Analyzing the Effects of Caffeine on Cognitive Performance*

Sinan Ma   Chengxin Li   Haoyu Wang   Kevin Luo
Zhekai Yu

2025-04-11

Caffeine is a natural stimulant found in coffee, tea, and some other plants. It is commonly used to boost reaction time and attention performance. However people who consume caffeine more often may not experience the same level of improvement as those who rarely consume caffeine. This study is to analyze the relationship between caffeine intake levels and Cognitive performance. To analyze this effect, we design a controlled experiment by using *The Island*. We randomly select 90 participants and assign them into three caffeine intake levels. Participants would be asked to complete a 10 minutes attention test and minutes mental arithmetic task. The results show that higher caffeine intake may lead to better attention performance, but caffeine does not significantly improve or impair arithmetic ability. In the future study, we may consider more independent factors that may affect test results, such as genetic factors or habitual caffeine intake history.

Table of contents

*We would like to thank Anna Ly for her invaluable guidance throughout this project. The source code and data can be found on our GitHub repository.

1 Introduction

Caffeine is one of the most widely used stimulants worldwide and is known for its potential to help people stay awake and affect various cognitive functions. While many people believe that moderate caffeine intake can improve mental sharpness and maintain concentration, the actual impact on different cognitive abilities remains controversial.

In this study, we examined the effects of three different levels of caffeine intake: decaffeinated (0 mg), regular (approximately 95 mg), and espresso (approximately 150 mg), on attention (measured by the number of missed letters) and mental arithmetic performance (quantified by test scores). We aim to study the following research questions:

- (RQ1) What are the impacts of caffeine intake on attention test performance?
  - Null hypothesis: caffeine intake has no association with attention test performance.
  - Alternative hypothesis: caffeine intake is significantly correlated with improved attention test performance.
- (RQ2) What are the impacts of caffeine intake on math (arithmetic) test score?
  - Null hypothesis: caffeine intake has no association with math test performance.
  - Alternative hypothesis: caffeine intake is significantly correlated with higher math test scores.

The structure of the paper is as follows: Section 2 describes the data collection process, experimental design, and statistical methods employed in this analysis. Section 3 explores the dataset through summary statistics, visualization, and model assumptions. Section 4 provides interpretations of the results, discuss the findings from the data and addresses our research questions. Section 5 discusses potential challenges encountered during the study and suggests areas for improvement. Finally, Section 6 summarizes the major insights and concludes our analysis.

2 Methodology

In the second-to-last week of February 2025, our group (Sinan, Chengxin, Haoyu, Kevin, and Zhekai) conducted a controlled experiment using the simulation platform *The Island* to investigate the effect of Caffeine Intake Level on attention performance and arithmetic ability.

In this experiment, We employed a between-subjects design using one-way ANOVA, where the experiment unit was an individual participant. The explanatory variable was *Caffeine Intake Level*, with three treatment groups: decaffeinated coffee (250 ml), which served as the control group; regular coffee (250 ml); and espresso (60 ml $\times$ 2). The response variables were *Attention Test Performance* and *Arithmetic Test Result*. Attention performance was measured by the number of missed letters in a 10-minute attention test, where a lower score indicated better performance. Arithmetic ability was assessed by the proportion of correct answers in a 4-minute mental arithmetic task, rounded to two decimal places.

In the simulation platform *The Island*, 90 participants were evenly distributed into three groups (n = 30 per group) to ensure a balanced design. Group assignment was conducted manually within the simulated environment by drawing participants from different islands and allocating them evenly across treatment groups.

To ensure a controlled comparison, all participants are 30 to 40 years old, university degree holders, non-smokers, and had not consumed alcohol in the past 24 hours. Before the test, each participant completed a 30-minute nap. After the nap, participants consumed their assigned caffeinated beverage and then waited for a 15-minute caffeine absorption period to control the variability. After the pre-test routine, participants completed both cognitive tasks. All outcomes were recorded as numeric values and were suitable for statistical analysis.

We did not implement blocking in this experiment, as all participants were homogeneous in terms of age, education level, and lifestyle factors. While our group attempted to randomize the assignment of participants by selecting from different islands and houses within the simulation platform, the randomization process was not fully formalized. Visual elements such as house size, color, or layout may have unconsciously influenced our manual selection, introducing potential selection bias. Despite these limitations, group sizes remained balanced and efforts were made to distribute participants evenly across conditions. The implications of this partial randomization are further discussed in section Section 5.

3 Analysis

3.1 Summary Statistics

This section presents the summary statistics for Attention Test Scores, Mental Arithmetic Test Scores, and Caffeine Treatment Groups. Table 1 and Table 2 display the mean, median, standard deviation, interquartile range, confidence intervals, minimum, and maximum. Table 3 shows the number of participants in each caffeine group.

3.1.1 Summary of Attention Test Scores

Table 1: Summary of Attention Test Scores

	Mean	Median	SD	IQR	CI Lower	CI Upper	Min	Max
Value	8.1	5	9.377297	9	6.135963	10.06404	0	42

The results indicate that the mean (8.1) is higher than the median (5), this suggests that the data is mostly concentrated on the right-side. This indicates the distribution is right-skewed. The interquartile range (IQR) of 9 is close to the standard deviation (9.38), indicating a wide spread of the data.

3.1.2 Summary of Mental Arithmetic Test Scores

Table 2: Summary of Mental Arithmetic Test Scores

	Mean	Median	SD	IQR	CI Lower	CI Upper	Min	Max
Value	0.6527778	0.68	0.2350187	0.4	0.6035541	0.7020015	0.1	1

The mean (0.65) and median (0.68) for the Mental Arithmetic Test Scores are relatively close. The standard deviation (0.235) may indicates the variability is relatively low. We may examine the box plots for further information.

3.1.3 Summary of Caffeine Treatment Groups

Table 3: Table 3: Summary of Participants in Each Treatment Group

	Decaffeinated	Regular	Espresso x2
Count	30.0	30.0	30.0
Percentage	33.3	33.3	33.3

The table confirms that each caffeine intake group (Decaffeinated, Regular Coffee, and two shots of Espresso) contains 30 participants, ensuring a balanced design.

3.2 Visualization

The following visualizations illustrate the distribution of test performance across caffeine treatment groups. Figure 1 illustrates the distribution of missed letters in the attention test, while Figure 2 visualizes performance on the mental arithmetic test.

3.2.1 Boxplot for Attention Test Scores

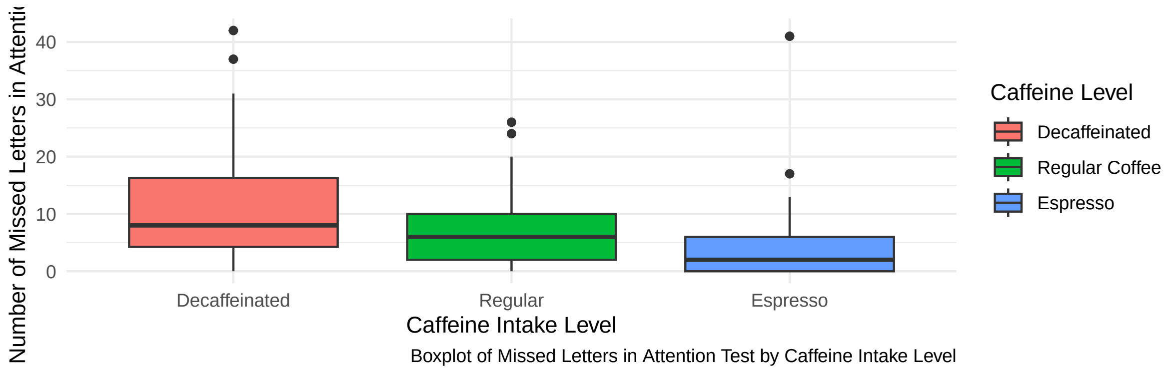

Figure 1: Boxplot of Missed Letters in Attention Test by Caffeine Intake Level

As shown in Figure 1, the box plot above visualizes how caffeine intake affects the participants' concentration of awareness. The median of the number of missed letters decreases as caffeine intake increases. The participants within the decaffeinated group had overall, a higher number of missed letters in the attention test. Furthermore, one may observe that the variability decreases as caffeine intake increases, indicated by the length of the boxes. The overall vertical placement of the boxes confirms our analysis of the summary table for the attention test result above, the distribution of the results are indeed right-skewed.

3.2.2 Boxplot for Mental Arithmetic Test Scores

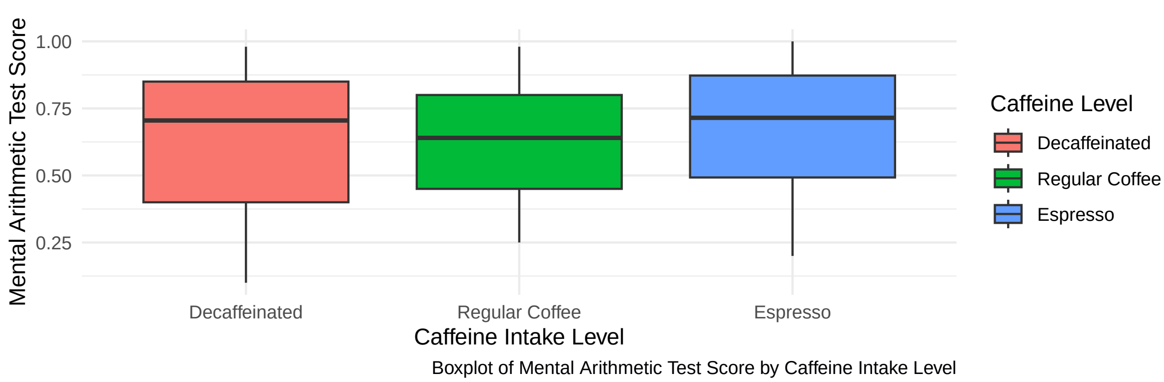

Figure 2

As shown in Figure 2, the box plot above visualizes how caffeine intake affects the participants' ability to perform mental arithmetic tests. There are no clear trends in how the median changes as caffeine intake increased, indicating that caffeine intake did not have a strong effect on mental arithmetic performance. Evidenced by the length of the boxes, the decaffeinated group had a slightly higher variability within their results. Overall, the spread of scores is not identical between the groups, yet there does not exist extreme outliers. These results show that the participants' mental arithmetic ability was not significantly influenced by caffeine intake.

3.3 Model Assumptions

This section evaluates whether the assumptions required for ANOVA are satisfied for both the attention test and arithmetic test models. We use Q-Q plots (Figure 3 and Figure 4) and Shapiro-Wilk tests (Table 4 and Table 8) to assess normality of residuals. Bartlett tests (Table 5 and Table 9) are used to examine homogeneity of variances. When these assumptions are violated, Kruskal-Wallis tests (Table 6 and Table 10) are employed as non-parametric alternatives, followed by pairwise comparisons (Table 7) when necessary.

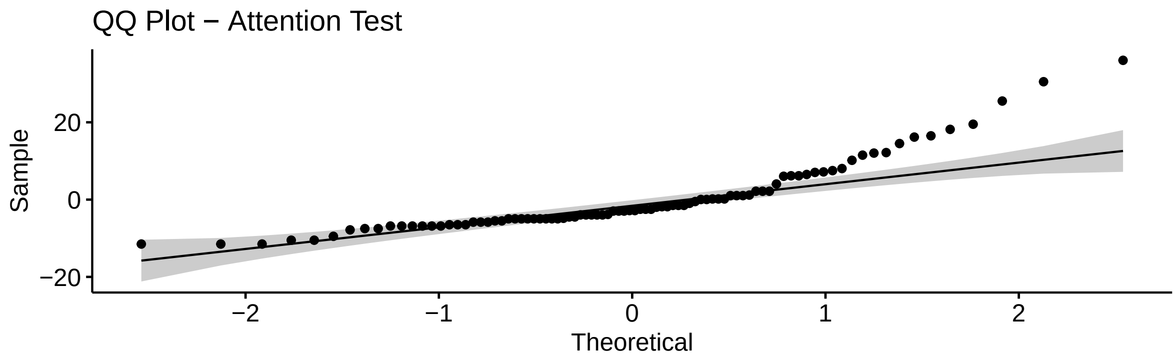

Figure 3

As shown in Figure 3, the Q-Q plot helps assess the normality of residuals from the attention test model. Evidenced by the plot, the data points on the top right deviate significantly from the straight line, suggesting that the data is right-skewed and the normality assumption may not be satisfied.

Table 4: Shapiro-Wilk Test

	Statistic	p.value
W	0.834	1.19e-08

Table 5: Bartlett Test

	Statistic	p.value
Bartlett's K-squared	6.682	0.0354

As shown in Table 4, the Shapiro-Wilk test yields a p-value approximately equal to zero, leading us to reject the null hypothesis and conclude that the residuals do not follow a normal distribution. In addition, the Bartlett's test result in Table 5 gives a p-value of 0.0354. Since this is below the 0.05 significance level, we reject the null hypothesis and conclude that the variances across groups are not equal.

Table 6: Kruskal-Wallis Test for Attention Test Scores

	ChiSq	df	p.value
Kruskal-Wallis chi-squared	10.401	2	0.00552

Due to the tests conducted earlier in this report, the results of the attention test violate both the assumption of normality and the assumption of homogeneity of variance. Therefore, it is more appropriate to use a non-parametric method. As shown in Table 6, the Kruskal-Wallis test was performed to determine if there are significant differences among group means. The test provides a p-value of approximately 0.0055. At the 0.05 significance level, we reject the null hypothesis and conclude that significant differences exist between the groups, justifying the need for post-hoc testing.

Table 7: Pairwise Comparisons with Bonferroni Adjustment

Group 1	Group 2	Adj. p-value
Regular Coffee	Decaffeinated	0.3651
Espresso	Decaffeinated	0.0197
Espresso	Regular Coffee	0.6751

As a follow-up to the Kruskal-Wallis test, we conducted pairwise t-tests with Bonferroni adjustment to correct for Type I error inflation. The results in Table 7 show that the only statistically significant comparison is between the decaffeinated group and the espresso group, with an adjusted p-value of 0.02. All other comparisons resulted in p-values greater than 0.05. Therefore, we conclude that the significant difference lies specifically between these two groups.

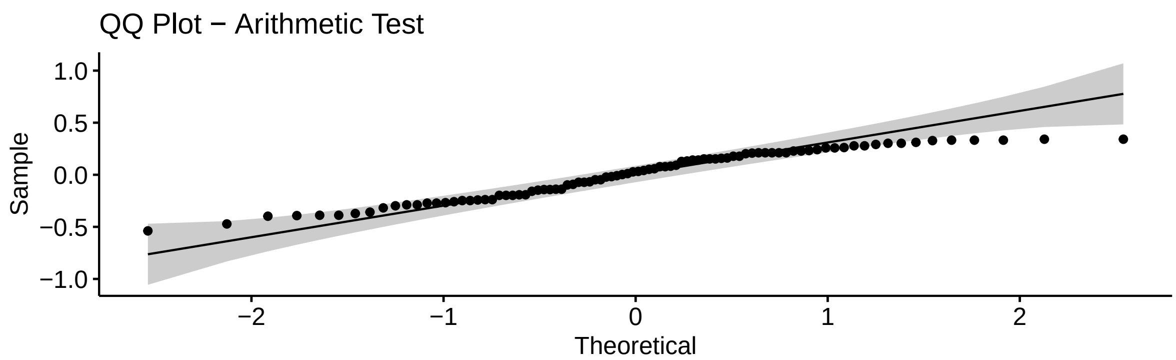

Figure 4: Q-Q Plot of Residuals from Mental Arithmetic Test Model

The Q-Q plot in Figure 4 provides a visual assessment of normality for the residuals from the mental arithmetic test model. Data points on both the right and left ends deviate from the straight line, suggesting that the assumption of normality may not be satisfied. We will confirm this using the Shapiro-Wilk test below.

Table 8: Shapiro-Wilk Test for Normality (Arithmetic Test)

	Statistic	p.value
W	0.947	0.00108

The Shapiro-Wilk test provides a p-value of 0.00108. the null hypothesis is rejected using the significance level of 0.05, the conclusion that the normality assumption is not satisfied can be made.

Table 9: Bartlett Test for Homogeneity of Variance (Arithmetic Test)

	Statistic	p.value
Bartlett's K-squared	0.921	0.631

Bartlett's test for homogeneity of variances provides a p-value of 0.631. The null hypothesis is failed to be rejected using a significance level of 0.05. The conclusion that there exists homogeneity of variances can be made.

Table 10: Kruskal-Wallis Test for Arithmetic Test Scores

	ChiSq	df	p.value
Kruskal-Wallis chi-squared	0.296	2	0.862

As shown in Table 10, due to the tests conducted earlier in this report, the results of the mental arithmetic test violate the assumption of normality. Therefore, it would be more appropriate to perform the Kruskal-Wallis test to verify if there exists any significant differences within the means of each group.

The results of the Kruskal-Wallis test provide a p-value of 0.8623, which is higher than the significance level of 0.05. It can be concluded that there does not exist any significant differences within the mean values of each group, and a follow-up post-hoc test is not necessary.

4 Discussion

The participants' results for both the attention test and the mental arithmetic test both violated the assumption of normality. Therefore, the Kruskal-Waliis test was deployed to verify if there existed any significant differences between the means of the results for each of the three groups of caffeine intake level. The test result for the mental arithmetic test indicates that caffeine intake level had no significant impact on the participants' ability to perform mental arithmetic. However, the Kruskal-Wallis test indicated that there did exist a significant difference between the means of the attention test for the three groups.

Upon further verification using pairwise comparisons using t-tests and the bonferroni adjustment method, there was a significant difference between the mean values of the decaffeinated group and the espresso. However, both the comparison between the regular coffee and decaffeinated coffee groups and the comparison between the regular coffee and espresso groups was shown to be statistically insignificant.

It can be concluded that only high levels on caffeine intake will induce a significant impact on a participant's attention levels. Furthermore, caffeine intake has no significant impact on the participant's ability to perform mental arithmetic.

5 Limitations

Our experiment used *The Island* tool to simulate 90 participants and assigned each person to one of three caffeine intake levels, with each group containing 30 participants. While we intended to randomize group assignment by manually selecting participants from different islands and houses, this process did not constitute formal randomization. Visual cues such as house size or color may have unintentionally influenced our selections, introducing potential allocation bias.

In the attention test, participants in the Espresso group missed fewer letters on average, suggesting a possible relationship between caffeine intake and improved attention. However, this observed effect may be confounded by unmeasured variables such as sleep quality the night before or mental stress due to personal circumstances. We attempted to reduce such variability by enforcing a 30-minute nap for all participants prior to the test.

Another limitation is that our cognitive tests were relatively short: a 10-minute attention task and a 4-minute mental arithmetic test. These simplified assessments may not fully capture the complexity of real-world tasks, which often require sustained attention and cognitive effort over longer periods. The ability to focus for 10 minutes does not necessarily indicate the capacity for long-term concentration in everyday settings.

For the arithmetic test, we failed to reject the null hypothesis, suggesting no statistically significant relationship between caffeine intake and mental arithmetic performance in this context.

In future studies, we recommend incorporating more rigorous randomization procedures (e.g., automated random assignment), implementing blocking strategies to control for additional variables, and applying double-blind procedures to minimize experimenter and participant bias. Additionally, designing tasks that more realistically reflect attention and arithmetic capabilities in real-life contexts would improve the ecological validity of the findings.

6 Conclusion

In summary, we investigated the effects of caffeine on cognitive performance and came to two main conclusions. First, there were significant differences between the three caffeine groups in the attention test, but only the comparison between the decaffeinated and espresso conditions reached statistical significance. This suggests that consuming a higher dose of caffeine (espresso) can significantly reduce the number of letters missed, while consuming a moderate dose of caffeine (regular coffee) does not have a significant advantage, which may indicate a threshold effect or individual differences in caffeine metabolism.

Second, regarding the mental arithmetic test, we found no significant differences between the three treatments. This result suggests that caffeine may not significantly alter more complex or computationally intensive tasks in our experimental design, or that any potential effect is too subtle to detect with the current sample size and measurement methods.

Based on these findings, future research could explore a wider range of cognitive functions and refine the range of caffeine doses. One particularly useful approach would be a repeated measures design, in which each participant receives multiple caffeine conditions. This design controls for individual differences, such as varying levels of caffeine tolerance, by comparing each participant's performance across treatment conditions. It allows for a more precise assessment of the direct effects of caffeine and helps clarify the relationship between dose and response.

Overall, our findings highlight that caffeine can improve some types of attentional performance, but its effects on more complex reasoning tasks are uncertain or small. Therefore, researchers and consumers should consider specific cognitive goals, whether it is improving short-term alertness or solving complex problems, when deciding on the optimal dose or usage scenario for caffeine.

7 Appendix

Loading Libraries

```
# Loading the required R package
library(tidyverse)
library(performance)
library(knitr)
library(ggpubr)
library(ggplot2)
library(car)
library(rstatix)
```

Loading Data

```
# Reading the cleaned data
data <- read.csv("data/01-clean_data_in_num.csv")
```

Summary of Attention Test Scores

```
# Summarizing the attention test scores
m1 = mean(data$num_of_missed_in_attention_test)
m2 = median(data$num_of_missed_in_attention_test)
m3 = sd(data$num_of_missed_in_attention_test)
m4 = IQR(data$num_of_missed_in_attention_test)

# Calculating a 95% confidence interval for mean
ci1 <- t.test(data$num_of_missed_in_attention_test)$conf.int

# Recording the minimum/maximum values
min_1 <- min(data$num_of_missed_in_attention_test)
max_1 <- max(data$num_of_missed_in_attention_test)

# Creating a summary table
attention_summary <- data.frame(Value = c(m1, m2, m3, m4, ci1[1], ci1[2],
                                          min_1, max_1))

kable(
  t(attention_summary),
  col.names = c("Mean", "Median", "SD", "IQR", "CI Lower", "CI Upper",
                "Min", "Max"),
```

```r
  caption = "Summary of Attention Test Scores",
  align = "c"
)
```

Summary of Mental Arithmetic Test Scores

```r
# Summarizing the mental arithmetic test scores
m5 <- mean(data$difficult_arithmetic_test_score, na.rm = TRUE)
m6 <- median(data$difficult_arithmetic_test_score, na.rm = TRUE)
m7 <- sd(data$difficult_arithmetic_test_score, na.rm = TRUE)
m8 <- IQR(data$difficult_arithmetic_test_score, na.rm = TRUE)

# Calculating a 95% confidence interval for mean
ci2 <- t.test(data$difficult_arithmetic_test_score, na.rm = TRUE)$conf.int

# Recording the minimum/maximum values
min_2 <- min(data$difficult_arithmetic_test_score, na.rm = TRUE)
max_2 <- max(data$difficult_arithmetic_test_score, na.rm = TRUE)

# Creating a summary table
math_summary <- data.frame(Value = c(m5, m6, m7, m8,
                                     ci2[1], ci2[2], min_2, max_2))

kable(
  t(math_summary),
  col.names = c("Mean", "Median", "SD", "IQR", "CI Lower", "CI Upper",
                "Min", "Max"),
  caption = "Summary of Mental Arithmetic Test Scores",
  align = "c",
  booktabs = TRUE
)
```

Summary of Caffeine Treatment Groups

```r
# Calculating the distribution of three caffeine intake groups
treatment_counts <- table(data$caffeine_intake_level)
treatment_percent <- prop.table(treatment_counts) * 100

# Creating a summary table
treatment_summary <- rbind(
  Count = round(as.numeric(treatment_counts), 1),
  Percentage = round(as.numeric(treatment_percent), 1)
```

```
)
colnames(treatment_summary) <- names(treatment_counts)

kable(treatment_summary,
      col.names = c("Decaffeinated", "Regular", "Espresso x2" ),
      caption = "Table 3: Summary of Participants in Each Treatment Group")
```

Boxplot for Attention Test Scores

```
# Creating a boxplot
ggplot(data, aes(
  x = as.factor(caffeine_intake_level),
  y = num_of_missed_in_attention_test,
  fill = as.factor(caffeine_intake_level)
)) +
  geom_boxplot() +
  scale_x_discrete(labels = c(
    "0" = "Decaffeinated",
    "1" = "Regular",
    "2" = "Espresso"
  )) +
  scale_fill_discrete(
    name = "Caffeine Level",
    labels = c(
      "0" = "Decaffeinated",
      "1" = "Regular Coffee",
      "2" = "Espresso"
    )
  ) +
  labs(
    x = "Caffeine Intake Level",
    y = "Number of Missed Letters in Attention Test",
    caption = "Boxplot of Missed Letters in Attention Test by
    Caffeine Intake Level"
  ) +
  theme_minimal()
```

Boxplot for Mental Arithmetic Test Scores

```
# Creating a box plot
ggplot(data, aes(
```

```r
  x = as.factor(caffeine_intake_level),
  y = difficult_arithmetic_test_score,
  fill = as.factor(caffeine_intake_level)
)) +
  geom_boxplot() +
  scale_x_discrete(labels = c(
    "0" = "Decaffeinated",
    "1" = "Regular Coffee",
    "2" = "Espresso"
  )) +
  scale_fill_discrete(
    name = "Caffeine Level",
    labels = c(
      "0" = "Decaffeinated",
      "1" = "Regular Coffee",
      "2" = "Espresso"
    )
  ) +
  labs(
    x = "Caffeine Intake Level",
    y = "Mental Arithmetic Test Score",
    caption = "Boxplot of Mental Arithmetic Test Score
    by Caffeine Intake Level"
  ) +
  theme_minimal()
```

Model Assumptions - QQ plot for Attention Test

```r
# Checking the Attention Test Model Assumptions
# Ensure caffeine intake level is a factor
data$caffeine_intake_level <- as.factor(data$caffeine_intake_level)

# Fit the ANOVA model for attention test
model_att <- lm(num_of_missed_in_attention_test ~ caffeine_intake_level,
                data = data)

# QQ Plot for normality check
ggqqplot(residuals(model_att), main = "QQ Plot - Attention Test")
```

Shapiro-Wilk test for Attention Test

```r
# Run tests
shapiro_test <- shapiro.test(residuals(model_att))

# Create compact results display
shapiro_results <- data.frame(
  Statistic = formatC(shapiro_test$statistic, format = "f", digits = 3),
  p.value = format.pval(shapiro_test$p.value, digits = 3)
)
knitr::kable(shapiro_results, align = 'c', caption = "Shapiro-Wilk Test")
```

Bartlett's test for Attention Test

```r
# Run tests
bartlett_test <- bartlett.test(num_of_missed_in_attention_test ~
                                 caffeine_intake_level, data = data)

# Create compact results display
bartlett_results <- data.frame(
  Statistic = formatC(bartlett_test$statistic, format = "f", digits = 3),
  p.value = format.pval(bartlett_test$p.value, digits = 3)
)

knitr::kable(bartlett_results, align = 'c', caption = "Bartlett Test")
```

Kruskal-Wallis Test for Attention Test

```r
# Convert caffeine intake level to words
data$caffeine_intake_level <- factor(data$caffeine_intake_level,
                                     levels = c(0, 1, 2),
                                     labels = c("Decaffeinated",
                                                "Regular Coffee", "Espresso"))

# Run Kruskal-Wallis test
kw_test <- kruskal.test(num_of_missed_in_attention_test ~
                          caffeine_intake_level, data = data)

# Format Kruskal-Wallis results
kw_df <- data.frame(
  ChiSq = formatC(kw_test$statistic, digits = 3, format = "f"),
  df = kw_test$parameter,
  p.value = format.pval(kw_test$p.value, digits = 3)
```

```r
)

knitr::kable(kw_df)
```

Pairwise t-test for Attention Test

Group 1	Group 2	Adj. p-value
Regular Coffee	Decaffeinated	0.3651
Espresso	Decaffeinated	0.0197
Espresso	Regular Coffee	0.6751

Model Assumptions - QQ plot for Arithmetic Test

```r
# Checking the Mental Arithmetic Model Assumptions
# Ensure caffeine intake level is a factor
data$caffeine_intake_level <- as.factor(data$caffeine_intake_level)

# Fit the ANOVA model for arithmetic test
model_math <- lm(difficult_arithmetic_test_score ~ caffeine_intake_level,
                 data = data)

# QQ Plot for normality check
ggqqplot(residuals(model_math), main = "QQ Plot - Arithmetic Test")
```

Shapiro-Wilk Test for Arithmetic Test

	Statistic	p.value
W	0.947	0.00108

Kruskal-Wallis Test for Arithmetic Test

```r
# Run Kruskal-Wallis test to check if there are significant differences
kw_test <- kruskal.test(difficult_arithmetic_test_score ~
                          caffeine_intake_level, data = data)

# Create display
kw_df <- data.frame(
  ChiSq = formatC(kw_test$statistic, digits = 3, format = "f"),
  df = kw_test$parameter,
```

```r
  p.value = format.pval(kw_test$p.value, digits = 3)
)

knitr::kable(kw_df, align = 'c')
```