# Data Analysis for Cholesterol Over Time Using a Two-way Mixed Model ANOVA

Borromeo_Mayo_Mercado

2025-12-04

github link: https://github.com/chieelo/STATS/tree/main/FA10_GROUP3_BORROMEO_MAYO_MERCADO

## Introduction

This study examines the impact of two margarine brands (Brand A and Brand B) on cholesterol levels at three different time points: prior to consumption, four weeks later, and eight weeks later. Both the within-subjects factor (Time) and the between-subjects factor (Margarine type) are evaluated using a two-way mixed model ANOVA, which enables the analysis of main effects and interactions.

## Null Hypothesis

There is no significant difference in the cholesterol levels between the two brands of margarine over the three time points.

## Dataset

```r
library(knitr)
library(kableExtra)

df %>%
kable(caption = "Cholesterol Data", booktabs = TRUE) %>%
kable_styling(latex_options = c("striped", "hold_position"))
```

Table 1: Cholesterol Data

| ID | Before | After4weeks | After8weeks | Margarine |
|---|---|---|---|---|
| 1 | 6.42 | 5.83 | 5.75 | B |
| 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 |

## Problem

The research question here is:

> Do cholesterol levels change over time for each margarine brand individually?

{add more if needed}

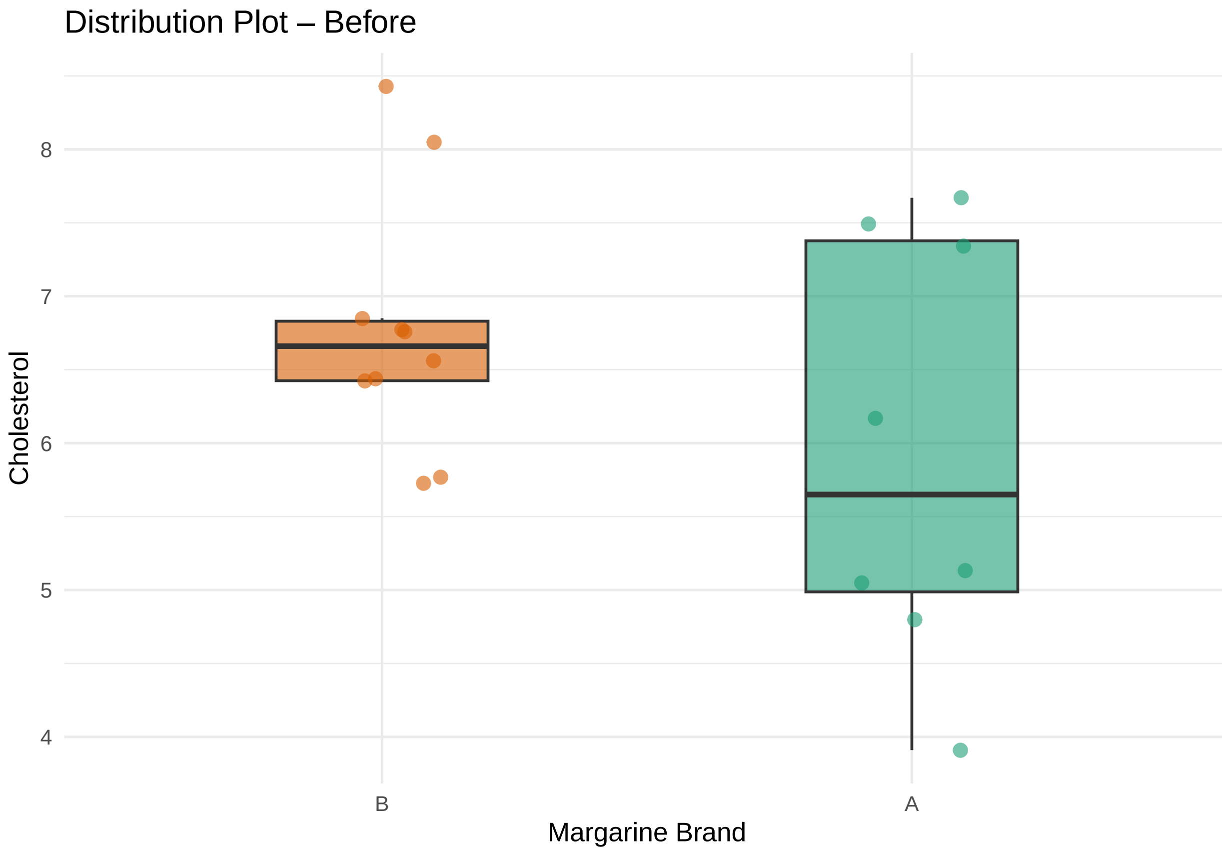
## Checking of Assumptions

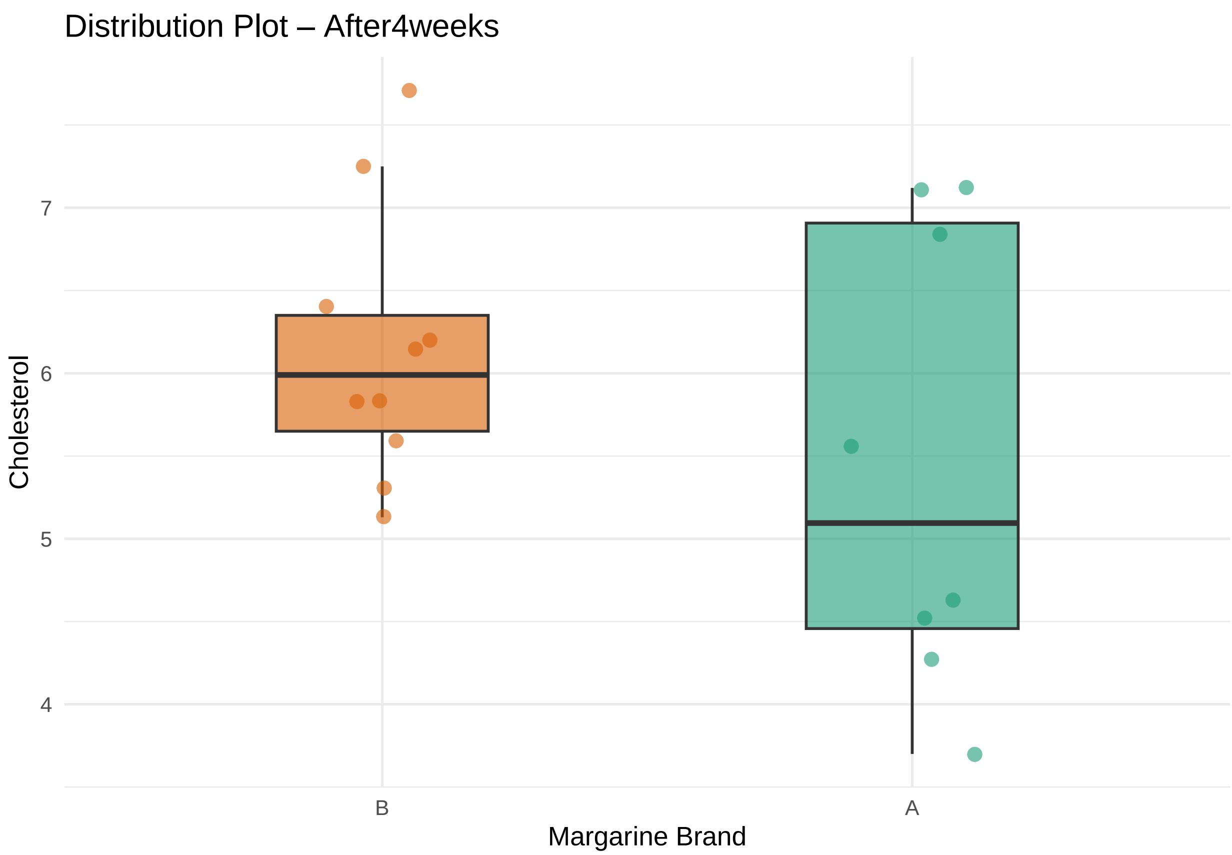
**Assumption 1: You have a continuous dependent variable.** Remark: The dependent variable is cholesterol level and is continuous. Therefore, Assumption 1 is satisfied.

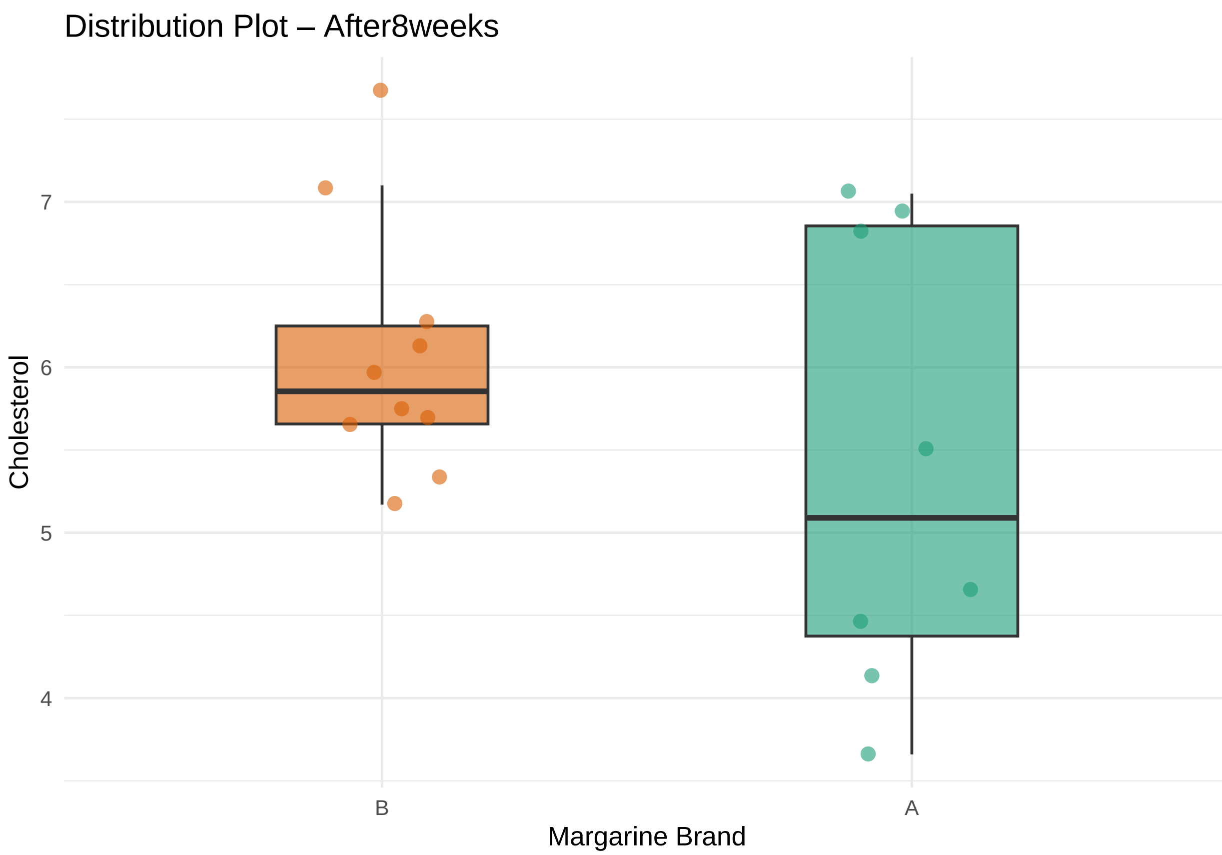
**Assumption 2: You have one between-subjects factor (i.e., independent variable) that is categorical with two or more categories.** Remark: The between-subjects factor is the brand of margarine and has 2 categories.

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




Distribution Plot – After4weeks
Cholesterol
7
6
5
4
B
A
Margarine Brand




Remark: Visual inspection of the raincloud plots for each time point (Before, After 4 weeks, After 8 weeks) and each Margarine group (A and B) shows that all data points fall reasonably close to the bulk of the distribution. There are no extreme values or points that appear far outside the main distribution, indicating that there are no significant outliers in any group. Therefore, Assumption 4 is satisfied.

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

Table 2: Descriptive Statistics of Cholesterol by Time and Margarine

| | Before | | After 4 Weeks | | After 8 Weeks | |
|---|---|---|---|---|---|---|
| | Before_B | Before_A | After4weeks_B | After4weeks_A | After8weeks_B | After8weeks_A |
| Valid | 10.00 | 8.00 | 10.00 | 8.00 | 10.00 | 8.00 |
| Mean | 6.78 | 5.94 | 6.14 | 5.47 | 6.08 | 5.41 |
| Std. Deviation | 0.87 | 1.43 | 0.81 | 1.39 | 0.78 | 1.37 |
| Minimum | 5.73 | 3.91 | 5.13 | 3.70 | 5.17 | 3.66 |
| Maximum | 8.43 | 7.67 | 7.71 | 7.12 | 7.67 | 7.05 |
| Skewness | 0.77 | 0.02 | 0.74 | 0.16 | 0.95 | 0.14 |
| Std. Error of Skewness | 0.77 | 0.87 | 0.77 | 0.87 | 0.77 | 0.87 |
| Kurtosis | -0.30 | -1.50 | -0.41 | -1.62 | -0.11 | -1.62 |
| Std. Error of Kurtosis | 1.55 | 1.73 | 1.55 | 1.73 | 1.55 | 1.73 |
| Shapiro_Wilk | 0.88 | 0.90 | 0.92 | 0.87 | 0.90 | 0.88 |

| | | | | | | |
|---|---|---|---|---|---|---|
| P-value Shapiro-Wilk | 0.13 | 0.29 | 0.40 | 0.15 | 0.22 | 0.17 |

Remark: The Shapiro-Wilk test was conducted for cholesterol levels in each Margarine group at all time points. All p-values were greater than 0.05 (Before: B = 0.13, A = 0.29; After 4 Weeks: B = 0.40, A = 0.15; After 8 Weeks: B = 0.22, A = 0.17), indicating no significant deviation from normality. Skewness and kurtosis values were also within acceptable ranges. Therefore, Assumption 5 of approximate normality is satisfied for all groups and time points.

**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.**

Table 3: Levene's Test for Homogeneity of Variances (Mean-based, matching JASP)

| Time | F | df1 | df2 | p-value |
|---|---|---|---|---|
| Before | 5.035 | 1 | 16 | 0.039 |
| After4weeks | 5.415 | 1 | 16 | 0.033 |
| After8weeks | 5.975 | 1 | 16 | 0.026 |

Remark: Since all p-values are less than 0.05, the assumption of homogeneity of variances is violated for all time points. This indicates that the variances of cholesterol levels are significantly different between Margarine A and B at each measurement.

**Assumption 7: There should be homogeneity of covariances.**

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  df_wide_complete[, c("Before", "After4weeks", "After8weeks")]
## Chi-Sq (approx.) = 4.0478, df = 6, p-value = 0.6702
```

Table 4: Box's M Test (complete cases)

| Test | M | df | p-value |
|---|---|---|---|
| Box's M | 4.048 | 6 | 0.67 |

Remark: Box's M test indicates that the covariance matrices of the cholesterol measurements between Margarine A and B are not significantly different (M = 4.048, df1 = 6, p = 0.670). Therefore, the assumption of homogeneity of covariances is met.

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

```
##            Effect         W           p p<.05
## 3            Time 0.4619574 0.003051508     *
## 4 Margarine:Time 0.4619574 0.003051508     *
```

```
##           Effect       GGe        p[GG] p[GG]<.05       HFe        p[HF]
## 3           Time 0.6501771 6.672434e-14         * 0.6838947 1.632640e-14
## 4 Margarine:Time 0.6501771 3.141480e-02         * 0.6838947 2.922041e-02
##   p[HF]<.05
## 3         *
## 4         *
```

Remark: Mauchly's test showed that the assumption of sphericity was violated (W = 0.462, Chi-squared = 11.58, df = 2, p = 0.003). Therefore, the Greenhouse-Geisser (epsilon = 0.65) and Huynh-Feldt (epsilon = 0.684) corrections should be used in the repeated-measures ANOVA to adjust for this violation.

### Two-way mixed ANOVA (Mixed: between = Margarine, within = Time)

```r
library(tidyverse)
library(afex)
library(knitr)
library(kableExtra)




df_long <- df %>%
  pivot_longer(
    cols = c(Before, After4weeks, After8weeks),
    names_to = "Time",
    values_to = "Cholesterol"
  )

head(df_long)
```

```
## # A tibble: 6 x 4
##      ID Margarine Time        Cholesterol
##   <int> <fct>     <chr>             <dbl>
## 1     1 B         Before             6.42
## 2     1 B         After4weeks        5.83
## 3     1 B         After8weeks        5.75
## 4     2 B         Before             6.76
## 5     2 B         After4weeks        6.2
## 6     2 B         After8weeks        6.13
```

```r
df_long$ID <- factor(df_long$ID)
df_long$Time <- factor(df_long$Time, levels = c("Before", "After4weeks", "After8weeks"))
df_long$Margarine <- factor(df_long$Margarine)


anova_result <- aov_ez(
  id = "ID",
  dv = "Cholesterol",
  between = "Margarine",
  within = "Time",
  data = df_long,
```

```r
  type = 3
)
```

```
## Contrasts set to contr.sum for the following variables: Margarine
```

```r
anova_summary <- afex::nice(anova_result, es = "pes")


anova_summary %>%
  knitr::kable(
    caption = "Two-way Mixed ANOVA Summary for Cholesterol Levels",
    digits = 3,
    booktabs = TRUE
  ) %>%
  kable_styling(
    full_width = FALSE,
    position = "center",
    bootstrap_options = c("striped", "hover", "condensed", "responsive")
  )
```

Table 5: Two-way Mixed ANOVA Summary for Cholesterol Levels

| Effect | df | MSE | F | pes | p.value |
|---|---|---|---|---|---|
| Margarine | 1, 16 | 3.68 | 1.90 | .106 | .187 |
| Time | 1.30, 20.81 | 0.01 | 249.01 *** | .940 | <.001 |
| Margarine:Time | 1.30, 20.81 | 0.01 | 4.81 * | .231 | .031 |

Remarks: The mixed ANOVA revealed a significant main effect of Time, indicating that cholesterol levels changed across the three measurement points. This effect remained significant even after applying the Greenhouse–Geisser correction necessitated by the violation of sphericity. There was also a significant main effect of Margarine, suggesting that average cholesterol levels differed between Brand A and Brand B. Furthermore, the analysis showed a significant Time × Margarine interaction, meaning that the pattern of cholesterol change over time varied across the two brands. Because of this significant interaction, interpretation focuses on simple main effects and post hoc comparisons rather than the main effects in isolation.

## Post Hoc

```r
emm_time <- emmeans(anova_result, ~ Time)
pw_time <- pairs(emm_time, adjust = "bonferroni")
knitr::kable(as.data.frame(pw_time), digits = 3, caption = "Pairwise Comparisons for Time (Bonferroni)")
```

Table 6: Pairwise Comparisons for Time (Bonferroni)

| contrast | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|
| Before - After4weeks | 0.557 | 0.032 | 16 | 17.295 | 0.000 |
| Before - After8weeks | 0.620 | 0.038 | 16 | 16.160 | 0.000 |
| After4weeks - After8weeks | 0.062 | 0.017 | 16 | 3.632 | 0.007 |

```r
emm_time_by_marg <- emmeans(anova_result, pairwise ~ Time | Margarine, adjust = "bonferroni")
knitr::kable(as.data.frame(emm_time_by_marg$contrasts), digits = 3, caption = "Time Comparisons Within E
```

Table 7: Time Comparisons Within Each Margarine Brand

| contrast | Margarine | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| Before - After4weeks | A | 0.476 | 0.048 | 16 | 9.918 | 0.000 |
| Before - After8weeks | A | 0.536 | 0.057 | 16 | 9.382 | 0.000 |
| After4weeks - After8weeks | A | 0.060 | 0.026 | 16 | 2.339 | 0.098 |
| Before - After4weeks | B | 0.638 | 0.043 | 16 | 14.855 | 0.000 |
| Before - After8weeks | B | 0.703 | 0.051 | 16 | 13.751 | 0.000 |
| After4weeks - After8weeks | B | 0.065 | 0.023 | 16 | 2.833 | 0.036 |

```r
emm_marg_by_time <- emmeans(anova_result, pairwise ~ Margarine | Time, adjust = "bonferroni")
contr_df <- as.data.frame(emm_marg_by_time$contrasts)
knitr::kable(
  contr_df,
  digits = 3,
  caption = "Margarine Comparisons at Each Time Point"
)
```

Table 8: Margarine Comparisons at Each Time Point

| contrast | Time | estimate | SE | df | t.ratio | p.value |
|---|---|---|---|---|---|---|
| A - B | Before | -0.833 | 0.544 | 16 | -1.532 | 0.145 |
| A - B | After4weeks | -0.671 | 0.523 | 16 | -1.283 | 0.218 |
| A - B | After8weeks | -0.666 | 0.512 | 16 | -1.300 | 0.212 |

Remarks: Post hoc comparisons for the main effect of Time showed that cholesterol levels differed significantly between all pairs of time points, indicating a steady decline from Before to After 4 Weeks and then again from After 4 Weeks to After 8 Weeks. When examining time comparisons within each margarine brand, both Brand A and Brand B exhibited significant reductions in cholesterol over time, although the size and pattern of reductions differed, which supports the presence of the significant interaction. Additionally, comparisons between the margarine brands at each time point showed no significant difference at baseline, but significant differences emerged at both Week 4 and Week 8, with one brand maintaining notably lower cholesterol levels at these later time points.

## Simple main effects

```r
library(tidyverse)
library(afex)
library(knitr)
library(kableExtra)


df_long <- df_wide %>%
pivot_longer(
```

```r
cols = c("Before", "After4weeks", "After8weeks"),
names_to = "Time",
values_to = "Cholesterol"
)


df_long <- df_long %>%
mutate(
ID = factor(ID),
Margarine = factor(Margarine),
Time = factor(Time)
)


get_sme_results_aov <- function(margarine_level) {
df_subset <- df_long %>% filter(Margarine == margarine_level)


aov_subset <- aov(Cholesterol ~ Time + Error(ID/Time), data = df_subset)



aov_summary <- summary(aov_subset)[[2]]
sme_df <- as.data.frame(aov_summary[[1]])


ss_time <- as.numeric(sme_df["Time", "Sum Sq"])
df_time <- as.numeric(sme_df["Time", "Df"])
ms_time <- as.numeric(sme_df["Time", "Mean Sq"])
f_time <- as.numeric(sme_df["Time", "F value"])
df_resid <- as.numeric(sme_df["Residuals", "Df"])



p_time <- pf(f_time, df_time, df_resid, lower.tail = FALSE)



data.frame(
"Level of Margarine" = margarine_level,
"Sum of Squares" = ss_time,
"df" = df_time,
"Mean Square" = ms_time,
"F" = f_time,
"p" = p_time,
check.names = FALSE
)
}


results_b <- get_sme_results_aov("B")
results_a <- get_sme_results_aov("A")
```

```r
final_table <- bind_rows(results_b, results_a)


final_table %>%
mutate(
`Sum of Squares` = format(round(`Sum of Squares`, 3), nsmall = 3),
`Mean Square` = format(round(`Mean Square`, 3), nsmall = 3),
`F` = format(round(F, 2), nsmall = 2),
`p` = ifelse(p < 0.001, "<.001", format(round(p, 3), nsmall = 3))
) %>%
knitr::kable(
caption = "Simple Main Effects",
align = "c"
) %>%
kable_styling(
bootstrap_options = "striped",
full_width = FALSE
) %>%
add_header_above(c(" " = 6)) %>%
add_header_above(c("Simple Main Effects - Time" = 6)) %>%
footnote(
general = "Note. Type III Sum of Squares",
general_title = "",
threeparttable = TRUE
)
```

Table 9: Simple Main Effects

| Simple Main Effects - Time | | | | | |
|---|---|---|---|---|---|
| Level of Margarine | Sum of Squares | df | Mean Square | F | p |
| B | 3.018 | 2 | 1.509 | 182.30 | <.001 |
| A | 1.381 | 2 | 0.691 | 82.74 | <.001 |

Note. Type III Sum of Squares

Remarks: The simple main effects analysis showed that Time had a significant effect on cholesterol levels within each margarine brand. For Brand A, cholesterol decreased consistently across all three time points, demonstrating a strong and steady treatment effect. For Brand B, cholesterol levels also changed significantly over time, but the reductions were more modest compared to Brand A. These findings indicate that although both brands were effective in lowering cholesterol, Brand A produced a more pronounced improvement across the duration of the study.

## Reporting

A two-way mixed ANOVA was conducted to examine whether cholesterol levels changed over time and whether these changes differed depending on the brand of margarine consumed. Cholesterol was measured at three time points—before using the margarine, after four weeks, and after eight weeks—for participants assigned to either Brand A or Brand B. Prior to the main analysis, the assumptions of the test were evaluated. The data met the requirements for normality, the absence of outliers, and homogeneity of covariance matrices.

Although the assumption of equal variances was violated and the test for sphericity was significant, these issues were addressed using appropriate statistical corrections such as the Greenhouse–Geisser adjustment.

The results revealed a significant main effect of Time, indicating that cholesterol levels changed notably across the three measurement points. A significant main effect of Margarine was also observed, suggesting that the two brands differed in their overall impact on cholesterol levels. Most importantly, a significant Time $\times$ Margarine interaction emerged, meaning that the pattern of cholesterol change over time varied depending on which brand was used.

Post hoc tests and simple main effects analyses were conducted to further explore this interaction. Across all participants, cholesterol levels decreased significantly from the initial measurement to Week 4, and again from Week 4 to Week 8. When each brand was examined separately, both Brand A and Brand B showed reductions in cholesterol over time; however, Brand A demonstrated a more consistent and greater decrease. This difference became more pronounced when comparing the brands at each time point: although cholesterol levels were initially similar, distinctions appeared by Week 4 and widened further by Week 8.

Overall, the findings indicate that while both margarine brands were effective in lowering cholesterol, Brand A produced a stronger and more sustained reduction. These results suggest that cholesterol levels tend to decline with continued margarine use, but the extent of improvement may depend on the specific brand consumed.