

Formative Assessment 10

APM1111 Statistical Theory

Lopez, Ian Darylle C. Pineda, Erianne L. Sindayen, Lorenzo Danilo V.

GitHub Link:
<https://github.com/ChevyGnome/APM1111>

Dataset

```
cholesterol_data <- data.frame(
  ID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18),
  Before = c(6.42, 6.76, 6.56, 4.8, 8.43, 7.49, 8.05, 5.05, 5.77,
            3.91, 6.77, 6.44, 6.17, 7.67, 7.34, 6.85, 5.13, 5.73),
  After4Weeks = c(5.83, 6.2, 5.83, 4.27, 7.71, 7.12, 7.25, 4.63, 5.31,
                 3.7, 6.15, 5.59, 5.56, 7.11, 6.84, 6.4, 4.52, 5.13),
  After8Weeks = c(5.75, 6.13, 5.71, 4.15, 7.67, 7.05, 7.1, 4.67, 5.33,
                 3.66, 5.96, 5.64, 5.51, 6.96, 6.82, 6.29, 4.45, 5.17),
  Margarine = c("B", "B", "B", "A", "B", "A", "B", "A", "B",
               "A", "B", "B", "A", "A", "B", "A", "B")
)

df_long <- cholesterol_data %>%
  pivot_longer(
    cols = c("Before", "After4Weeks", "After8Weeks"),
    names_to = "Time",
    values_to = "Cholesterol"
  )

df_long$ID <- as.factor(df_long$ID)
df_long$Margarine <- as.factor(df_long$Margarine)
df_long$Time <- factor(df_long$Time, levels = c("Before", "After4Weeks", "After8Weeks"))

kable(cholesterol_data, caption = "Cholesterol Over Time Data Set")
```

Table 1: Cholesterol Over Time Data Set

| ID | Before | After4Weeks | After8Weeks | Margarine |
|----|--------|-------------|-------------|-----------|
| 1 | 6.42 | 5.83 | 5.75 | B |

| ID | Before | After4Weeks | After8Weeks | Margarine |
|----|--------|-------------|-------------|-----------|
| 2 | 6.76 | 6.20 | 6.13 | B |
| 3 | 6.56 | 5.83 | 5.71 | B |
| 4 | 4.80 | 4.27 | 4.15 | A |
| 5 | 8.43 | 7.71 | 7.67 | B |
| 6 | 7.49 | 7.12 | 7.05 | A |
| 7 | 8.05 | 7.25 | 7.10 | B |
| 8 | 5.05 | 4.63 | 4.67 | A |
| 9 | 5.77 | 5.31 | 5.33 | B |
| 10 | 3.91 | 3.70 | 3.66 | A |
| 11 | 6.77 | 6.15 | 5.96 | B |
| 12 | 6.44 | 5.59 | 5.64 | B |
| 13 | 6.17 | 5.56 | 5.51 | A |
| 14 | 7.67 | 7.11 | 6.96 | A |
| 15 | 7.34 | 6.84 | 6.82 | A |
| 16 | 6.85 | 6.40 | 6.29 | B |
| 17 | 5.13 | 4.52 | 4.45 | A |
| 18 | 5.73 | 5.13 | 5.17 | B |

Introduction

A requirement in the management of one's cardiovascular health is following consumption-based interventions aimed at managing lipid levels, specifically cholesterol, which is an essential biological indicator for heart health. This study explores the effects of two different margarine brands (Brand A and Brand B) on cholesterol levels over the span of 8 weeks.

Cholesterol levels were measured at three different time points. The first is measured before consumption, the second is measured after 4 weeks of consumption, while the last is measured after 8 weeks of consumption. Each participant belonged to one of two groups depending on which margarine was consumed. With the experiment design consisting of a between-subjects factor (Margarine) and a within-subjects factor (Time), a two-way mixed model Analysis of Variance (ANOVA), otherwise known as split-plot ANOVA, was applicable to the study.

In this analysis, the primary objective was determining the change in cholesterol levels over a period of two months and whether the specific margarine brand consumed factored on that change in cholesterol level. The study also aims to explore whether cholesterol levels decreased over time regardless of margarine brand, if one brand was more effective than the other, and whether there was an interaction between margarine and time.

Assumptions

Assumption #1: You have a continuous dependent variable.

Assumption #2: You have one between-subjects factor (i.e., independent variable) that is categorical with two or more categories.

Assumption #3: You have one within-subjects factor (i.e., independent variable) that is categorical with two or more categories.

Assumption #4: There should be no significant outliers in any cell of the design.

Assumption #5: The dependent variable should be approximately normally distributed for each cell of the design.

Assumption #6: The variance of your dependent variable should be equal between the groups of the between-subjects factor, referred to as the assumption of homogeneity of variances.

Assumption #7: There should be homogeneity of covariances.

Assumption #8: The variance of the differences between groups should be equal, referred to as the assumption of sphericity.

Hypotheses

A. Main Effect of Between-Subjects Factor (Margarine)

Null Hypothesis (H_0): There is no significant difference in mean cholesterol levels between Margarine A and Margarine B.

Alternative Hypothesis (H_A): There is a significant difference in mean cholesterol levels between Margarine A and Margarine B.

B. Main Effect of Within-Subjects Factor (Time)

Null Hypothesis (H_0): There is no significant difference in the mean cholesterol levels across the three time points.

Alternative Hypothesis (H_A): There is a significant difference in the mean cholesterol levels across the three time points.

C. Interaction Effect (Time \times Margarine)

Null Hypothesis (H_0): There is no significant interaction between time and margarine on cholesterol levels.

Alternative Hypothesis (H_A): There is a significant interaction between time and margarine on cholesterol levels.

Checking of Assumptions

Assumption #1: You have a continuous dependent variable.

Remark: The dependent variable is cholesterol level and is continuous.

Assumption #2: You have one between-subjects factor (i.e., independent variable) that is categorical with two or more categories.

Remark: The between-subjects factor is the brand of margarine and has 2 categories.

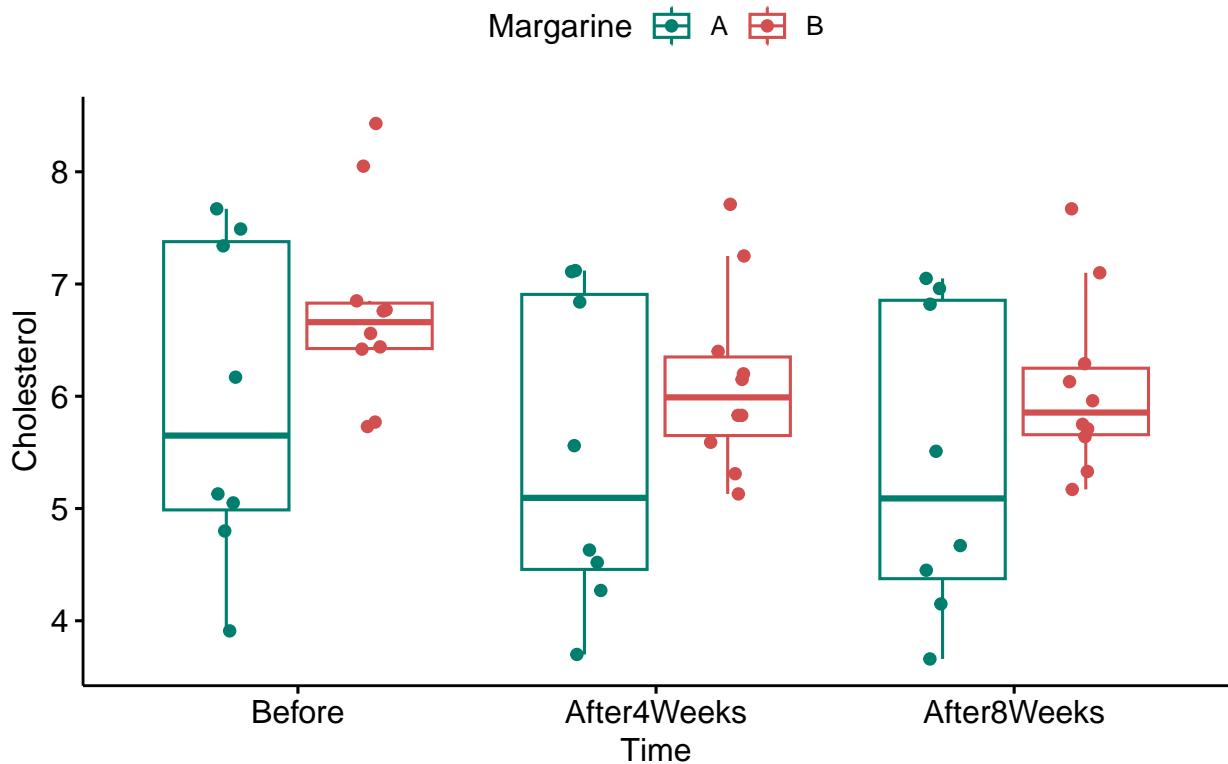
Assumption #3: You have one within-subjects factor (i.e., independent variable) that is categorical with two or more categories.

Remark: The within-subjects factor is time and has 3 categorical levels (before, after 4 weeks, after 8 weeks).

Assumption #4: There should be no significant outliers in any cell of the design.

```
ggbboxplot(  
  df_long,  
  x = "Time",  
  y = "Cholesterol",  
  color = "Margarine",  
  palette = c("#007E6E", "#D34E4E"),  
  add = "jitter"  
) +  
  labs(title = "Boxplot of Cholesterol Distribution")
```

Boxplot of Cholesterol Distribution



Remark: There are no significant outliers as assessed by visual inspection of the boxplot.

Assumption #5: The dependent variable should be approximately normally distributed for each cell of the design.

```
normality_results <- df_long %>%
  group_by(Margarine, Time) %>%
  shapiro_test(Cholesterol)

kable(normality_results, caption = "Shapiro-Wilk Test for Normality")
```

Table 2: Shapiro-Wilk Test for Normality

| Margarine | Time | variable | statistic | p |
|-----------|-------------|-------------|-----------|-----------|
| A | Before | Cholesterol | 0.9005331 | 0.2921657 |
| A | After4Weeks | Cholesterol | 0.8710561 | 0.1543526 |
| A | After8Weeks | Cholesterol | 0.8763769 | 0.1738318 |
| B | Before | Cholesterol | 0.8809547 | 0.1338272 |
| B | After4Weeks | Cholesterol | 0.9249442 | 0.4000403 |
| B | After8Weeks | Cholesterol | 0.8994147 | 0.2158757 |

Remark: The dependent variable, cholesterol level, is approximately normally distributed for each combination of the groups of margarine and time, as assessed by Shapiro-Wilk test of normality, ($p > 0.05$)

Assumption #6: The variance of your dependent variable should be equal between the groups of the between-subjects factor, referred to as the assumption of homogeneity of variances.

```
levene_results <- df_long %>%
  group_by(Time) %>%
  levene_test(Cholesterol ~ Margarine)

kable(levene_results, caption = "Levene's Test for Homogeneity of Variances")
```

Table 3: Levene's Test for Homogeneity of Variances

| Time | df1 | df2 | statistic | p |
|-------------|-----|-----|-----------|-----------|
| Before | 1 | 16 | 4.399481 | 0.0521897 |
| After4Weeks | 1 | 16 | 4.175257 | 0.0578472 |
| After8Weeks | 1 | 16 | 4.718317 | 0.0452175 |

Remark: Homogeneity of variances was met for the 'Before' ($p > .05$) and 'After 4 weeks' ($p > .05$) time points. However, the assumption was not met for the 'After 8 weeks' time point ($p < .05$).

Assumption #7: There should be homogeneity of covariances.

```
box_m_result <- box_m(cholesterol_data[, c("Before", "After4Weeks", "After8Weeks")], cholesterol_data$M)

kable(box_m_result, caption = "Box's M-test for Homogeneity of Covariances")
```

Table 4: Box's M-test for Homogeneity of Covariances

| statistic | p.value | parameter | method |
|-----------|-----------|-----------|---|
| 4.047784 | 0.6702099 | 6 | Box's M-test for Homogeneity of Covariance Matrices |

Remark: Homogeneity of covariances was met ($p > .05$) which indicates that the covariance matrices are equal across groups.

Assumption #8: The variance of the differences between groups should be equal, referred to as the assumption of sphericity.

```
res.aov <- df_long %>%
  anova_test(dv = Cholesterol, wid = ID, within = Time, between = Margarine)

sphericity_table <- res.aov[["Mauchly's Test for Sphericity"]]
kable(sphericity_table, caption = "Mauchly's Test for Sphericity")
```

Table 5: Mauchly's Test for Sphericity

| Effect | W | p | p<.05 |
|----------------|-------|-------|-------|
| Time | 0.462 | 0.003 | * |
| Margarine:Time | 0.462 | 0.003 | * |

Remark: The assumption of sphericity was not met ($p < .05$) which indicates that the variances of the differences among the three time points are not equal. Thus, a correction must be applied to account for the violation. In this case, the Greenhouse-Geisser correction was utilized.

Descriptive Statistics

```
descriptives <- df_long %>%
  group_by(Margarine, Time) %>%
  get_summary_stats(Cholesterol, type = "mean_sd")

kable(descriptives, caption = "Descriptive Statistics")
```

Table 6: Descriptive Statistics

| Margarine | Time | variable | n | mean | sd |
|-----------|-------------|-------------|----|-------|-------|
| A | Before | Cholesterol | 8 | 5.945 | 1.428 |
| A | After4Weeks | Cholesterol | 8 | 5.469 | 1.388 |
| A | After8Weeks | Cholesterol | 8 | 5.409 | 1.374 |
| B | Before | Cholesterol | 10 | 6.778 | 0.866 |
| B | After4Weeks | Cholesterol | 10 | 6.140 | 0.815 |
| B | After8Weeks | Cholesterol | 10 | 6.075 | 0.779 |

Two-Way Mixed ANOVA Test

```
anova_table <- get_anova_table(res.aov, correction = "GG")

kable(anova_table, caption = "Two-Way Mixed ANOVA (Greenhouse-Geisser Corrected)")
```

Table 7: Two-Way Mixed ANOVA (Greenhouse-Geisser Corrected)

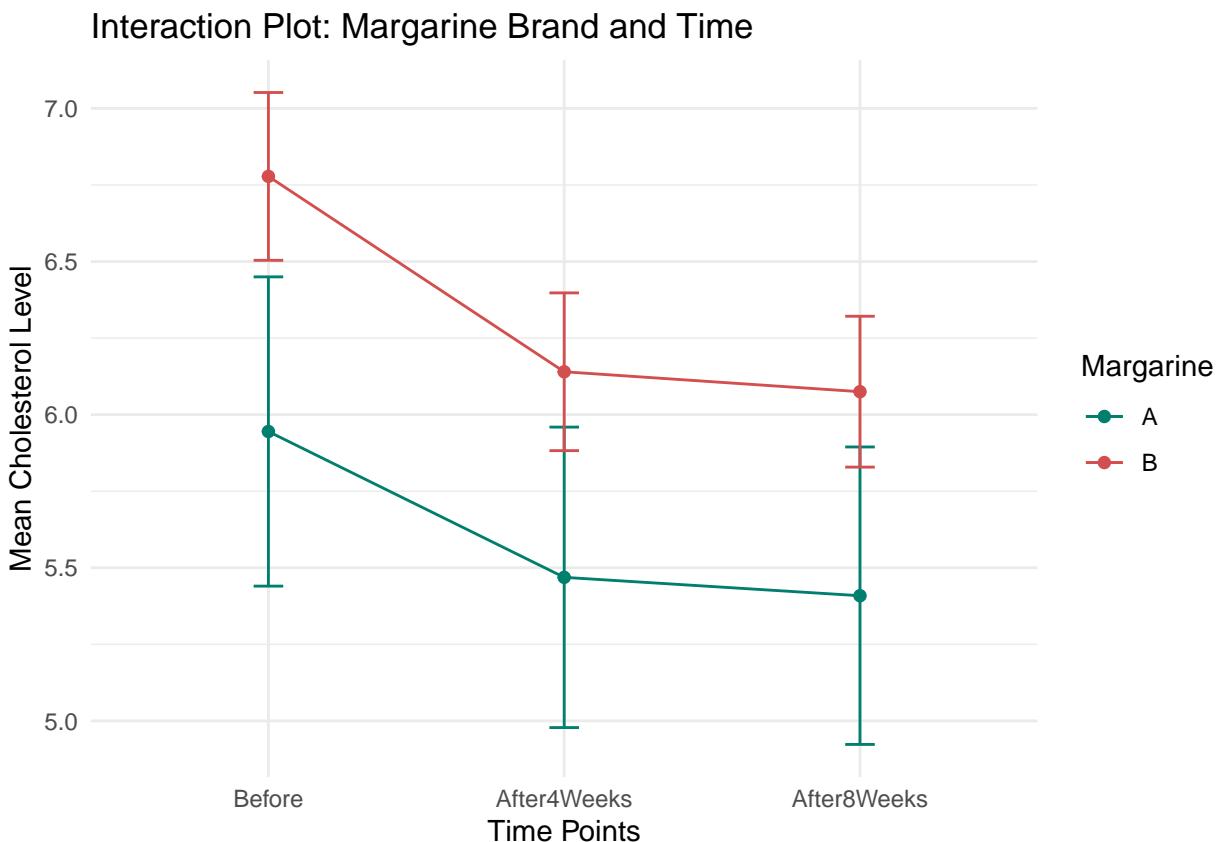
| Effect | DFn | DFd | F | p | p<.05 | ges |
|----------------|-----|-------|---------|-------|-------|-------|
| Margarine | 1.0 | 16.00 | 1.896 | 0.187 | | 0.106 |
| Time | 1.3 | 20.81 | 249.008 | 0.000 | * | 0.065 |
| Margarine:Time | 1.3 | 20.81 | 4.814 | 0.031 | * | 0.001 |

Remark: As observed, there was a significant main effect of Time ($p < .05$), indicating that cholesterol levels changed across time points. However, there was no significant main effect of Margarine, ($p > .05$), indicating that the overall mean cholesterol levels did not differ between brands. There was also a significant Time \times Margarine interaction ($p < .05$). Thus, the effect of time differed between Margarine A and B.

```

ggline(
  df_long,
  x = "Time",
  y = "Cholesterol",
  color = "Margarine",
  add = "mean_se",
  palette = c("#007E6E", "#D34E4E"),
  title = "Interaction Plot: Margarine Brand and Time",
  ylab = "Mean Cholesterol Level",
  xlab = "Time Points"
) +
  theme_minimal()

```



Simple Main Effects Analysis

(a) Simple Main Effects Analysis of Time per Margarine Brand

```

one_way_time <- df_long %>%
  group_by(Margarine) %>%
  anova_test(dv = Cholesterol, wid = ID, within = Time) %>%
  get_anova_table() %>%
  adjust_pvalue(method = "bonferroni")

kable(one_way_time, caption = "Simple Main Effect of Time per Margarine Brand")

```

Table 8: Simple Main Effect of Time per Margarine Brand

| Margarine | Effect | DFn | DFd | F | p | p<.05 | ges | p.adj |
|-----------|--------|------|-------|---------|----------|-------|-------|----------|
| A | Time | 1.16 | 8.09 | 82.741 | 1.18e-05 | * | 0.033 | 2.36e-05 |
| B | Time | 2.00 | 18.00 | 182.303 | 0.00e+00 | * | 0.142 | 0.00e+00 |

(b) Simple Main Effects Analysis of Margarine per Time Point

```
one_way_margarine <- df_long %>%
  group_by(Time) %>%
  anova_test(dv = Cholesterol, wid = ID, between = Margarine) %>%
  get_anova_table() %>%
  adjust_pvalue(method = "bonferroni")

kable(one_way_margarine, caption = "Simple Main Effect of Margarine per Time Point")
```

Table 9: Simple Main Effect of Margarine per Time Point

| Time | Effect | DFn | DFd | F | p | p<.05 | ges | p.adj |
|-------------|-----------|-----|-----|-------|-------|-------|-------|-------|
| Before | Margarine | 1 | 16 | 2.346 | 0.145 | | 0.128 | 0.435 |
| After4Weeks | Margarine | 1 | 16 | 1.647 | 0.218 | | 0.093 | 0.654 |
| After8Weeks | Margarine | 1 | 16 | 1.691 | 0.212 | | 0.096 | 0.636 |

Post Hoc Analysis

(a) Pairwise Comparisons for Time (per Margarine)

```
pwc_time <- df_long %>%
  group_by(Margarine) %>%
  pairwise_t_test(
    Cholesterol ~ Time,
    paired = TRUE,
    p.adjust.method = "bonferroni"
  )

kable(pwc_time, caption = "Pairwise Comparisons: Time levels within Margarine")
```

Table 10: Pairwise Comparisons: Time levels within Margarine

| Margarine | .y. | group1 | group2 | n1 | n2 | statistic | df | p | p.adj | p.adj.signif |
|-----------|-------------|--------|-------------|----|----|-----------|----|---------|---------|--------------|
| A | Cholesterol | Before | After4Weeks | 8 | 8 | 9.824754 | 7 | 2.4e-05 | 7.2e-05 | **** |
| A | Cholesterol | Before | After8Weeks | 8 | 8 | 9.087783 | 7 | 4.0e-05 | 1.2e-04 | *** |

| Margarine | .y. | group1 | group2 | n1 | n2 | statistic | df | p | p.adj | p.adj.signif |
|-----------|-------------|-------------|-------------|----|----|-----------|----|---------|---------|--------------|
| A | Cholesterol | After4Weeks | After8Weeks | 8 | 8 | 2.898275 | 7 | 2.3e-02 | 6.9e-02 | ns |
| B | Cholesterol | Before | After4Weeks | 10 | 10 | 14.965919 | 9 | 1.0e-07 | 3.0e-07 | **** |
| B | Cholesterol | Before | After8Weeks | 10 | 10 | 14.116264 | 9 | 2.0e-07 | 6.0e-07 | **** |
| B | Cholesterol | After4Weeks | After8Weeks | 10 | 10 | 2.512211 | 9 | 3.3e-02 | 1.0e-01 | ns |

(b) Pairwise Comparisons for Margarine (per Time Point)

```
pwc_margarine <- df_long %>%
  group_by(Time) %>%
  pairwise_t_test(
    Cholesterol ~ Margarine,
    paired = FALSE, # Independent t-test
    p.adjust.method = "bonferroni"
  )

kable(pwc_margarine, caption = "Pairwise Comparisons: Margarine levels within Time")
```

Table 11: Pairwise Comparisons: Margarine levels within Time

| Time | .y. | group1 | group2 | n1 | n2 | p | p.signif | p.adj | p.adj.signif |
|-------------|-------------|--------|--------|----|----|-------|----------|-------|--------------|
| Before | Cholesterol | A | B | 8 | 10 | 0.145 | ns | 0.145 | ns |
| After4Weeks | Cholesterol | A | B | 8 | 10 | 0.218 | ns | 0.218 | ns |
| After8Weeks | Cholesterol | A | B | 8 | 10 | 0.212 | ns | 0.212 | ns |

Reporting

In this study, the researchers employed the two-way mixed ANOVA test to evaluate the effect of two Margarine brands (brand A and brand B) and three different time points (Before, After 4 Weeks, and After 8 Weeks) on cholesterol levels. Assumptions were verified before performing analysis. Outliers were assessed by visual inspection of a boxplot; normality was assessed by Shapiro-Wilk's normality test; homogeneity of variances was assessed by Levene's test; homogeneity of covariances was assessed by Box's M-test. Results showed that there were no outliers, residuals were normally distributed ($p > .05$), homogeneity of variances was met for the 'Before' and 'After 4 weeks' time points ($p > .05$) but was not met for the 'After 8 weeks' ($p < .05$) time point. Homogeneity of covariances was also met ($p > .05$). However, Mauchly's test indicated that the assumption of sphericity was violated for the Time factor, ($W = 0.462, p = .003$). Due to these findings, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

ANOVA results showed that there was no statistically significant main effect of margarine brand on cholesterol levels, $F(1, 16) = 1.896, p = .187, \eta^2_G = .106$. These statistics support the notion that cholesterol levels did not differ significantly between participants who used Margarine A and Margarine B. However, there was a statistically significant main effect of time on cholesterol levels, $F(1.30, 20.81) = 249.008, p < .001, \eta^2_G = .065$. This indicates that the cholesterol levels have major differences that span over the three different time points. The interaction between Margarine brand and Time was also statistically significant,

$F(1.30, 20.81) = 4.814, p = .031, \eta^2_G = .001$. This suggests that the pattern of change in cholesterol level over time differed between the two margarine brands.

Simple main effects analysis was executed to inspect the effect of time within each margarine brand. For Brand A, there was a significant effect of time on cholesterol levels, $p < .001$. Pairwise comparisons with Bonferroni correction revealed that cholesterol levels significantly decreased from the beginning ($M = 5.95$) to 4 Weeks ($M = 5.47, p < .01$) and to 8 Weeks ($M = 5.41, p < .01$). Similarly, for Brand B, there was also a significant effect of time on cholesterol levels, $p < .001$, where cholesterol levels significantly decreased from the beginning ($M = 6.78$) to 4 Weeks ($M = 6.14, p < .001$) and to 8 Weeks ($M = 6.08, p < .001$).

Conclusion

The results of this study indicated that both Margarine A and Margarine B were associated with significant reductions in cholesterol levels over time. While the interaction effect was statistically significant ($p = .031$), the effect size was negligible, and post-hoc analysis showed that both margarine brands had similar trajectories of having a major drop on cholesterol level during the first four weeks but stabilized in the time leading up to 8 weeks. Thus, time proved to have a significant effect on improving cholesterol levels, while the specific margarine brand did not.