Impact of Caffeinated and Caffeine-Free Beverages on the Problem Solving Skills of A Virtual Population*

Bhagatinder Longia   Musab Muhammad   Muntasir Munem
Shea Munson   Chloe Syriac

March 17, 2025

Caffeine is the most popular psychoactive substance in the world, particularly due to its widely known stimulating effects. As a stimulant, caffeine has mostly commonly been used to increase mental alertness, which is why the goal of our study is to determine if there is an association between caffeine use and problem solving skills. We collected data using a virtual population on The Islands. After collecting a sample of 150 virtual participants and splitting them into three age groups, a caffeinated or caffeine-free drink, with water as a control, were administered to each participant. Using a balanced Kruskal-Wallis test, we found that [summary of key findings]. [brief overview of result implications, improvments needed].

Table of contents

*Data available at https://github.com/shea-m/Beverage-Impact-on-Problem-Solving-STA305H5S2025

1 Introduction

As the world's most popular psychoactive substance, caffeine's widely known effects have been studied extensively. Despite certain drawbacks like digestion and anxiety issues in some people, due to being a stimulant, it has most commonly been used to temporarily increase mental alertness which improves things like memory and lessens fatigue, thus providing an energy boost.

In this study, we are analyzing whether there is an association between caffeine use and problem solving skills by giving coffee, energy drinks, and caffeine-free versions of both to participants from the The Islands, a virtual population developed by the University of Queensland for learning and teaching in statistics. Our sample was split into three age groups: 18-35, 36-54, and 55+. Considering the effects of caffeine, participants assigned a caffeinated drink would be expected to achieve a higher score that those that are not.

We aim to study the following research questions:

- RQ1: Does beverage type have an impact on problem solving score?
  - H_0: no association between beverage type and problem solving score.
  - H_a: association between beverage type and problem solving score exists.
- RQ2: Do caffeinated coffee and energy drinks have a higher impact on problem solving scores compared to their caffeine-free counterparts?
  - H_0: no association between caffeinated drinks and higher score
  - H_a: association between caffeinated drinks and higher score exists.

This paper consist of our methodology, analysis, results, limitations, and a conclusion for our study.

2 Methodology

2.1 Design and Setup

In February 2025, we conducted an experiment to determine if different caffeinated and caffeine-free drinks effect the problem solving skills of virtual participants on The Islands. Our experiment utilized a balanced one-way randomized complete block (RCB) design. Participants were blocked out into one of three age groups, where they were given one of five drinks (the treatment) and had their score on a 20 minute problem solving test (the response) recorded.

Each of the three age group blocks contained an equal number of participants for each treatment, per our RCB design. While we did not have any major factors that we could easily control for, to help account for nuisance factors treatments were randomly assigned among participants using an R script (See Listing 1 in Appendix).

The experiment included a total of 150 participants ($n = 150$), with 50 in each of the following age blocks: 18-35, 36-54, and 55+. Each participant was given one of five 250 millilitre drinks ($a = 5$): Water (Control), Energy Drink, Energy Drink Decaffeinated, Coffee, or Coffee Decaffeinated. Each treatment factor contained 30 participants ($n_j = 30$), split evenly among the three age blocks thus giving a balanced design. This also provided for enough repetitions to further account for nuisance factors that might not have been dealt with through the blocking or randomization. Since each participant was exposed to only one treatment and randomly assigned to a treatment group, this experiment utilized a between-subjects design.

2.2 Data Collection

Each participant was given their randomly assigned drink and then administered a 20 minute problem solving test. The test was scored from 0 to 100 and after completion, each participants score was recorded. Once all data collection was completed, it was cleaned, stripping any personally identifying information and assigning factor levels and names more easily utilizable and readable within R.

3 Analysis

3.1 Summary Statistics and Checking Assumptions

We began by summarizing statistics of our data, presented in Table 1. The high mean and median indicate that participants generally scored well on the problem solving test. A low standard deviation indicates tight clustering around the mean and a low IQR indicates tight clustering around the median. This confirms the indication of similar mean and median. The minimum of 67 lies well outside of $1.5 \times$ IQR, indicating a presence of outliers on the low side, with further inspection necessary to determine the amount. The maximum of 100 along with the data being skewed means that there are no outliers present on the high end.

Table 1: Summary of Problem Solving Test Scores

	Min	Max	Mean	Median	SD	IQR
Value	67	100	93.4867	96	6.9618	6.75

More preliminary information is obtained plotting the treatments in relation to scores in Figure 1.

From the boxplot, we can make an initial inference that the means are likely similar among groups, however the variances are different. Further analysis will confirm these findings. Before we can confirm our initial inference of means, we must check the three assumptions for our planned one-way ANOVA: independence, normally distributed, and equal variance. The first assumption of independence is satisfied by design. All data points are collected from unique individuals whose results do not depend on the others. We next check normality and homogeneity of variance. From the test below in Figure 2 we see that both of these assumptions are violated.

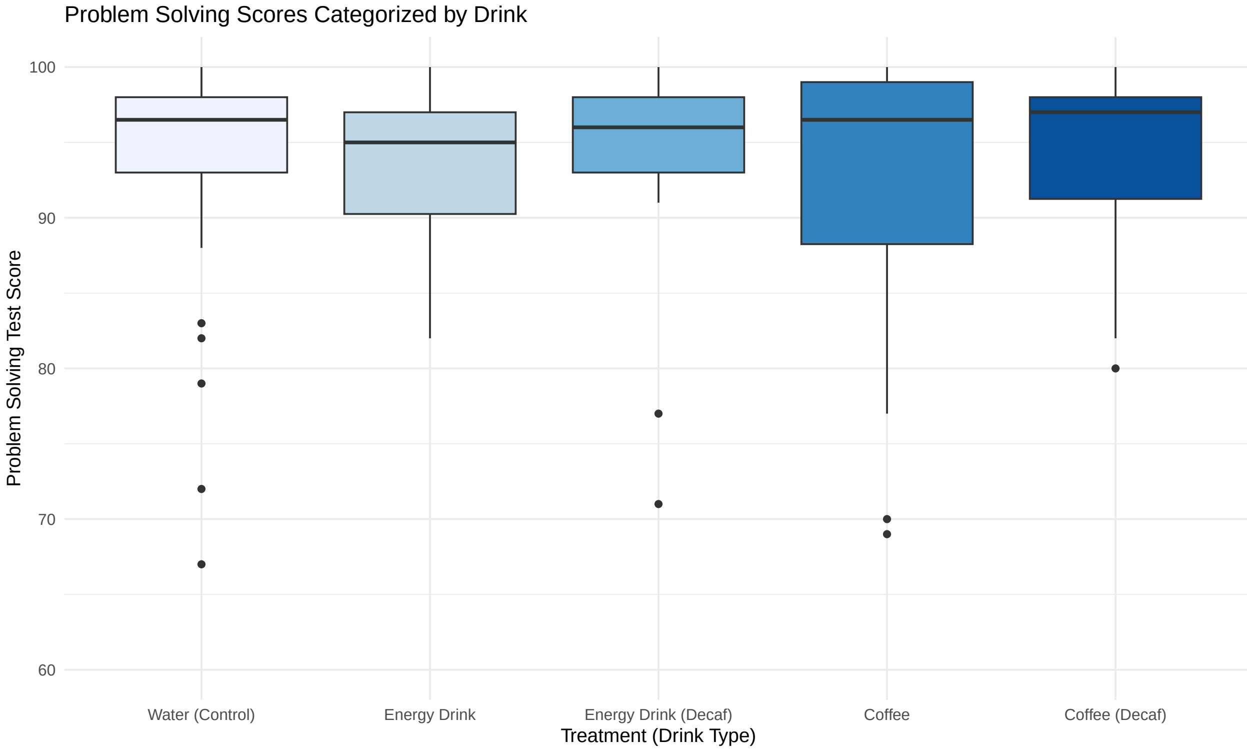

Figure 1: Boxplot of Scores by Treatment

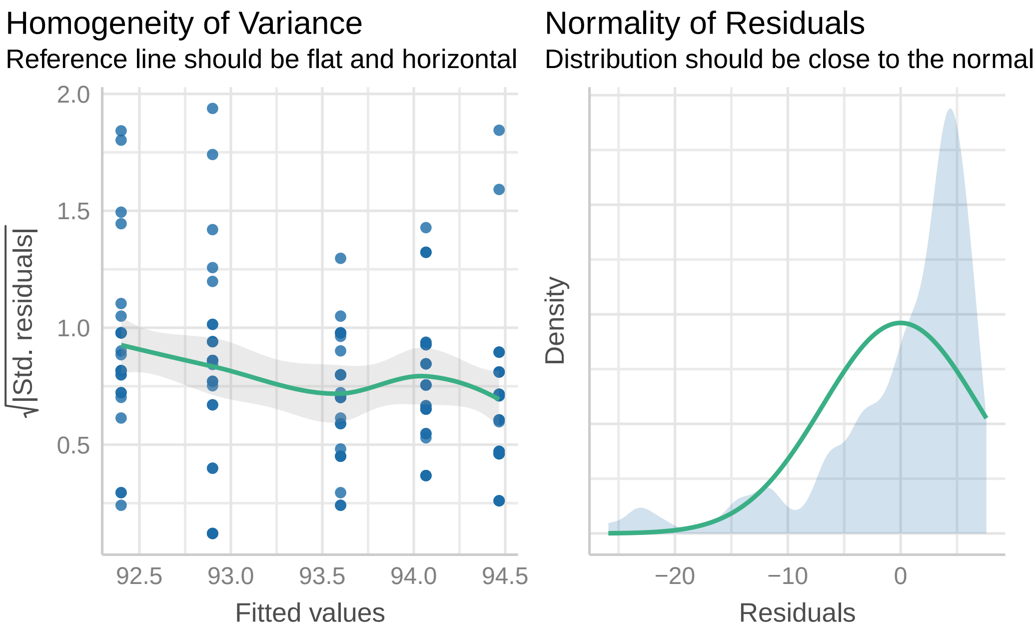

Figure 2: Checking One-Way ANOVA Assumptions

3.2 Statistical Tests

As the assumption of normality and homogeneity of variance have been violated, we are unable to use a one-way ANOVA test. Instead, we opt to use Kruskal-Wallis non-parametric H-test utilizing the treatment type (beverage given) as our factor. The results of the Kruskal-Wallis H-test are displayed below in Table 2.

Table 2: Kruskal-Wallis H-Test

H-Statistic	P Value	DF
1.811637	0.7703527	4

4 Results

From Table 2 we find a p-value for the Kruskal-Wallis H-test is 0.7735. This p-value is significantly greater than our significance level of $\alpha = 0.05$. As a result, there is not any statistically significant level evidence to suggest the mean problem solving score is different among treatment groups. This result is to be expected as it confirms the initial results inferred from Figure 1.

Since there we're no significant results, no further followup was necessary for RQ2. If there were statistically significant differences in caffeinated vs non-caffeinated beverages, we would have also seen statistically significant evidence that the means among drinks was not equal.

5 Limitations

This study had a number of limitations. One thing that should have done differently was waiting after giving each participant their beverage before administering the problem solving test. We administered the test immediately after giving participants their treatment, which could have impacted results. Waiting would have allowed for more absorption of caffeine and therefore a better look into its impact.

Additionally, despite the depth provided by The Islands for conducting an experiment, using a virtual population is a major limitation in assessing the impact of caffeine on the problem solving skills of real people. This is because it does not take all of the countless real life factors that impact caffeine into account. Further, we do not know exactly the implementation of the problem solving quiz or the drinks, resulting in a significant level of unknown. As a result, our experiment may not be incredibly relevant to the real world.

The final major limitation we had was the model used. We would have like to utilize a two-way ANOVA model, with drink and age as factors, as Figure 3 indicates there might be some difference among different drinks between age groups.

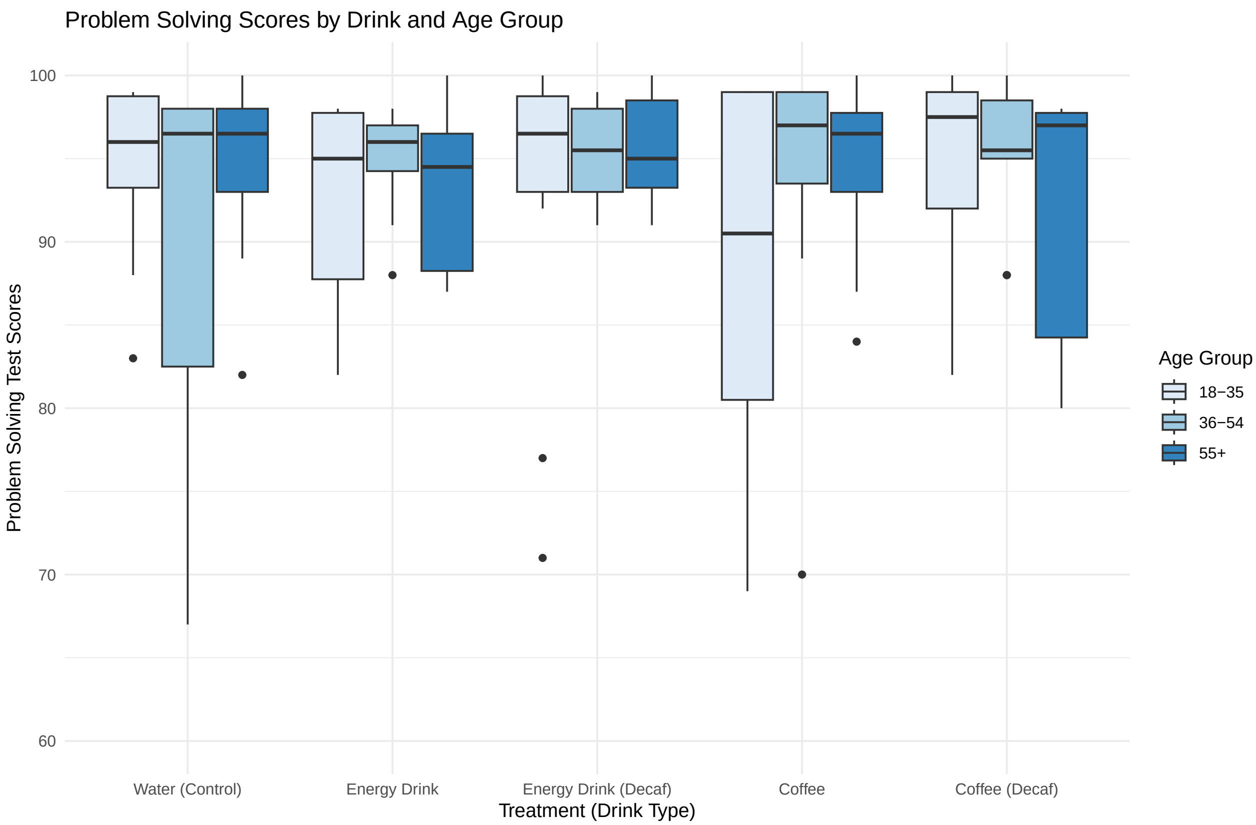

Figure 3: Box Plot Showing Problem Solving Scores by Drink and Age Group

Unfortunately, similar to one-way, our assumptions for both normality and homogeneity of variance were violated, shown in Figure 4. Due to the challenge of implementing a non-parametric two-way ANOVA test and because of time constraints, we were unfortunately unable to implement this model and had to instead utilize the one-way Kruskal-Wallis model.

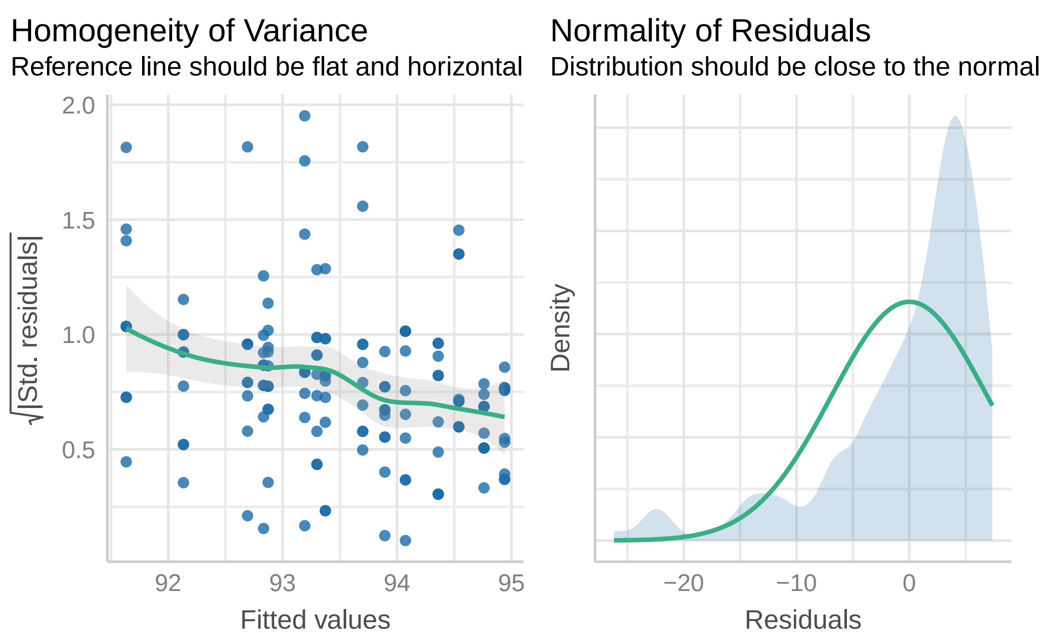

Figure 4

6 Conclusion

This study aimed to examine the potential association between caffeine consumption through various beverages and problem-solving performance scores among virtual participants in The Islands. Analysis from a sample of 150 participants showed [include key findings.]

7 Appendix

Listing 1 Randomization Script for Assigning Treatments

```r
library(clipr)
groups <- c("Water", "Coffee", "Coffee (Decaf)", "Energy Drink",
          "Energy Drink (Decaf)")
labels <- rep(groups, 10)
shuffled <- sample(labels)
write_clip(shuffled, type = "text")
```

```r
# Ensure necessary packages are installed and loaded
packages <- c("tidyverse", "patchwork", "performance", "knitr", "see",
              "kableExtra", "broom")
for(pkg in packages) {
  if (!requireNamespace(pkg, quietly = TRUE)) {
    install.packages(pkg)
  }
}
library(tidyverse)
library(patchwork)
library(performance)
library(knitr)
library(see)
library(kableExtra)
library(broom)

# Load the Data
data <- read.csv("cleaned_data.csv", header = TRUE)

# Put factor levels in desired order
data$age_group <- factor(data$age_group,
                         levels = c("Y", "M", "O"),
                         labels = c("18-35", "36-54", "55+"))
data$treatment <- factor(data$treatment,
                         levels = c("W", "E", "ED", "C", "CD"),
                         labels = c("Water (Control)", "Energy Drink",
                                    "Energy Drink (Decaf)", "Coffee",
                                    "Coffee (Decaf)"))
attach(data)
```

```r
# Generate table for score summary data
sum_data <- data.frame(Value = c(min(score), max(score), round(mean(score), 4),
                                 median(score), round(sd(score), 4), IQR(score)))
labs <- c("Min", "Max", "Mean", "Median", "SD", "IQR")
kable(t(sum_data), col.names = labs, align = "c")

# Generate boxplot for scores explained by drink
ggplot(data, aes(x = treatment, y = score, fill = treatment)) +
  geom_boxplot() +
  scale_y_continuous(limits = c(60, 100)) +
  scale_fill_brewer(palette = "Set1") +
  labs(title = "Problem Solving Scores Categorized by Drink",
  x = "Treatment (Drink Type)", y = "Problem Solving Test Score") +
  theme(legend.position = "none")

# Check assumptions for one-way model
one_factor_model <- aov(score ~ treatment, data = data)
check_model(one_factor_model, check = c("normality", "homogeneity"))

# Perform non-parametric K-W test
non_par <- kruskal.test(score~treatment, data)
formatted <- as.data.frame(tidy(non_par))
formatted$method <- NULL
col_names <- c("H-Statistic", "P Value", "DF")
kable(formatted, col.names = col_names,
      caption = "Kruskal-Wallis H-Test")

# Generate boxplot for scores explained by drink and age
age <- c("18-35", "36-54", "55+")
ggplot(data, aes(x = treatment, y = score, fill = age_group)) +
  geom_boxplot(position = position_dodge(width = 0.8)) +
  scale_y_continuous(limits = c(60, 100)) +
  labs(title = "Problem Solving Scores by Drink and Age Group",
       x = "Treatment (Drink Type)", y = "Problem Solving Test Scores",
       fill = "Age Group") +
  scale_fill_brewer(palette = "Blues", breaks = age, labels = age) +
  theme_minimal()

# Check assumptions for 2 factor model
two_factor_model <- aov(score ~ treatment + age_group, data = data)
check_model(two_factor_model, check = c("normality", "homogeneity"))
```