Exploratory Data Analysis of Caffeine and Cognitive Performance

Sinan Ma CHengxin Li Haoyu Wang Kevin Luo
Zhekai Yu

2025-02-27

Summary Statistics

This section shows the summary statistics for Attention Test Scores, Mental Arithmetic Test Scores, and Caffeine Treatment Groups. The first two tables give the mean, median, standard deviation, and interquartile range. The third table shows the number of participants in each caffeine group.

Summary of Attention Test Scores

Table 1: Table 1: Summary of Attention Test Scores

	Mean	Median	SD	IQR
Value	8.1	5	9.377297	9

In the Attention Test, a higher number of missed letters means lower attention. The results show that the mean (8.1) is higher than the median (5), indicating that some participants had much higher missed letter counts, which skews the data. The IQR (9) is close to the standard deviation (9.38), indicating a relatively spread-out distribution.

Summary of Mental Arithmetic Test Scores

Table 2: Table 2: Summary of Mental Arithmetic Test Scores

	Mean	Median	SD	IQR
Value	0.6527778	0.68	0.2350187	0.4

In the Mental Arithmetic Test, a higher score means better performance. The results show that the mean (0.65) and median (0.68) are very close, indicating that most participants performed similarly. The IQR (0.4) is larger than the standard deviation (0.24), indicating that the middle 50% of scores are more spread out compared to the overall variation.

Summary of Caffeine Treatment Groups

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

	0	1	2
Count	30.00000	30.00000	30.00000
Percentage	33.33333	33.33333	33.33333

The treatment summary table confirms that the Decaffeinated, Regular Coffee, and Double Espresso groups each have 30 participants, making up 33.3% of the total sample. This ensures the groups evenly distributed.

Visualization

Boxplot for Attention Test Scores

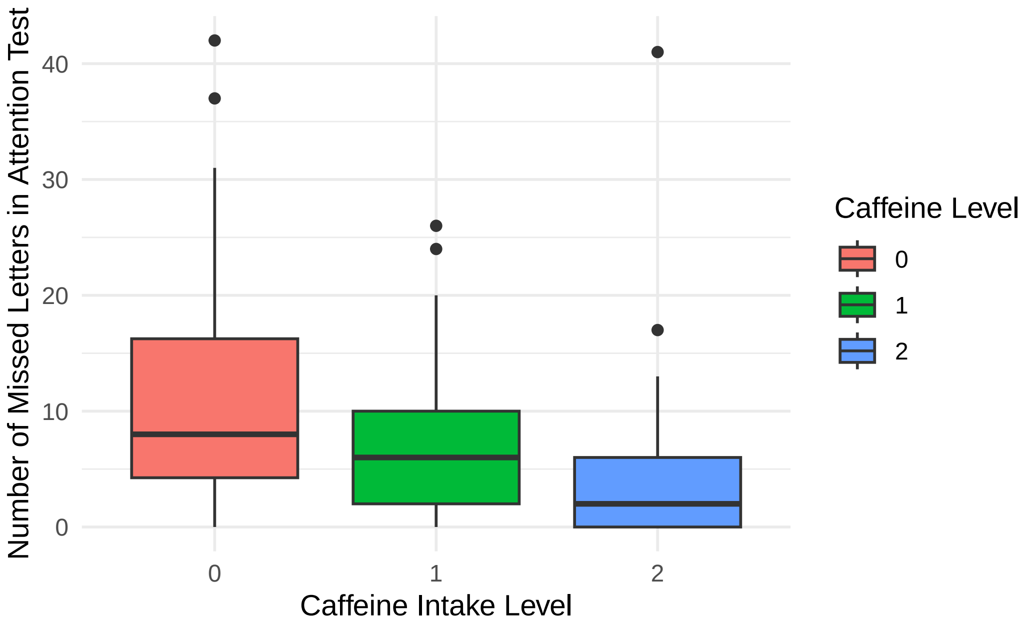


Figure 4: Boxplot of Missed Letters in Attention Test by Caffeine Intake Level

This boxplot visualizes how caffeine intake affects attention. The median of missed letters goes down when caffeine intake goes up. Participants in the decaffeinated group missed the most letters, while those in the espresso group missed the least.It indicates that higher caffeine intake leads to better attention performance.

The spread of data (IQR) is widest in the decaffeinated group, meaning their performance was less consistent.The regular and espresso groups were more stable.

Outliers appears in all groups, but the decaffeinated group has the most extreme ones, showing that some participants performed much worse than others.

Boxplot for Mental Arithmetic Test Scores

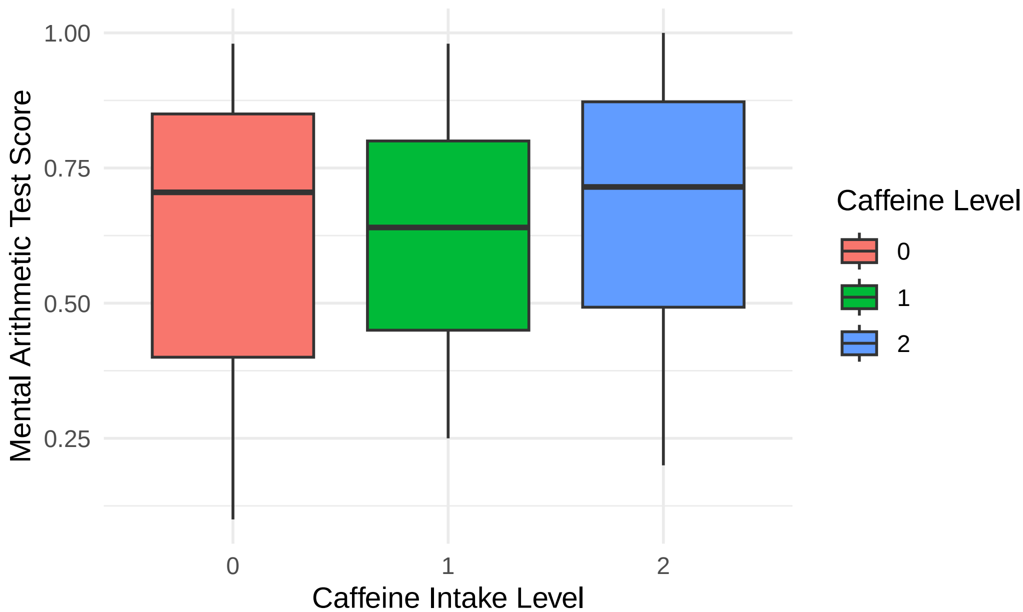


Figure 5: Boxplot of Mental Arithmetic Test Score by Caffeine Intake Level

This boxplot shows that caffeine intake level does not have a strong effect on mental arithmetic performance. The median scores across the decaffeinated (0), regular coffee (1), and espresso (2) groups are similar, suggesting that caffeine does not significantly improve or impair arithmetic ability.

The interquartile ranges (IQRs) and whiskers are comparable across groups, indicating that the variability in scores is consistent. There are no extreme outliers, showing that participants' performance was stable across different caffeine levels. This shows that: unlike attention performance, mental arithmetic ability remains unaffected by caffeine intake.

ANOVA Analysis

```
                      Df Sum Sq Mean Sq F value  Pr(>F)   
caffeine_intake_level  1    640   640.3   7.841 0.00628 **
Residuals             88   7186    81.7                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                      Df Sum Sq Mean Sq F value Pr(>F)
caffeine_intake_level  1  0.017 0.01667   0.299  0.586
Residuals             88  4.899 0.05567               
```

The ANOVA results show that caffeine intake level significantly affects attention performance ($p = 0.00628$). The F-value (7.841) suggests that differences between caffeine groups are not due to random chance. Participants who consumed more caffeine performed better on the Attention Test, missing fewer letters.

For the Mental Arithmetic Test, the ANOVA results do not show a significant effect ($p = 0.586$). The F-value (0.299) is low, meaning caffeine intake does not strongly impact arithmetic performance. Scores are similar across caffeine levels.

These results indicate that caffeine improves attention but does not affect arithmetic ability.

Attention Test Visualizations

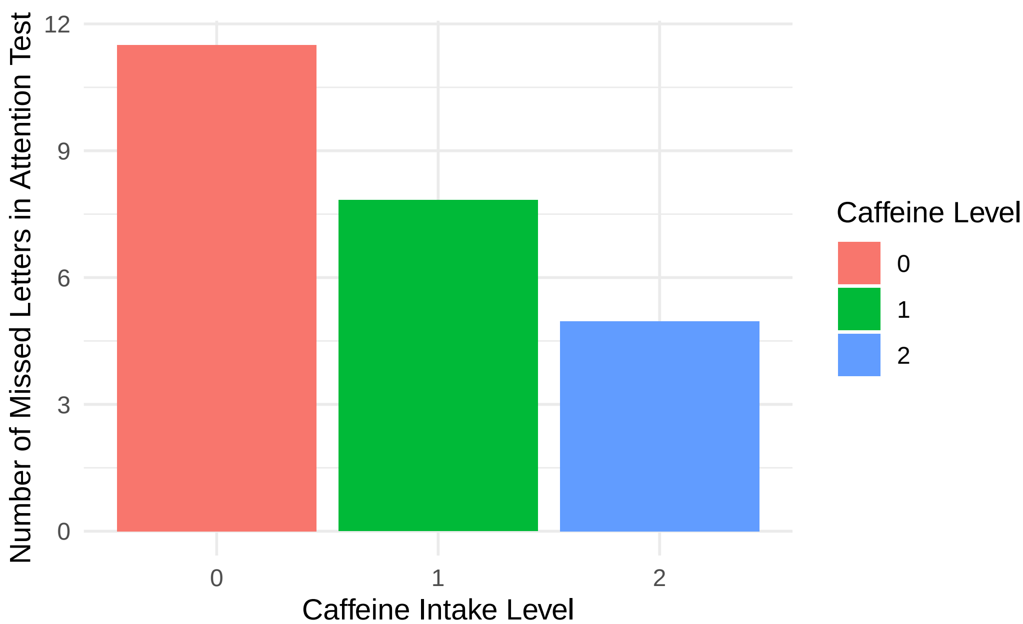


Figure 1: Missed Letters in Attention Test by Caffeine Intake Level

The bar chart visualizes the number of missed letters in the Attention Test for different caffeine intake levels (Decaffeinated = 0, Regular Coffee = 1, Espresso = 2). The decaffeinated group missed the most letters. The regular coffee group missed fewer. The espresso group missed the least.

The pattern shows that higher caffeine intake leads to better attention performance. The difference between groups is clear. The decaffeinated group had the worst performance, while the espresso group had the best. This result supports the idea that caffeine helps maintain focus.

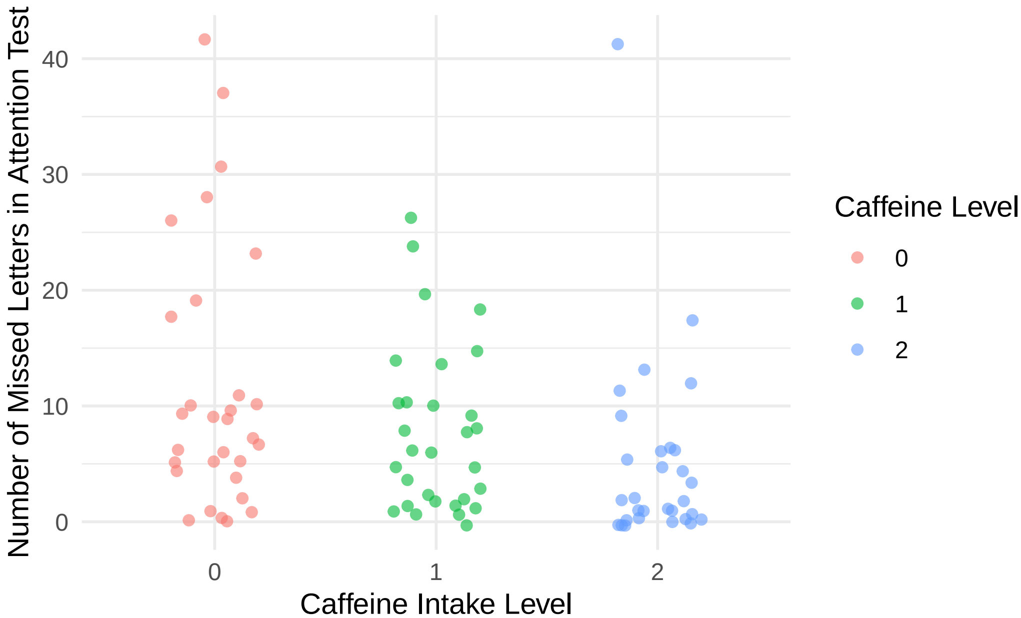


Figure 2: Missed Letters in Attention Test by Caffeine Intake Level

The scatter plot further confirms this pattern. The decaffeinated group has the highest variation, with some participants missing many more letters than others. The regular coffee group is more consistent, and the espresso group has the most stable performance with fewer missed letters.

The scatter plot also shows differences between individuals. Some participants in the decaffeinated group performed much worse than others, while the espresso group showed less variation. This reinforces that more caffeine leads to better and more stable attention performance.

Mental Arithmetic Test Visualizations

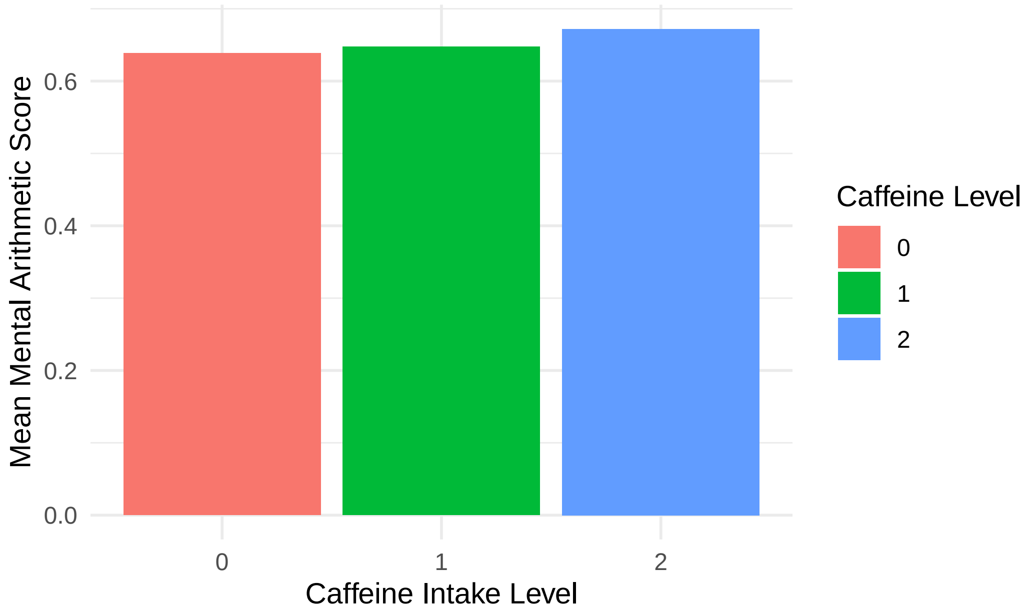
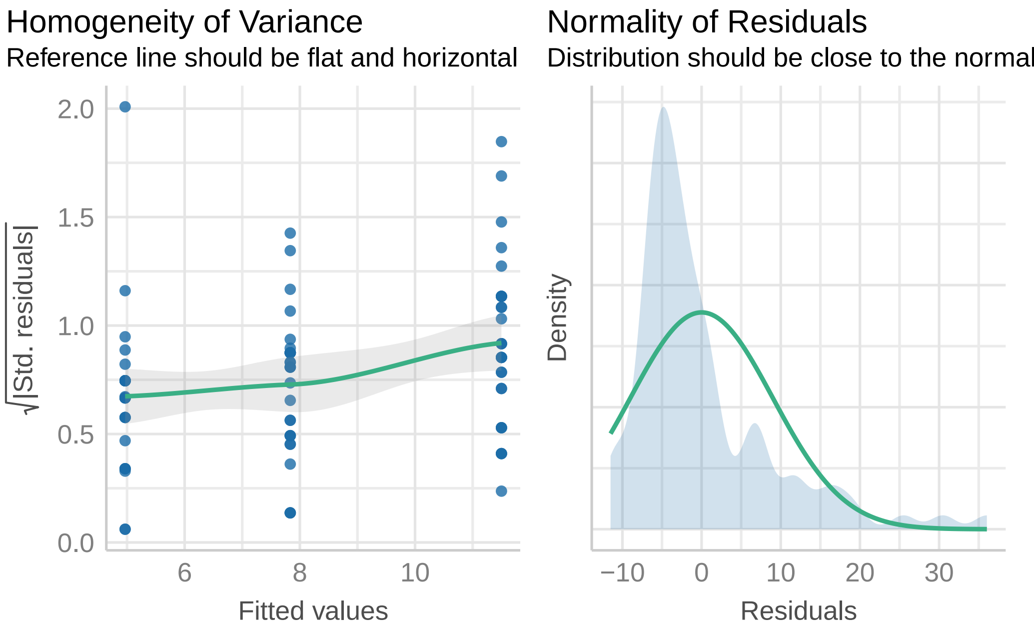


Figure 3: Mental Arithmetic Score by Caffeine Intake Level

The bar chart visualizes the mean mental arithmetic test scores for different caffeine intake levels (Decaffeinated = 0, Regular Coffee = 1, Espresso = 2). The scores are very close across all groups. The espresso group has the highest mean score. The decaffeinated group has the lowest, but the difference is small.

The results suggest that caffeine does not strongly impact mental arithmetic performance. Unlike the attention test, where higher caffeine intake clearly improved performance, the differences here are very minimal.

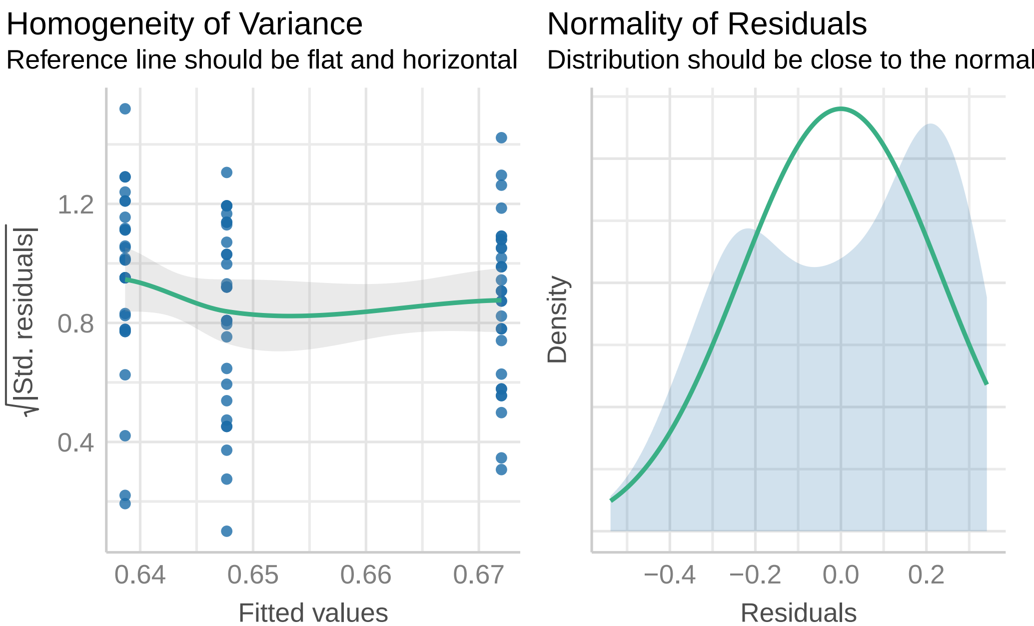
#Model Assumptions



The left plot (Homogeneity of Variance) checks if variance is the same across fitted values. The reference line is not completely flat, but it does not show a strong pattern. The spread of points is fairly even. This suggests that the homogeneity of variance assumption is likely met.

The right plot (Normality of Residuals) checks if residuals follow a normal distribution. The curve should be bell-shaped. The curve is not balanced. It skews to the right. This suggests that the normality assumption is violated.

The skewed distribution may be caused by the data. Some participants missed many more letters than others. This creates extreme outliers. The test scores may not follow a normal distribution. Since normality is violated, ANOVA results may not be valid. A transformation or Kruskal-Wallis Test may be needed.



The left plot (Homogeneity of Variance) shows if variance is the same across fitted values. The reference line is mostly flat. The spread of points is similar at different values. This suggests that the homogeneity of variance assumption holds.

The right plot (Normality of Residuals) checks if residuals follow a normal distribution. The curve is close to a bell shape. The peak is slightly uneven, but the distribution is mostly balanced. This suggests that the normality assumption is likely met.

Both assumptions appear to hold. ANOVA results should be valid for this model.

Appendix

Summary Statistics

```
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)

attention_summary = data.frame(Value = c(m1, m2, m3, m4))

kable(t(attention_summary),
      col.names = c("Mean", "Median", "SD", "IQR"),
      caption = "Table 1: Summary of Attention Test Scores",
      align = "c")
```

```
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)

math_summary <- data.frame(Value = c(m5, m6, m7, m8))

kable(t(math_summary), col.names = c("Mean", "Median", "SD", "IQR"),
      caption = "Table 2: Summary of Mental Arithmetic Test Scores",
      align = "c", booktabs = TRUE)
```

```
treatment_counts <- table(data$caffeine_intake_level)
treatment_percent <- prop.table(treatment_counts) * 100
treatment_summary <- rbind(
  Count = as.numeric(treatment_counts),
  Percentage = as.numeric(treatment_percent)
)
colnames(treatment_summary) <- names(treatment_counts)

kable(treatment_summary, caption = "Table 3: Summary of Participants in Each
      Treatment Group")
```

Visualization

```
ggplot(data, aes(x = as.factor(caffeine_intake_level),
                 y = num_of_missed_in_attention_test,
                 fill = as.factor(caffeine_intake_level))) +
  geom_boxplot() +
  labs(
    x = "Caffeine Intake Level",
    y = "Number of Missed Letters in Attention Test",
    fill = "Caffeine Level",
    caption = "Figure 4: Boxplot of Missed Letters in Attention Test by Caffeine
    Intake Level"
  ) +
  theme_minimal()
```

```
ggplot(data, aes(x = as.factor(caffeine_intake_level),
                 y = difficult_arithmetic_test_score,
                 fill = as.factor(caffeine_intake_level))) +
  geom_boxplot() +
  labs(
    x = "Caffeine Intake Level",
    y = "Mental Arithmetic Test Score",
    fill = "Caffeine Level",
    caption = "Figure 5: Boxplot of Mental Arithmetic Test Score by Caffeine
    Intake Level"
  ) +
  theme_minimal()
```

ANOVA Analysis

```
model_att <- aov(num_of_missed_in_attention_test ~ caffeine_intake_level,
                 data=data)
summary(model_att)

model_math <- aov(difficult_arithmetic_test_score ~ caffeine_intake_level,
                  data=data)
summary(model_math)
```

Attention Test Visualizations

```
att_summary <- data %>%
  group_by(caffeine_intake_level) %>%
  summarise(
    mean_att = mean(num_of_missed_in_attention_test),
    se_att   = sd(num_of_missed_in_attention_test) / sqrt(n())
  )

ggplot(att_summary, aes(x = as.factor(caffeine_intake_level), y = mean_att, fill = as.factor
  geom_col() +
  labs(
    x = "Caffeine Intake Level",
    y = "Number of Missed Letters in Attention Test",
    fill = "Caffeine Level",
    caption = "Figure 1: Missed Letters in Attention Test by Caffeine
    Intake Level"
    ) +
  theme_minimal()
```

```
ggplot(data, aes(x = as.factor(caffeine_intake_level),
                 y = num_of_missed_in_attention_test,
                 color = as.factor(caffeine_intake_level))) +
  geom_jitter(width = 0.2, alpha = 0.6) +
  labs(
    x = "Caffeine Intake Level",
    y = "Number of Missed Letters in Attention Test",
    color = "Caffeine Level",
    caption = "Figure 2: Missed Letters in Attention Test by Caffeine
    Intake Level"
  ) +
  theme_minimal()
```

Mental Arithmetic Test Visualizations

```
math_summary <- data %>%
  group_by(caffeine_intake_level) %>%
  summarise(
    mean_math = mean(difficult_arithmetic_test_score, na.rm=TRUE),
```

```r
    se_math    = sd(difficult_arithmetic_test_score, na.rm=TRUE) / sqrt(n())
  )

ggplot(math_summary, aes(x = as.factor(caffeine_intake_level),
                         y = mean_math,
                         fill = as.factor(caffeine_intake_level))) +
  geom_col() +
  labs(
    x = "Caffeine Intake Level",
    y = "Mean Mental Arithmetic Score",
    fill = "Caffeine Level",
    caption = "Figure 3: Mental Arithmetic Score by Caffeine Intake Level"
  ) +
  theme_minimal()
```

Model Assumptions

```r
data$caffeine_intake_level <- as.factor(data$caffeine_intake_level)

model_att <- aov(num_of_missed_in_attention_test ~ caffeine_intake_level,
                 data = data)
check_model(model_att, check = c("normality", "homogeneity"))
```

```r
data$caffeine_intake_level <- as.factor(data$caffeine_intake_level)

model_math <- aov(difficult_arithmetic_test_score ~ caffeine_intake_level,
                  data = data)
check_model(model_math, check = c("normality", "homogeneity"))
```