Khafaji, Mostafa
Rodillas, Christian Miguel

FA 10

Two-Way Mixed ANOVA (BW)

Assumptions:

1. You have a continuous dependent variable.
2. You have one between-subjects factor (i.e., independent variable) that is categorical with two or more categories.
3. You have one within-subjects factor (i.e., independent variable) that is categorical with two or more categories.
4. There should be no significant outliers in any level of the within-subjects factor.
5. Your dependent variable should be approximately normally distributed for each level of the within-subjects factor.
6. The variance of your dependent variable should be equal between the groups of the between-subjects factor.
7. There should be homogeneity of covariances.
8. The variance of the differences between groups should be equal.

Null and Alternative Hypotheses

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

Alternative Hypothesis: There is a significant difference in the cholesterol levels between the two brands of margarine over the three time points.

Dataset and Problem

This analysis utilizes R to explore and investigate the connection of 2 margarine brands and time points, at before, after 4 weeks, and after 8 weeks, to their respective cholesterol levels. We aim to determine if there are underlying statistical differences in the people's cholesterol level within the given time stamps with the independent group (Brand of Margarine) and if there is any interaction effect between the two groups.

Assumptions:

Assumption #1: You have a continuous dependent variable.

Remark. The dataset regarding cholesterol has a dependent variable, the cholesterol level. The stated variable is dependent on the Brand of margarine consumed, and is measured in a continuous level. This satisfies the assumption #1.

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

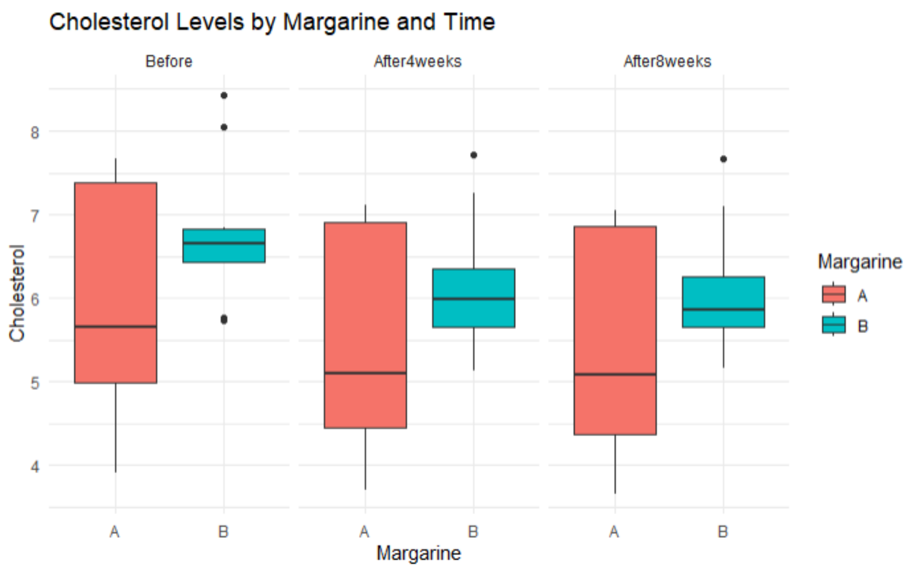
Remark. In the dataset, there is a singular between subject factor, the margarine, and has two categorical groups, which is given as its brand. This This satisfies the assumption #2.

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, the time with an interval of 4 weeks, is made with 3 time intervals that of the same subjects, meeting the assumption that the within-subjects factor should consist of at least two categorical, "related groups" or "matched pairs".

Assumption #4: There should be no significant outliers in any level of the within-subjects factor.

```{r}
ggplot(long_data, aes(x = Margarine, y = Cholesterol, fill = Margarine)) + geom_boxplot() + facet_wrap(~ Time) + theme_minimal() +
  labs(title = "Cholesterol Levels by Margarine and Time", y = "Cholesterol")
```



Remark. There were identifiable significant outliers in the cholesterol levels over time within the brand Margarine B, as assessed by visual inspection of boxplots. Therefore, Assumption #4 is violated.

Assumption #5: Your dependent variable should be approximately normally distributed for each level of the within-subjects factor.

The Descriptive Statistics given are filtered by the time and the brand of Margarine:

Margarine <fctr>	Time <fctr>	Valid <int>	Mode <chr>	Median <dbl>	Mean <dbl>	Std. Deviation <dbl>	Variance <dbl>	Skewness <dbl>	Std. Error of Skewness <dbl>
A	Before	8	3.91	5.650	5.94500	1.4281256	2.0395429	0.01545911	0.7521014
A	After4weeks	8	3.7	5.095	5.46875	1.3876026	1.9254411	0.13111186	0.7521014
A	After8weeks	8	3.66	5.090	5.40875	1.3737065	1.8870696	0.11278029	0.7521014
B	Before	10	5.73	6.660	6.77800	0.8664718	0.7507733	0.66052144	0.6870429
B	After4weeks	10	5.83	5.990	6.14000	0.8145892	0.6635556	0.63406333	0.6870429
B	After8weeks	10	5.17	5.855	6.07500	0.7788346	0.6065833	0.80926920	0.6870429

Kurtosis <dbl>	Std. Error of Kurtosis <dbl>	Minimum <dbl>	Maximum <dbl>	25th Percentile <dbl>	50th Percentile <dbl>	90th Percentile <dbl>
-1.8492904	1.732051	3.91	7.67	4.9875	5.650	7.544
-1.9418343	1.732051	3.70	7.12	4.4575	5.095	7.113
-1.9462760	1.732051	3.66	7.05	4.3750	5.090	6.987
-0.8113954	1.549193	5.73	8.43	6.4250	6.660	8.088
-0.9051956	1.549193	5.13	7.71	5.6500	5.990	7.296
-0.6567688	1.549193	5.17	7.67	5.6575	5.855	7.157

For the p-value:

```{r}
normality_results <- long_data %>%
  group_by(Margarine, Time) %>%
  summarize(p_value = shapiro.test(Cholesterol)$p.value)

print(normality_results)
```

R Console | grouped_df 6 x 3

A tibble: 6 × 3 | Groups: Margarine [2]

Margarine <fctr>	Time <fctr>	p_value <dbl>
A	Before	0.2921657
A	After4weeks	0.1543526
A	After8weeks	0.1738318
B	Before	0.1338272
B	After4weeks	0.4000403
B	After8weeks	0.2158757

6 rows

Remark. The dependent variable, the cholesterol levels, is approximately normally distributed for each combination of their time to the respective brand of margarine consumed. And as assessed by the Shapiro-Wilk test of normality, ($p > 0.05$), all fall within the prescribed basis of greater than 0.05.

Assumption #6: The variance of your dependent variable should be equal between the groups of the between-subjects factor.

```{r}
library(car)
levene_test_check <- leveneTest(Cholesterol ~ Margarine, data = long_data)
print(levene_test_check)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value   Pr(>F)
group  1  10.505 0.002079 **
      52
```

Remark. As assessed by Levene's test of equality of variances ($p < 0.05$), then the variances for each combination of the margarine with the respective cholesterol level per the time given is non-homogenous as p-value = 0.002079. Therefore, assumption 6 is violated.

Since assumption 6 has been violated, we can therefore reject the null hypothesis, and infer that there is a significant difference in the cholesterol levels between the two brands of margarine over the three time points.

Assumption #7: There should be homogeneity of covariances.

Remark. Since the null has already been rejected prior to assumption #7, this has been nullified.

Assumption #8: The variance of the differences between groups should be equal.

Remark. Since the null has already been rejected prior to assumption #, this has been nullified.

To proceed, we can model our analysis as the following:

- Friedman test to check for significant difference between the time points for each individual.
- Mann-Whitney U test to check for significant difference between margarine groups within each time point.

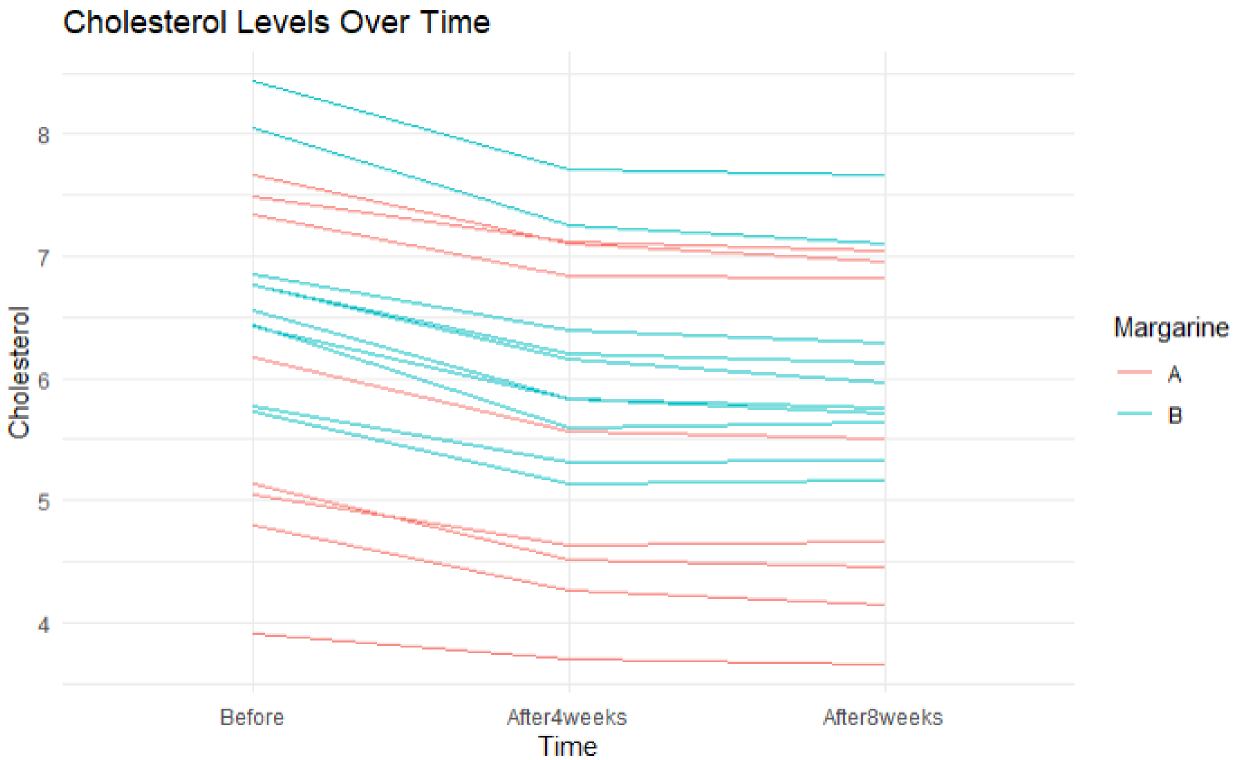
Friedman Test:

```
```{r}
friedman_test_result <- friedman.test(Cholesterol ~ Time | ID, data = long_data)
print(friedman_test_result)
```
```

```
        Friedman rank sum test

data:  Cholesterol and Time and ID
Friedman chi-squared = 29.778, df = 2, p-value = 3.419e-07
```

Remark. What the p-value at $3.419x10^{-7}$ It shows that there is an underlying significant difference between the given time points for each individual's cholesterol level, as assessed by the Friedman Test ($p < 0.05$). Thus, there is strong evidence that cholesterol levels differ significantly across the time points. Attached below is a visual representation of the cholesterol levels across different IDs.



Mann-Whitney Wilcoxon Test:

```{r}
mann_whitney_test_b4 <- wilcox.test(Cholesterol ~ Margarine, data = long_data %>% filter(Time == "Before"))
print(mann_whitney_test_b4)

mann_whitney_after4weeks <- wilcox.test(Cholesterol ~ Margarine, data = long_data %>% filter(Time == "After4weeks"))
print("Mann-Whitney U test for 'After4weeks':")
print(mann_whitney_after4weeks)

mann_whitney_after8weeks <- wilcox.test(Cholesterol ~ Margarine, data = long_data %>% filter(Time == "After8weeks"))
print("Mann-Whitney U test for 'After8weeks':")
print(mann_whitney_after8weeks)
```

```
	Wilcoxon rank sum exact test

data:  Cholesterol by Margarine
W = 26, p-value = 0.237
alternative hypothesis: true location shift is not equal to 0

Warning: cannot compute exact p-value with ties[1] "Mann-Whitney U test for 'After4weeks':"

	Wilcoxon rank sum test with continuity correction

data:  Cholesterol by Margarine
W = 26, p-value = 0.2301
alternative hypothesis: true location shift is not equal to 0

[1] "Mann-Whitney U test for 'After8weeks':"

	Wilcoxon rank sum exact test

data:  Cholesterol by Margarine
W = 26, p-value = 0.237
alternative hypothesis: true location shift is not equal to 0
```

Remark. The Mann-Whitney U test, or the Wilcoxon rank sum test, shows that there is no significant difference in the cholesterol levels of those that consume Margarine A and Margarine B, across all measurement for each 4 week interval, i.e. for the initial measurement (W = 26, p = 0.237), after 4 weeks (W = 26, p = 0.2301), and after 8 weeks (W = 26, p = 0.237). Thus, there is ultimately no significant difference between the two types of margarine regarding its effect on the cholesterol levels of its consumers.

Post-Hoc:

```
```{r}
pairwise.wilcox.test(long_data$Cholesterol, long_data$Time, paired = TRUE, p.adjust.method = "bonferroni")
```
```

```
Warning: cannot compute exact p-value with tiesWarning: cannot compute exact p-value with ties
        Pairwise comparisons using Wilcoxon signed rank test with continuity correction

data:  long_data$Cholesterol and long_data$Time

            Before  After4weeks
After4weeks 0.00064 -
After8weeks 2.3e-05 0.01117

P value adjustment method: bonferroni
```

Remark. The values under the Bonferroni continuity correction indicate that there are significant differences ($p < 0.05$) between the comparison of each time point. Particularly, the initial cholesterol level measurement was significantly higher than the measurement after 4 weeks (p = 0.00064), and after 8 weeks (p = 2.3e-5). The measurement after 4 weeks also differed greatly from the measurement taken after 8 weeks (p = 0.01117).

Reporting:

The Friedman rank sum test and the Mann-Whitney U Test was used to determine if there was a significant difference in the cholesterol levels of consumers of different kinds of margarine, and if there was an increase of their cholesterol levels over time.The groups of margarine consumers are those that consumed Margarine A (n = 8), and Margarine B (n =10). Their cholesterol levels were then taken over an 8 week interval (before, after 4 weeks, after 8 weeks). It was found that there were 2 significant outliers in those that consumed Margarine B, during the initial measurement. Using the Shapiro-Wilk Test, it was found that the combination of all groups were normally distributed($p>0.05$). However, the levene's test revealed that there was no homogeneity of variances of the dependent variable for all physical activity groups ($p = 0.002079$).

After failing the assumptions for a Two-Way Mixed Anova testing, the Friedman rank sum test and the Mann-Whitney U test, non-parametric alternatives of the Repeated Measures Anova and the Independent Samples T-Test, respectively, were immediately used.

Testing for a significant difference between the means of the cholesterol level triples taken in a time period of 8 weeks while consuming margarine (n=18). The Friedman rank sum test showed an astounding difference in cholesterol levels across the time period and their intervals ($\chi2 = 29.778$, df = 2, $p = 3.14e\text{-}7$). It was found that the cholesterol level decreased from Before ($M_{\text{Margarine A}} = 5.94$, $SD_{\text{Margarine A}} = 1.43$; $M_{\text{Margarine B}} = 6.78$, $SD_{\text{Margarine B}} = 0.866$), after 4 weeks ($M_{\text{Margarine A}} = 5.47$, $SD_{\text{Margarine A}} = 1.39$; $M_{\text{Margarine B}} = 6.14$, $SD_{\text{Margarine B}} = 0.815$), and after 8 weeks ($M_{\text{Margarine A}} = 5.41$, $SD_{\text{Margarine A}} = 1.37$; $M_{\text{Margarine B}} = 6.08$, $SD_{\text{Margarine B}} = 0.779$).

Post Hoc Analysis then shows that the initial cholesterol level measurement was significantly higher than the measurement after 4 weeks ($p = 0.00064$), and after 8 weeks ($p = 2.3e\text{-}5$). The measurement after 4 weeks also differed greatly from the measurement taken after 8 weeks ($p = 0.01117$).

Mann-Whitney U test was used to determine if there is a significant difference between the two groups of margarine, in each time interval. There was no significant difference between the cholesterol level of the two types of margarine in the before time interval ($W = 26$, $p = 0.237$). This continues in the measurement of cholesterol level after 4 weeks ($W = 26$, $p = 0.2301$), and in after 8 weeks ($W = 26$, $p = 0.237$).

To conclude, the test shows that margarine is margarine, i.e. there is no significant difference between the brand or type of margarine, at least in the case of Margarine A and Margarine B. However, statistical tests have found that the cholesterol level of the participants, who consumed margarine over the course of 8 weeks, decreased along that time period.