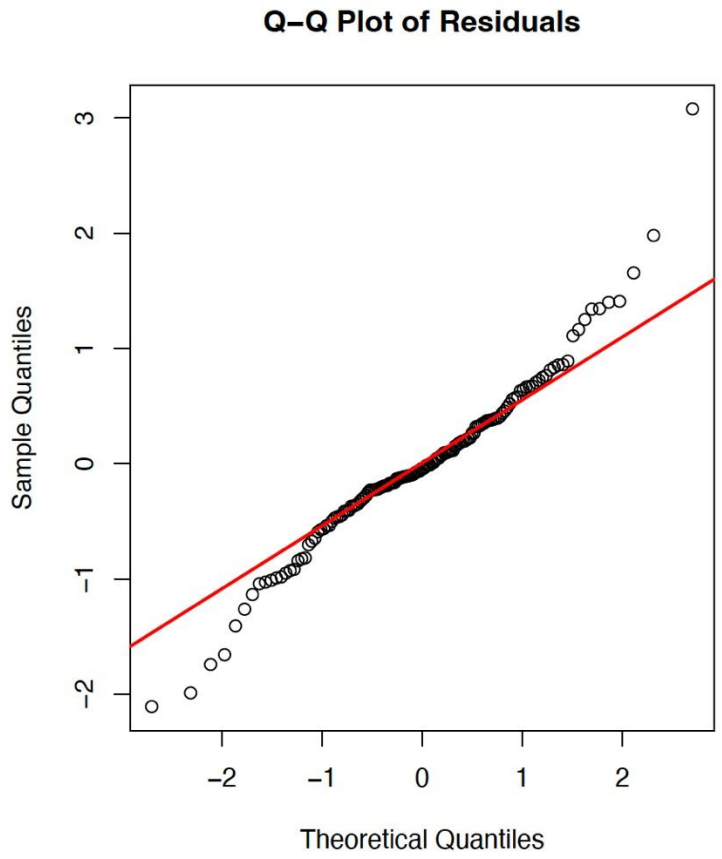
Step 3: Formal Test with Tukey's Test for Additivity

```r
#-------------------------------------------------
# Tukey's Test for NonAdditivity
#-------------------------------------------------

score_add <- aov(Score ~ Time + Location, data = study)
study$q3 <- fitted(score_add)^2
score_tukey_add <- aov(Score ~ Time + Location + q3, data = study)
anova(score_tukey_add)
```

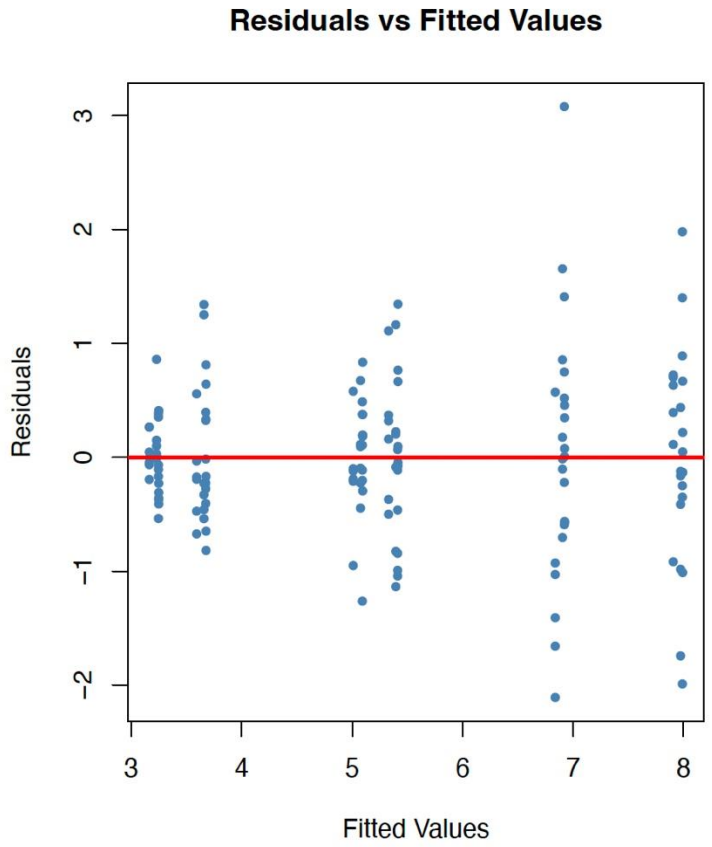
```
## Analysis of Variance Table
##
## Response: Score
##             Df Sum Sq Mean Sq  F value    Pr(>F)
## Time         1  13.20  13.201  23.7634 2.933e-06 ***
## Location     2 384.19 192.096 345.7939 < 2.2e-16 ***
## q3           1   2.68   2.683   4.8292   0.02964 *
## Residuals  139  77.22   0.556
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

❑ **Interpretation:** Test confirms that the interaction effect is significant, as also suggested by the (diagnostic) interaction plot

Step 4: Model Assumption Check



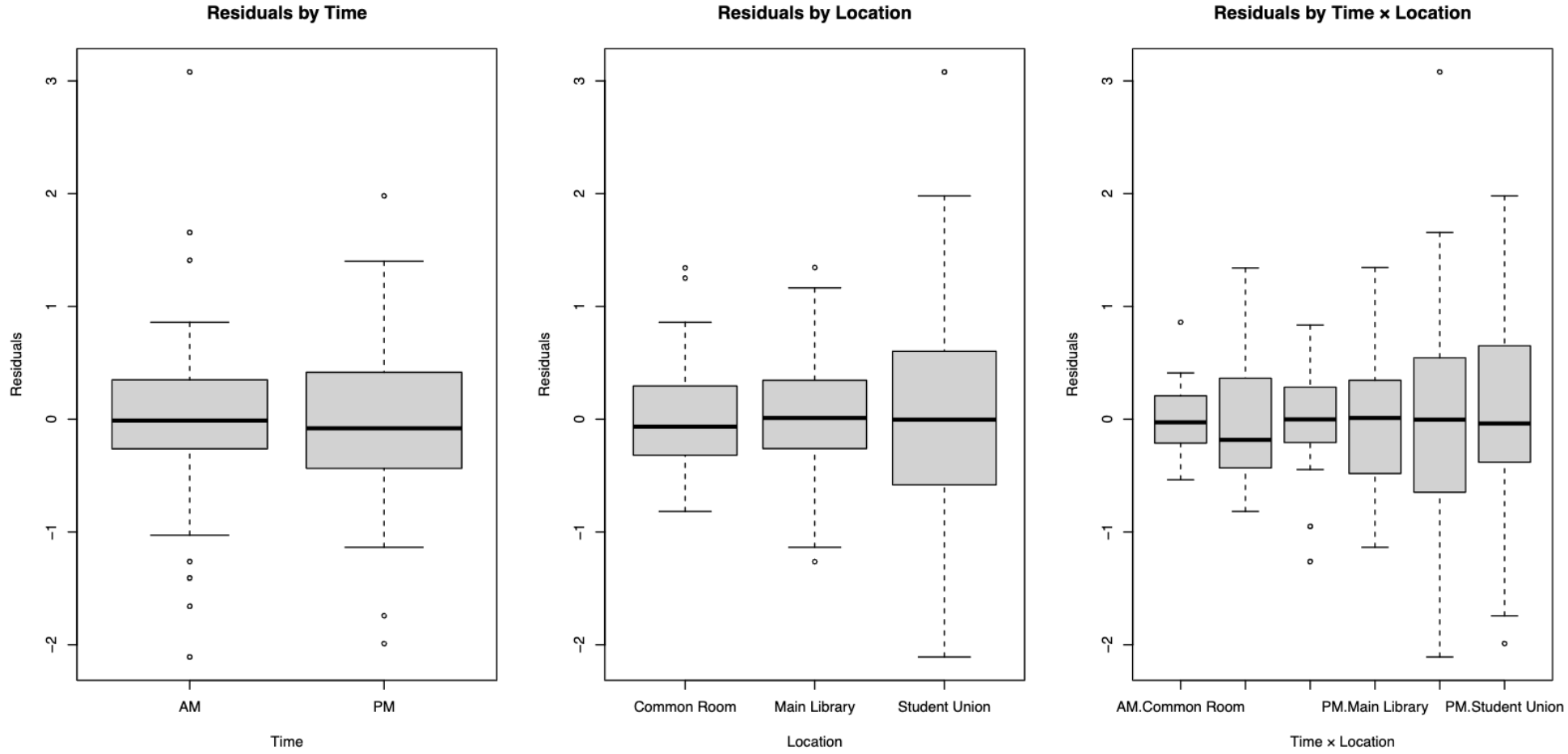
- ❑ Most points are wrapping well across the reference line.
- ❑ Some observations need attention.
- ❑ **The normality assumption is reasonably satisfied.**



- ❑ The residuals appear randomly scattered around the horizontal line at zero
- ❑ No clear pattern or shape is forming.
- ❑ **Constant variance assumption seems to be satisfied.**

Step 4: Model Assumption Check

Residual boxplots



- **Interpretation:** Both ‘AM × Student Union’ and ‘PM × Student Union’ show larger spreads and therefore may have more outliers.

Step 5: Tukey-adjusted confidence intervals

✓ Compute estimated marginal means (EMM) for the factor location from the reduced ANOVA model.

emm <- emmeans(score_model_red, ~ Location)

✓ Determine the pairwise comparisons between the levels of locations

```
summary(pairs(emm, adjust = "tukey"))
```

```
##  contrast                     estimate    SE  df t.ratio p.value
##  Common Room - Main Library      -1.79 0.151 135 -11.877  <.0001
##  Common Room - Student Union     -3.99 0.151 135 -26.480  <.0001
##  Main Library - Student Union    -2.20 0.151 135 -14.603  <.0001
##
## Results are averaged over the levels of: Time
## Degrees-of-freedom method: satterthwaite
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Interpretation:

❑ Common Room is significantly quieter (lower distraction) than the others

❑ Student Union is the loudest

Step 5: Tukey-adjusted confidence intervals

✓ Compute estimated marginal means (EMM) for the factor 'time' from the reduced ANOVA model

emm <- emmeans(score_model_red, ~ Time)

✓ Determine the pairwise comparisons between the levels of time

```
##  contrast estimate    SE  df t.ratio p.value
##  AM - PM    -0.606 0.123 135  -4.917  <.0001
##
## Results are averaged over the levels of: Location
## Degrees-of-freedom method: satterthwaite
```

Interpretation:

❑ AM is a quieter/better time to study on campus

Step 5: Tukey-adjusted confidence intervals

✓ Compute estimated marginal means (EMM) for the factor 'location: time' from the reduced ANOVA model.

emm <- emmeans(score_model_red, ~ Time:Location)

✓ Determine the pairwise comparisons between the levels of location and time

```
emm <- emmeans(score_model_red, ~ Time:Location)

## Cannot use mode = "kenward-roger" because *pbkrtest* package is not installed
summary(pairs(emm, adjust = "tukey"))

##  contrast                              estimate    SE  df t.ratio p.value
##  AM Common Room - PM Common Room         -0.428 0.213 135  -2.008  0.3432
##  AM Common Room - AM Main Library        -1.846 0.213 135  -8.654  <.0001
##  AM Common Room - PM Main Library        -2.165 0.213 135 -10.152  <.0001
##  AM Common Room - AM Student Union       -3.674 0.213 135 -17.223  <.0001
##  AM Common Room - PM Student Union       -4.742 0.213 135 -22.233  <.0001
##  PM Common Room - AM Main Library        -1.417 0.213 135  -6.645  <.0001
##  PM Common Room - PM Main Library        -1.737 0.213 135  -8.144  <.0001
##  PM Common Room - AM Student Union       -3.245 0.213 135 -15.215  <.0001
##  PM Common Room - PM Student Union       -4.314 0.213 135 -20.225  <.0001
##  AM Main Library - PM Main Library       -0.320 0.213 135  -1.498  0.6660
##  AM Main Library - AM Student Union      -1.828 0.213 135  -8.570  <.0001
##  AM Main Library - PM Student Union      -2.897 0.213 135 -13.580  <.0001
##  PM Main Library - AM Student Union      -1.508 0.213 135  -7.071  <.0001
##  PM Main Library - PM Student Union      -2.577 0.213 135 -12.082  <.0001
##  AM Student Union - PM Student Union     -1.069 0.213 135  -5.010  <.0001
##
## Degrees-of-freedom method: satterthwaite
```

Interpretation:

❑ AM Common Room is the quietest interaction combination (all point estimates are negative). Not significantly quieter than PM Common Room

Summary

Anticipated learning:

- ❑ Identified the quietest campus locations and times for effective studying
- ❑ Compared noise levels across the common room, student union, and main library.
- ❑ Determined optimal study conditions based on time and location

Challenges encountered:

- ❑ External factors (e.g., using a microphone/loudspeaker in different events, especially in the student union) influenced the objective rating
- ❑ Inconsistency in the noise level app installed on different smartphones
- ❑ Data collection on a specific date and time

Summary

Key findings:

- ❑ The common room in the morning (AM) is the quietest in all the study locations considered in the study
- ❑ The student union is the loudest area, less preferable for studying
- ❑ AM is generally quieter than PM across all locations

Future scope:

- ❑ Increase sample size and include more time slots
- ❑ Incorporate a qualitative survey for subjective ratings
- ❑ Control external factors by accounting for campus events
- ❑ Include noise level meters (decibel meters) instead of the mobile app.

Thank you