Final Project

PSTAT122: Design and Analysis of Experiments

Fall 2025

STUDENT NAME

- Yanxiu Jin (yanxiu_jin)
- Luke Maldonado (lukemaldonado)
- Jeff Loomis (jeffloomis)
- Erika Nares (erikanares)

Due Date

Due Date: Monday, December 8, 2025, 11:59 PM

1 Introduction

The purpose of this experiment is to investigate whether loud background music influences simple reaction time. Our research question is: **Do individuals respond faster or slower to a visual stimulus when loud music is playing compared to silence?**

Reaction time is widely used in psychology and human-performance research as an indicator of attention, processing speed, and readiness. Understanding how auditory stimulation influences reaction time is relevant for everyday activities such as driving, gaming, and working in noisy environments.

Prior research shows that music can influence cognitive processing, though the direction of the effect is not always consistent. High-arousal or cognitively demanding music has been shown to reduce performance on attention and memory tasks compared to silence (Cassidy and MacDonald 2007), and another study (Ünal et al. 2013) reported that background music increased cognitive load and reduced performance in a simulated driving task. These findings raise the possibility that loud music may interfere with even simple reaction-time performance.

Motivated by these prior results, we designed a controlled randomized complete block experiment in which each participant completed reaction-time trials under both a loud-music condition and a silence condition. This within-subject design isolates the effect of loud music while accounting for natural individual differences.

2 Experimental Design

2.1 Factors and treatment structure

Our experiment investigated whether loud background music affects simple reaction time. The experiment included one fixed factor, Condition, with two treatment levels: Loud Music and Silence. For the loud music condition, we chose "Flight of the Bumblebee" by Nikolai Rimsky-Korsakov because it is a high-intensity, complex piece that is cognitively demanding, providing a clear contrast to the silence condition. Each participant received both treatment levels.

2.2 Measurements and Units

The response variable was simple reaction time, measured in milliseconds (ms) using the Human Benchmark Reaction Time Test. The test recorded how quickly participants clicked when the color on their screen changed from red to green.

2.3 Fixed and Random Factors

Condition was treated as a fixed factor because its two levels (Loud Music vs. Silence) represent the specific treatments of interest. Participants were modeled as a random blocking factor. Because participants represent a random sample from a broader population, participants are not of direct inferential interest. Treating them as a random factor accounts for natural individual variability while allowing the estimated treatment effect to generalize at the population level.

2.4 Design Type

The design type for our experiment was a Randomized Complete Block Design (RCBD). This design was appropriate because each participant served as a block and received both treatment conditions, allowing for within-subject comparisons that reduces variability due to individual differences.

2.5 Implementation

Randomization was implemented by randomizing the order of treatments for each participant. Specifically through randomly assigning half of the participants to complete the experiment first with music followed by silence and the other half experienced the reverse, preventing systematic order effects.

For replication, each condition was measured across all participants. Additionally, each participant completed five experimental trials per treatment. Averaging out the results to produce one response per participant per condition helped reduce error and increase accuracy of the measured reaction times.

For blocking, participants formed blocks. This allowed us to control for variability between participants, which means Loud Music and Silence were compared within each individual rather than across different people.

The sample size for our experiment was 18. To determine this, we performed a power analysis using the effect size from a previous study (Farrell 2021) that analyzed simple reaction time under Loud Music and Silence conditions. Although our experimental conditions differed, the prior study provided a reference for estimating effect size. The power analysis indicated that at least 15 participants were needed to achieve 80% power for detecting a paired difference at $= 0.05$. Because our experiment was fairly straightforward to implement, we increased the sample size to 18 participants to improve accuracy in our final results.

```
library(pwr)

# Previous study components
n_study <- 20
mean_difference <- 36.63  # (Difference b|n 0dB and 74dB)
df_study <- n_study - 1
p_value_study <- 0.002

t_value_study <- qt( p_value_study/2 , df_study , lower.tail = FALSE )


# Calculating Standard Deviation
SD_D <- (mean_difference * sqrt(n_study)) / t_value_study
SD_D  # 45.76586
```

```
[1] 45.76586
```

```r
power_result <- pwr.t.test(d = mean_difference / SD_D,
                           power = 0.8,
                           sig.level = 0.05,
                           type = "paired",
                           alternative = "two.sided")
power_result # 14.29127
```

```
     Paired t test power calculation

              n = 14.29127
              d = 0.8003782
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*
```

3 Data Collection

3.1 Procedure

Our experiment which focuses on the relationship between human reaction time and loud background music was conducted usuing a variety of specifications. Firstly, the experiment was conducted in different locations using an online reaction time measurment website (https://humanbenchmark.com/tests/reactiontime) and a modern piano rendition of "Flight of the Bumblebee" from YouTube (https://youtu.be/M93qXQWaBdE?si=5r1ALkZ1AOeAU0mk). In addition, in order to reduce external variability, we all made sure to perform our experiments between the hours of 7PM to 9PM mostly on Thursday November 27, 2025 but some on Wednesday November 26, 2025. For further standardization, we made sure that participants completed the test on a mobile phone running Google Chrome, positioned approximately 40 cm (15.748 in) from the participant's face in a brightly lit room.

Within the experiment, we asked each participant to complete five warm-up trials without music to understand how the benchmark worked. By familiarizing participants with the interface, the trials can help reduce the common slow-start bias observed in reaction-time tasks. Warm-up data were not included in the analysis.

After the warm-up, participants completed ten experimental trials divided into two treatment conditions: Loud Music and Silence. Each treatment condition consisted of five trials, and the order of the treatment conditions was randomized for each participant (either Music → Silence or Silence → Music). During the Loud Music condition, participants listened to a loudly played modern piano rendition of Flight of the Bumblebee via YouTube. In the Silence condition, no audio was played.

To maintain consistent testing conditions, all ten trials were collected continuously in one session without breaks. For each participant, the five trials within each condition were later averaged to produce a single reaction-time measurement for Loud Music and for Silence to perform our statistics on.

3.2 Challenges/Adjustments

In the experiment, we did make it a point to cover as many sources of error variation as possible but given the nature of our project, we did run into some issues along the way. The most immediate issue was the date range in which the experiment was conducted. We conducted our experiment on the days of November 26 to November 27 which is the day before and the day of thanksgiving in the United States. With this, we were unable to meet in person, and each team member instead collected data independently from a few friends. This introduced unavoidable variation in measurement environments across participants.

One major challenge is device variability. Each of the team member used different mobile phones, which can differ in screen size, brightness, touch sensitivity, and input latency. Moreover, although we put the phone at approximately 40 cm from their face, this distance likely fluctuated naturally over the course of the session as participants adjusted their posture.

In addition, there were limitations in how well we could control the timing between trials. At the beginning of the experiment, participants were instructed that after completing one trial, they should immediately tap to start the next trial and continue in this manner until all ten trials were finished. This procedure ensured that the experiment ran in a single continuous session, but it also meant that the short intervals between trials were self-paced rather than strictly controlled. Some participants may have initiated the next trial almost instantly, while others may have taken slightly longer. Even within the same participant, the inter-trial intervals were not perfectly consistent across all the trials. These variations in timing could introduce additional variability in attention, readiness, or task engagement.

We also faced challenges with audio and environmental standardization. While the piece Flight of the Bumblebee was used consistently, the playback devices and setups varied: some experimenters used laptops, others relied on phone speakers, and the volume level and speaker quality were not perfectly matched across households. The relative position of the audio source to the participant also differed. Similarly, although we instructed that the room should be "bright," we each had different ideas of a "bright" room so we each had different light levels in the room.

3.3 Data Presentation

Table 1: Raw data after recording for Loud Music trials

Subject	music 1	music 2	music 3	music 4	music 5	Average
1	363	363	282	294	332	326.8
2	328	271	284	298	291	294.4
3	332	329	293	318	313	317.0
4	297	287	297	298	287	293.2
5	322	280	291	432	320	329.0
6	294	283	324	348	314	312.6
7	322	268	282	264	279	283.0
8	279	324	315	279	299	299.2
9	428	369	294	339	350	356.0
10	285	299	296	268	280	285.6
11	361	334	391	287	329	340.4
12	480	329	346	323	357	367.0
13	226	255	250	277	248	251.2
14	304	321	348	301	310	316.8
15	273	303	307	337	299	303.8
16	265	231	222	259	249	245.2
17	242	256	335	233	267	266.6
18	287	365	482	258	407	359.8

Table 2: Raw data after recording for Silence trials

Subject	silence 1	silence 2	silence 3	silence 4	silence 5	Average
1	279	307	296	367	307	311.2
2	278	270	412	356	165	296.2
3	313	281	292	337	302	305.0
4	252	292	289	303	280	283.2
5	304	301	288	316	295	300.8
6	288	349	324	312	322	319.0
7	290	244	288	330	303	291.0
8	313	296	272	315	294	298.0
9	270	240	221	240	245	243.2

Subject	silence 1	silence 2	silence 3	silence 4	silence 5	Average
10	325	291	325	281	301	304.6
11	351	272	349	309	375	331.2
12	308	278	301	297	403	317.4
13	268	241	277	244	265	259.0
14	289	316	295	261	281	288.4
15	339	301	310	336	394	336.0
16	260	268	245	219	247	247.8
17	242	285	283	361	296	293.4
18	253	326	304	240	296	283.8

```r
library(dplyr)

df_long <- read.csv("reaction_times_long.csv")

summary_stats <- df_long %>%
  group_by(treatment) %>%
  summarize(
    Mean = round(mean(rt), 2),
    SD   = round(sd(rt), 2),
    .groups = "drop"
  )

kable(
  summary_stats,
  caption = "Table 3: Summary of Mean and Standard Deviation by Condition"
)
```

Table 3: Table 3: Summary of Mean and Standard Deviation by Condition

treatment	Mean	SD
Music	308.20	48.93
Silence	294.96	40.63

4 Analysis

4.1 Exploratory Data Analysis

Participants' reaction times were summarized for each experimental condition to provide an initial overview of performance. Table 3 reports the mean and standard deviation of reaction times across the Loud Music and Silence conditions. On average, participants responded faster in the Silence condition ($M = 294.96$ ms) compared to the Loud Music condition ($M = 308.20$ ms). Variability was also somewhat higher in the Loud Music condition ($SD = 48.93$ ms) than in the Silence condition ($SD = 40.63$ ms).

Histogram of Reaction Times

To further examine the distributional characteristics of the reaction time data, we plotted histograms for each condition. These visualizations provide insight into the overall shape of the data and help identify potential deviations from normality. Across both the Silence and Loud Music conditions, the distributions appear fairly close to normal, with a smooth unimodal shape. While a few outliers are present, they do not appear numerous enough to substantially distort the overall distribution, supporting the use of RCBD ANOVA in the subsequent analysis.

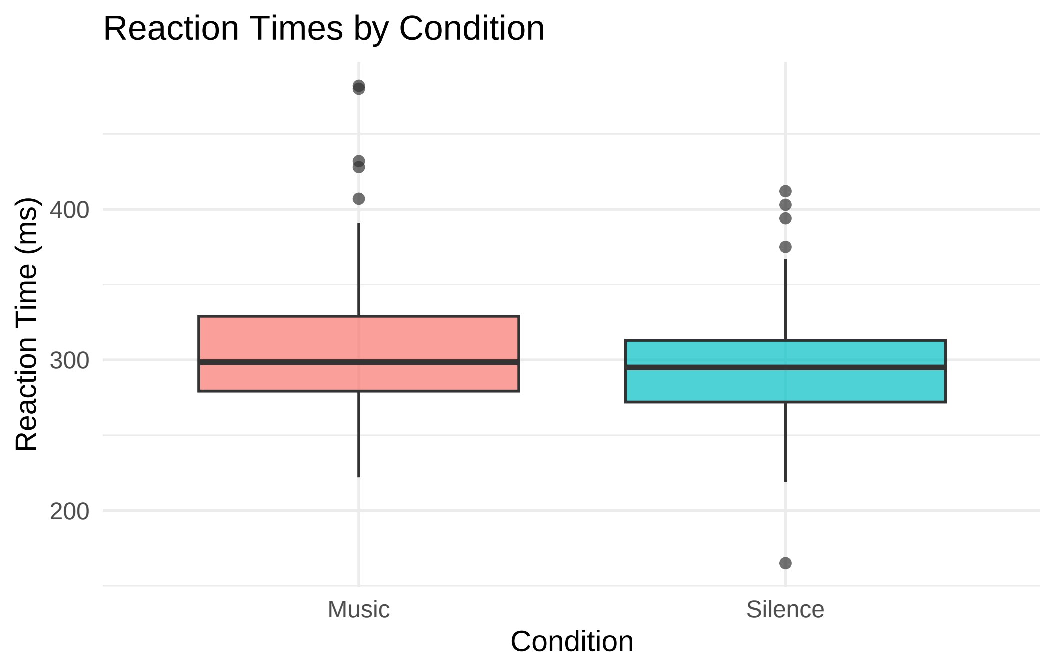
```r
library(ggplot2)

ggplot(df, aes(x = rt, fill = treatment)) +
  geom_histogram(alpha = 0.6, bins = 20, position = "identity") +
  facet_wrap(~ treatment) +
  labs(
    title = "Distribution of Reaction Times by Condition",
    x = "Reaction Time (ms)",
    y = "Count"
  ) +
  theme_minimal()
```

Boxplots of Reaction Times by Condition

The boxplots display the distribution of reaction times for each condition. The music condition shows a slightly wider spread in reaction times, and the median values for Loud Music and Silence appear similar. The Loud Music condition also includes a few higher outliers compared to the Silence condition.

```r
ggplot(df, aes(x = treatment, y = rt, fill = treatment)) +
geom_boxplot(alpha = 0.7) +
labs(
title = "Reaction Times by Condition",
x = "Condition",
y = "Reaction Time (ms)"
) +
theme_minimal() +
theme(legend.position = "none")
```

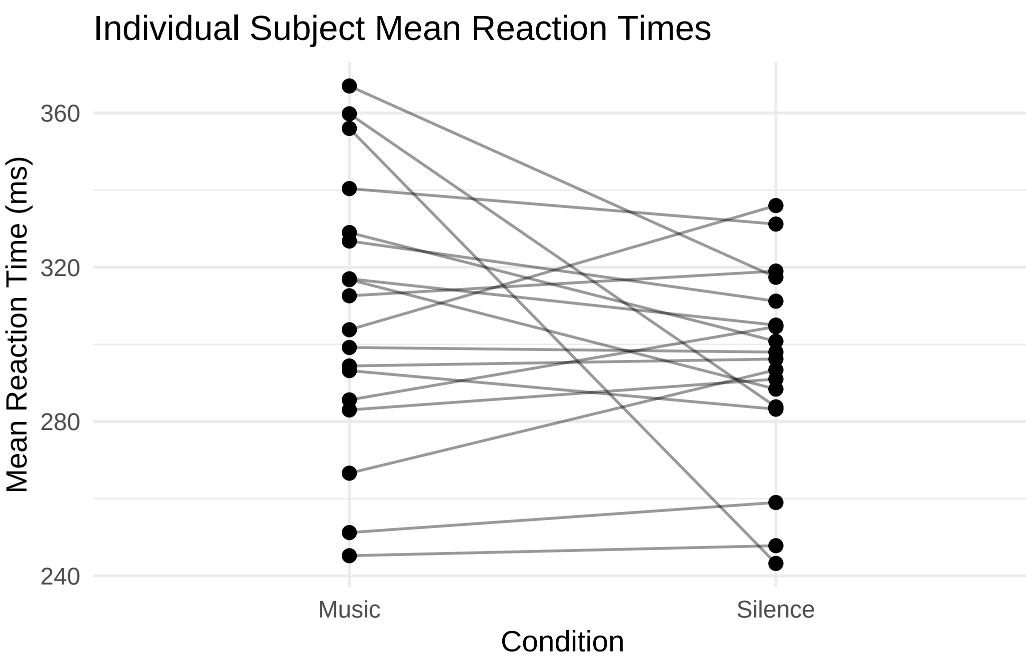


Subject-Level Mean Reaction Times

The subject-level mean plot displays one line per participant, connecting each individual's average reaction time in the Loud Music and Silence conditions. The figure shows a mixture of patterns: many subjects exhibit only slight changes between conditions, with lines that shift modestly upward or downward, while a few subjects display more pronounced differences. The overall collection of lines naturally produces some crisscrossing, and there is no visibly clear pattern.

```
# Compute means per subject per condition
subject_means <- df %>%
  group_by(Subject, treatment) %>%
  summarize(mean_rt = mean(rt), .groups = "drop")

# Line plot showing each subject's pattern
ggplot(subject_means, aes(x = treatment, y = mean_rt, group = Subject)) +
  geom_line(alpha = 0.4) +
  geom_point(size = 2) +
  labs(
    title = "Individual Subject Mean Reaction Times",
    x = "Condition",
    y = "Mean Reaction Time (ms)"
  ) +
  theme_minimal()
```



4.2 Hypothesis Testing

Following the exploratory analysis, we conducted a formal statistical test to determine whether reaction times differed between the Silence and Loud Music conditions. Because each participant completed both conditions, the study follows a randomized complete block design (RCBD) in which subjects serve as blocks. This approach accounts for participant-to-participant variability by treating individual differences as a block effect, thereby isolating the treatment effect of interest. The hypothesis for the RCBD analysis can be expressed as

$$H_0 : \tau_{\text{silence}} = \tau_{\text{music}} \text{ vs } H_A : \tau_{\text{silence}} \neq \tau_{\text{music}}$$

where τ_{silence} and τ_{music} represent the treatment effects associated with the Silence and Loud Music conditions, respectively. The RCBD ANOVA evaluates whether these treatment effects differ after accounting for subject-level blocking.

4.3 ANOVA Results

The ANOVA table is presented below.

```r
subject_means$Subject <- factor(subject_means$Subject)
fit <- aov(mean_rt ~ treatment + Subject, data = subject_means)
anova_table <- anova(fit)
kable(
  anova_table,
  digits = 4,
  caption = "ANOVA Table for RCBD (Subjects as Blocks)"
)
```

Table 4: ANOVA Table for RCBD (Subjects as Blocks)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	1	1578.738	1578.7378	2.4103	0.1390
Subject	17	20968.102	1233.4178	1.8831	0.1011
Residuals	17	11134.822	654.9895	NA	NA

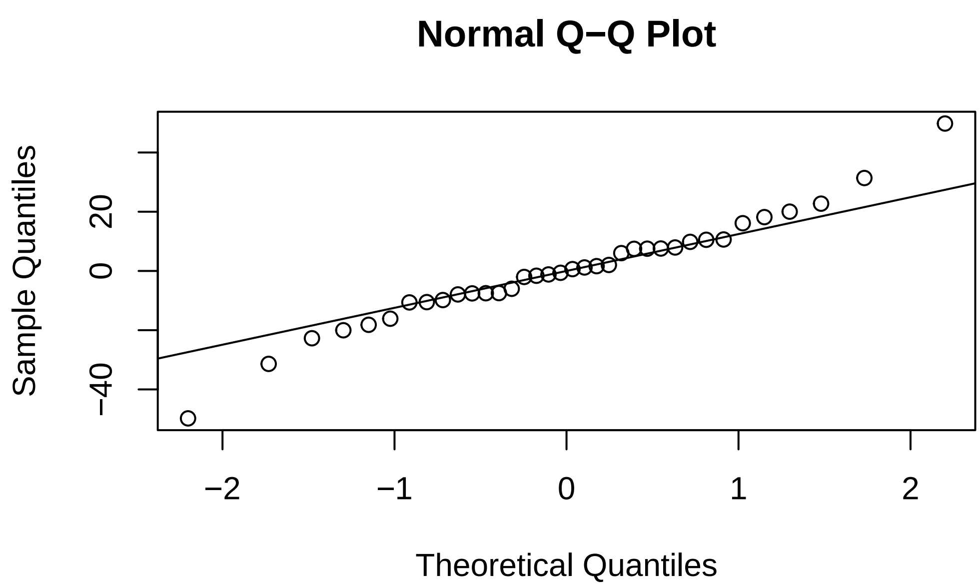
The RCBD ANOVA tested whether reaction times differed between the Silence and Loud Music conditions. The analysis showed that the effect of treatment was not statistically significant, $F(1, 33) = 2.41$, $p = 0.1390$. Because the result was not significant, we fail to reject the null hypothesis. Based on the averaged reaction times for each participant, there was no reliable evidence of a difference in reaction times between the Silence and Loud Music conditions.

Model Assumptions

After fitting the RCBD ANOVA model, we first examined the normal QQ plot of the residuals to assess whether the normality assumption was reasonably satisfied. In this case, the points follow the line fairly closely, suggesting that the residuals do not deviate strongly from normality.

```r
residuals_rt <- residuals(fit)

qqnorm(residuals_rt)
qqline(residuals_rt)
```



We also conducted the Shapiro–Wilk test to provide a numerical check of the normality assumption. The test yielded a statistic of $W = 0.9698$ with a p-value of 0.4203, and because the p-value exceeds the 0.05 threshold, we fail to reject the null hypothesis of normality. This result is consistent with the visual impression from the QQ plot.

```r
# Run Shapiro-Wilk test on ANOVA residuals
shap <- shapiro.test(residuals_rt)

# Convert to a tidy data frame for kable
shap_df <- data.frame(
  Statistic = shap$statistic,
  p_value   = shap$p.value
)

kable(
  shap_df,
  digits = 4,
  caption = "Shapiro-Wilk Normality Test for ANOVA Residuals"
)
```

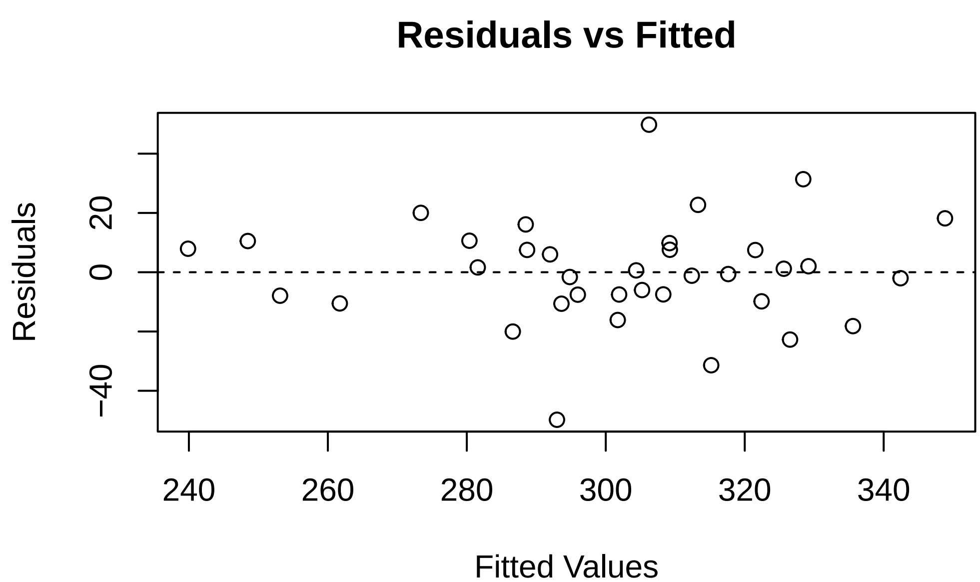
Table 5: Shapiro–Wilk Normality Test for ANOVA Residuals

	Statistic	p_value
W	0.9698	0.4203

To assess homoscedasticity, we examined the residuals versus fitted values plot. The residuals appear to be spread out fairly evenly across the range of fitted values, with no obvious funneling or systematic pattern. This suggests that the assumption of constant variance is reasonably satisfied for the RCBD ANOVA model.

```
fitted_rt <- fitted(fit)

plot(fitted_rt, residuals_rt,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted")
abline(h = 0, lty = 2)
```



Analysis summary

In summary, the exploratory analyses provided an overview of the reaction time distributions and subject-level patterns across the two conditions. A formal hypothesis test was then conducted using an RCBD ANOVA to evaluate whether reaction times differed between silence and music. The ANOVA indicated that the treatment effect was not statistically significant, and the assumption checks based on residual diagnostics showed no substantial departures from normality or homoscedasticity. Overall, the results do not provide evidence of a difference in reaction times between the two conditions.

Although participants serve as random blocks conceptually, the RCBD analysis in this project was conducted using a standard fixed-effects ANOVA model. In balanced RCBD designs, this approach yields the same treatment inference as a mixed-effects model with a random block effect. The following code proves the statisitical equivalence by giving Its t-statistic of -1.553 corresponds to an F-statistic of 2.41, exactly matching the treatment F-value obtained from the RCBD ANOVA table.

```
df_long <- read.csv("reaction_times_long.csv")

subject_means <- df_long %>%
  group_by(Subject, treatment) %>%
  summarize(mean_rt = mean(rt), .groups = "drop")
```

```r
subject_means$Subject <- factor(subject_means$Subject)
library(lme4)
m <- lmer(mean_rt ~ treatment + (1 | Subject), data = subject_means)
summary(m)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: mean_rt ~ treatment + (1 | Subject)
   Data: subject_means

REML criterion at convergence: 333.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9860 -0.5191  0.1226  0.4295  1.9040

Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 289.2    17.01
 Residual             655.0    25.59
Number of obs: 36, groups:  Subject, 18

Fixed effects:
                 Estimate Std. Error t value
(Intercept)       308.200      7.243  42.554
treatmentSilence  -13.244      8.531  -1.553

Correlation of Fixed Effects:
            (Intr)
tretmntSlnc -0.589
```

```r
(-1.553)^2
```

```
[1] 2.411809
```

5 Conclusions

The goal of this experiment was to determine whether loud background music affects simple reaction time compared to silence condition. Using an RCBD in which each participant completed both conditions, we found no statistically significant difference between Loud Music and Silence. Although the mean reaction times were slightly slower under loud music, this difference was not large enough to be meaningful. Assumption checks and a mixed-effects model confirmed the same conclusion.

Several limitations of the experiment may have introduced additional variability into the reaction-time measurements. Because data were collected individually, testing environments differed across participants. Device differences, small fluctuations in viewing distance, inconsistent inter-trial timing, and non-standardized audio and lighting conditions all added noise to the measurements. These uncontrolled factors may have diluted the true treatment effect, making it harder to detect. As a result, even though the sample size was informed by a power analysis, the actual effect size under our conditions may have been smaller than anticipated, reducing the effective statistical power of the study.

Future studies could benefit from a more controlled laboratory environment, standardized devices, controlled audio setting, consistent lighting, and automated timing between trials. Increasing the number of trials per condition or collecting more participants could also improve precision. Additionally, exploring other types of music (e.g., varying tempo, complexity, or genre) or including a low-volume condition may provide a more nuanced understanding of how auditory stimulation affects reaction-time performance.

References

Cassidy, Gianna, and Raymond AR MacDonald. 2007. "The Effect of Background Music and Background Noise on the Task Performance of Introverts and Extraverts." *Psychology of Music* 35 (3): 517–37.

Farrell, Jane. 2021. "The Effect of Increasing Music Volume on Reaction Time." *The Journal of Science and Medicine*.

Ünal, Ayça Berfu, Dick de Waard, Kai Epstude, and Linda Steg. 2013. "Driving with Music: Effects on Arousal and Performance." *Transportation Research Part F: Traffic Psychology and Behaviour* 21: 52–65.