

# Formative Assessment 2 | MANOVA

Baybayon, Darlyn Antoinette B.

```
suppressPackageStartupMessages({  
  library(tidyverse)  
  library(readr)  
  library(ggplot2)  
  library(dplyr)  
  library(MVN)  
  library(mice)  
  library(heplots)  
})
```

## The Data

```
df <- read_csv("manova_mancova_practice.csv", show_col_types = FALSE)  
head(df)
```

```
## # A tibble: 6 x 12  
##   ID      Treatment School Gender   Age SES_Index Pre_Math Pre_Reading Pre_Science  
##   <chr> <chr>      <chr> <chr> <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 C000~ Control   East   F    17.1     0.163     50.3     64.1     63.8  
## 2 C001~ Control   North  F    15.9     0.586     41.9     54.7     53.1  
## 3 C002~ Control   West   F    17.5     0.711     54.9     60.7     37.3  
## 4 C003~ Control   North  M    17.6     0.793     65.8     55.8     58.9  
## 5 C004~ Control   North  F    15.0    -0.349     63.2     60.3     62  
## 6 C005~ Control   West   M    15.6    -0.462     53.9     58.1     56.2  
## # i 3 more variables: Post_Math <dbl>, Post_Reading <dbl>, Post_Science <dbl>
```

The dataset contains multivariate data of students from different schools. It has 12 features including each student's ID, Treatment group (Instructional method: Control, MethodA, MethodB), School, Gender, Age, SES\_index, Pre-test Scores, and Post-test Scores.

## Objective

Students are grouped into three instructional methods, Control, MethodA, and MethodB. They were tested in Math, Reading, and Science before and after implementing the instructional methods. To examine whether the teaching methods significantly affect students' performance on the three subjects, multivariate analyses of variance (MANOVA) will be performed. To determine if the results hold after accounting for their initial performance, their pre-test scores will be used as covariates in a multivariate analysis of covariance (MANCOVA).

## Data Preprocessing

### Convert data types

```
df <- df %>%
  select(-"ID") %>%
  mutate(across(c("Treatment", "Gender", "School"), as.factor))
```

## Check nulls and handle them

```
colSums(is.na(df))/nrow(df)*100
```

```
##      Treatment      School      Gender      Age      SES_Index      Pre_Math
##      0.000000      0.000000      0.000000      0.000000      4.891304      2.717391
## Pre_Reading Pre_Science Post_Math Post_Reading Post_Science
##      3.260870      4.347826      4.347826      5.434783      2.173913
```

There are some missing data in numeric columns. According to the given design of the data, the missingness is MCAR, so we can impute the missing values with a simple mean imputation.

```
df_imputed <- df %>%
  mutate(
    SES_Index = ifelse(is.na(SES_Index), mean(SES_Index, na.rm=TRUE), SES_Index),
    Pre_Math = ifelse(is.na(Pre_Math), mean(Pre_Math, na.rm=TRUE), Pre_Math),
    Pre_Reading = ifelse(is.na(Pre_Reading), mean(Pre_Reading, na.rm=TRUE), Pre_Reading),
    Pre_Science = ifelse(is.na(Pre_Science), mean(Pre_Science, na.rm=TRUE), Pre_Science),
    Post_Math = ifelse(is.na(Post_Math), mean(Post_Math, na.rm=TRUE), Post_Math),
    Post_Reading = ifelse(is.na(Post_Reading), mean(Post_Reading, na.rm=TRUE), Post_Reading),
    Post_Science = ifelse(is.na(Post_Science), mean(Post_Science, na.rm=TRUE), Post_Science),
  )
colSums(is.na(df_imputed))
```

```
##      Treatment      School      Gender      Age      SES_Index      Pre_Math
##           0           0           0           0           0           0
## Pre_Reading Pre_Science Post_Math Post_Reading Post_Science
##           0           0           0           0           0
```

Mean and variance after imputation

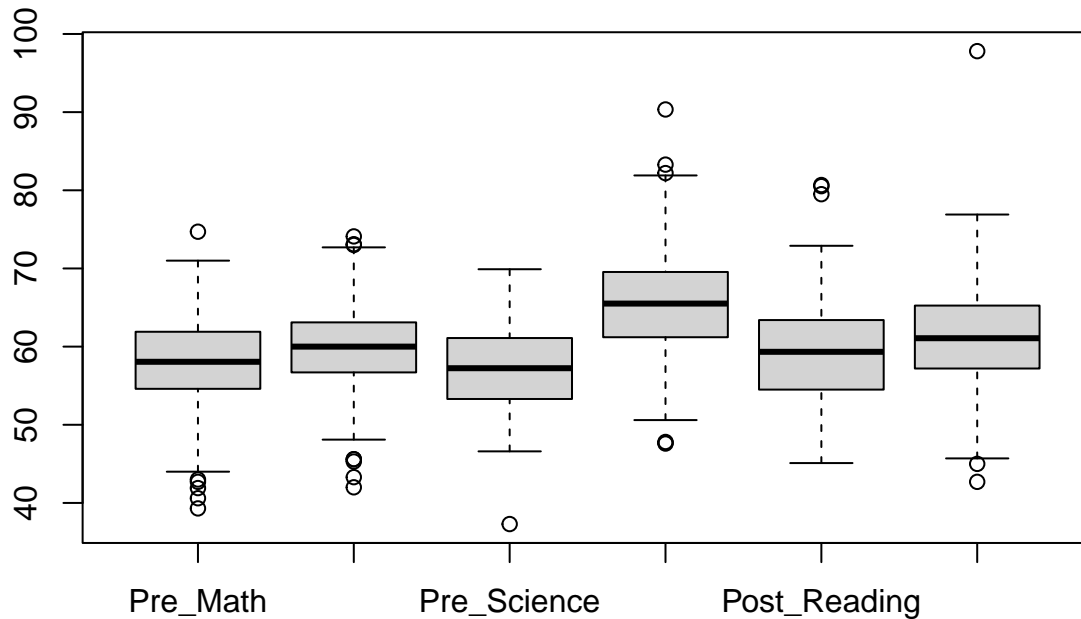
```
round(data.frame(mean = sapply(select(df, where(is.numeric)), mean, na.rm=TRUE),
  mean_imp = sapply(select(df_imputed, where(is.numeric)), mean),
  var = sapply(select(df, where(is.numeric)), var, , na.rm=TRUE),
  var_imp = sapply(select(df_imputed, where(is.numeric)), var)), 2)
```

```
##      mean mean_imp  var var_imp
## Age      16.80    16.80 0.77   0.77
## SES_Index -0.03    -0.03 0.86   0.82
## Pre_Math   58.06    58.06 38.60  37.54
## Pre_Reading 59.65    59.65 32.78  31.70
## Pre_Science 57.24    57.24 27.79  26.57
## Post_Math   65.51    65.51 53.52  51.18
## Post_Reading 59.34    59.34 43.73  41.34
## Post_Science 61.08    61.08 54.94  53.74
```

## Check and handle outliers

*Univariate outliers*

```
boxplot(df_imputed[6:11])
```



```
lapply(select(df_imputed[6:11], where(is.numeric)), function(x) boxplot.stats(x)$out)
```

```
## $Pre_Math
## [1] 41.9 43.0 40.6 39.3 42.7 74.7
##
## $Pre_Reading
## [1] 73.0 45.3 43.3 42.0 45.6 74.1 73.1 45.6
##
## $Pre_Science
## [1] 37.3
##
## $Post_Math
## [1] 47.80000 47.70000 47.60000 82.20000 90.34177 83.28444
##
## $Post_Reading
## [1] 80.67652 79.50072 80.52707
##
## $Post_Science
## [1] 45.00000 42.70000 97.80108
```

By visual inspection of boxplots and Tukey's rule, outliers for each test score distribution were identified. Univariate outliers may affect the data's shape and may potentially lead to issues with multivariate normality which is a crucial assumption for MANOVA. Further investigation of multivariate outliers will be conducted before proceeding with MANOVA. If violations occur, transformations will be considered to address the assumption and ensure the validity of the analysis

### Multivariate outliers

```
# Post-Test
post <- df_imputed[,9:11]
post$dist <- mahalanobis(post, colMeans(post), cov(post))
threshold <- qchisq(1 - 0.001, df = 3)
outliers <- which(post$dist > threshold)
outliers
```

```
## [1] 114
```

```
# Pre-Test
pre <- df_imputed[,6:8]
pre$dist <- mahalanobis(pre, colMeans(pre), cov(pre))
threshold <- qchisq(1 - 0.001, df = 3)
outliers <- which(pre$dist > threshold)
outliers
```

```
## [1] 3 68
```

```
df_no_outliers <- df_imputed[-c(114,3,68),]
```

Mahalanobis distance was computed separately for the three pre-test and three post-test dependent variables to identify multivariate outliers. Using a chi-square cutoff at  $p < .001$  ( $df = 3$ ), one case was flagged as a multivariate outlier in the post-test group while two cases were found in the pre-test group. Since MANOVA assumes multivariate normality, this outlier may be problematic as it may distort the data's center and shape leading to inaccurate conclusions. To preserve this assumption, these outliers will be excluded from the final dataset, resulting in 181 cases for MANOVA and MANCOVA analyses.

## MANOVA

**Conduct a MANOVA to test whether the three dependent variables (PostTest1, PostTest2, PostTest3) differ by Teaching Method**

Treatment → DVs = Post\_Math, Post\_Reading, Post\_Science.

### MANOVA Result

```
manova_result <- manova(
  cbind(Post_Math, Post_Reading, Post_Science) ~ Treatment,
  data = df_no_outliers
)
summary(manova_result)
```

```
##           Df Pillai approx F num Df den Df    Pr(>F)
## Treatment   2 0.0961    2.978      6   354 0.007521 **
## Residuals 178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A one-way multivariate analysis of variance (MANOVA) was conducted to examine the effect of Treatment on scores in Math, Reading, and Science. The results indicate a statistically significant difference in the mean test scores across the different treatment groups (Pillai = 0.096,  $F(6, 354) = 2.978$ ,  $p = 0.008$ ). Thus, we reject the null hypothesis, and conclude that the instructional method has a significant effect on the combined outcomes of the scores.

### Test Statistics

```
wilks <- summary(manova_result, test ="Wilks")
pillai <- summary(manova_result, test ="Pillai")
roy <- summary(manova_result, test ="Roy")
lh<- summary(manova_result, test ="Hotelling-Lawley")

data.frame( Wilks = wilks$stats[1,2],
             Pillai = pillai$stats[1,2],
             Lawley_Hotell = lh$stats[1,2],
             Roy = roy$stats[1,2])
```

```
##           Wilks      Pillai Lawley_Hotell      Roy
## 1 0.9052173 0.09609952    0.1032525 0.08641966
```

The Wilks' Lambda quantifies how much of the variance in the combination of the dependent variables is not explained by the independent variable. The Wilks' Lambda of 0.905 indicates that 90.5% of the variance in test scores is not explained by the treatment group, suggesting a weak effect.

The Pillai's Trace measures how much of the multivariate variance is explained by the independent variable. The value of 0.096 indicates that about 9.6% of the variance in test scores is explained by the treatment group. Similar to Wilks' Lambda, this suggests that the treatment has a weak effect on the scores.

The Lawley–Hotelling Trace value of 0.103 similarly indicates that the treatment explains a low percentage of the combined variance across the test measures, again pointing to a small effect.

The Roy's Largest Root value of 0.086 reflects the maximum variance explained in a single dimension where group differences are strongest. This also suggests a weak effect of the treatment on test scores.

## Follow-up ANOVA

```
summary.aov(manova_result)
```

```
## Response Post_Math :
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    2  123.2   61.621   1.3225 0.2691
## Residuals   178 8294.0   46.595
##
## Response Post_Reading :
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment    2  237.4  118.718   3.0554 0.04958 *
## Residuals   178 6916.3   38.856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response Post_Science :
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Treatment    2  567.9  283.956   6.3962 0.002077 **
## Residuals   178 7902.2   44.395
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

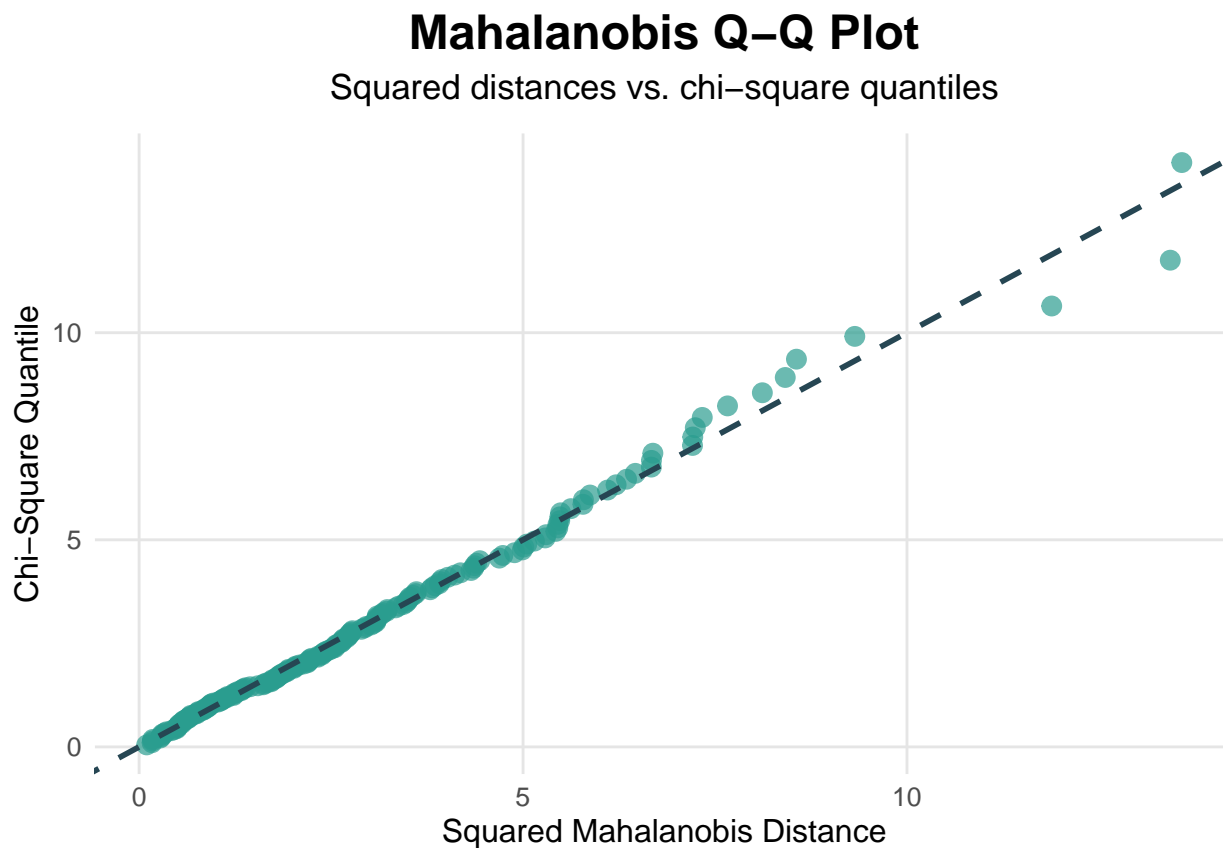
To further examine group differences, separate univariate ANOVAs were conducted for each post-test score. Results showed that Post-Science scores differed significantly between treatment groups ( $p = 0.0021$ ), as did Post-Reading scores ( $p = 0.0496$ ). In contrast, Post-Math scores did not show a significant difference across groups. According to the results of MANOVA and follow-up ANOVAs, the teaching methods have significant but minimal effect on

overall student performance. The effect was significant for Science and Reading but not so much for Math. As Math scores did not seem to differ significantly between different instructional methods, student performance in Math may be less sensitive to changes in teaching style.

## Check assumptions

### Multivariate Normality

```
multivariate_diagnostic_plot(df_no_outliers[,9:11], type="qq")
```



The Q-Q plot shows that most points align with the theoretical quantiles with some observations in the upper tail deviating slightly from the diagonal reference line. This indicates good agreement with the expected normal distribution. Hence, based on the Q-Q plot of Mahalanobis distances, the assumption of multivariate normality appears to be satisfied.

```
mardia(df_no_outliers[,9:11])
```

```
##           Test Statistic    p.value    Method
## 1 Mardia Skewness 18.0245546 0.05455085 asymptotic
## 2 Mardia Kurtosis -0.2374066 0.81234140 asymptotic
```

A Mardia's SKewness and Kurtosis test was conducted to confirm if the dependent variables follow multivariate normality. The skewness statistic was 17.192 ( $p = 0.07$ ) and the kurtosis statistic was -0.172 ( $p = 0.863$ ). Since both p-values exceed the common significance level of 0.05, we can conclude that there were no significant skewness and excess kurtosis. The Mardia's test results support our observation in the Q-Q plot, so we can reasonably assume that the post test scores follow a multivariate normal distribution.

### Homogeneity of Variance-Covariance Matrices

```
boxM(cbind(Post_Math, Post_Reading, Post_Science) ~ Treatment,
     data = df_no_outliers)
```

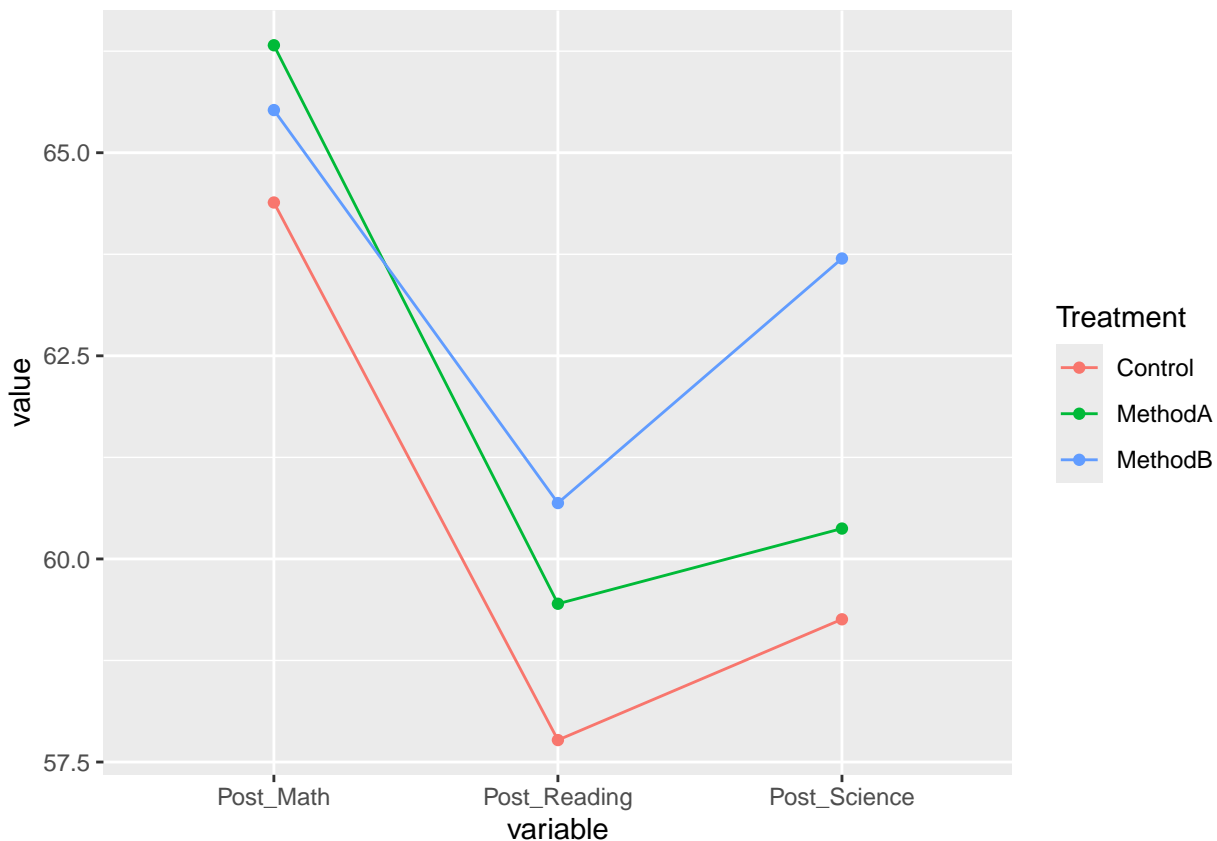
```
##
## Box's M-test for Homogeneity of Covariance Matrices
##
## data: df_no_outliers
## Chi-Sq (approx.) = 8.834, df = 12, p-value = 0.717
```

The Box's M test did not reveal a significant difference ( $p > 0.05$ ) in covariance matrices across the treatment groups. Therefore the assumption of homogeneity of covariance matrices is met and the MANOVA is valid.

## Profile Plots

```
fitted_vals <- fitted(manova_result)
fitted_means <- aggregate(fitted_vals, by = list(Treatment = df_no_outliers$Treatment), FUN = mean)
fitted_long <- reshape2::melt(fitted_means, id.vars = "Treatment")

ggplot(fitted_long, aes(x = variable, y = value, group = Treatment, color = Treatment)) +
  geom_line() + geom_point()
```



```
df_no_outliers %>%
  group_by(Treatment) %>%
  summarise(
    Mean_Science = mean(Post_Science, na.rm = TRUE),
```

```

Mean_Reading = mean(Post_Reading, na.rm = TRUE),
Mean_Math = mean(Post_Math, na.rm = TRUE),
.groups = "drop"
)

```

```

## # A tibble: 3 x 4
##   Treatment Mean_Science Mean_Reading Mean_Math
##   <fct>         <dbl>         <dbl>         <dbl>
## 1 Control          59.3          57.8          64.4
## 2 MethodA          60.4          59.4          66.3
## 3 MethodB          63.7          60.7          65.5

```

The profile plots compares the average performance of students from each instructional methods in post-tests in Math, Reading, and Science. All treatment groups perform relatively well in Math, all groups also get their lowest average scores in Reading. MethodA students performed best in Math while MethodB students performed best in Reading and Science. Control group consistently lagged below other groups for all three subjects. Overall, the group profiles are not parallel as the difference between the groups are not constant across the subject scores. This plot provides a visual support of the findings from the MANOVA, showing that the scores across treatment groups are significantly different. It also visualizes how close the Math scores are between groups compared to Reading and Science scores as shown in the follow up univariate ANOVA results.

## Two-Way MANOVA

**Test for the effect of Teaching Method (3 levels), Gender (2 levels), and their interaction on the three post-test variables**

```

manova_2w_result <- manova(
  cbind(Post_Math, Post_Reading, Post_Science) ~ Treatment*Gender,
  data = df_no_outliers
)

summary(manova_2w_result)

```

```

##              Df    Pillai approx F num Df den Df  Pr(>F)
## Treatment      2 0.096677  2.94603     6    348 0.00811 **
## Gender          1 0.024825  1.46800     3    173 0.22499
## Treatment:Gender 2 0.011928  0.34799     6    348 0.91089
## Residuals      175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A Two-Way MANOVA was conducted to test for the effect of Treatment, Gender, and their interaction on the three post-test scores in Math, reading, and Science. The results indicate statistically significant difference in the mean test scores across the treatment groups (Pillai = 0.097,  $F(6, 348) = 2.946$ ,  $p = 0.008$ ). Gender (Pillai = 0.025,  $F(3, 183) = 1.468$ ,  $p = 0.225$ ) main effect was not statistically significant indicating that the overall combined test scores did not differ significantly between male and female students. Gender's interaction with Treatment (Pillai = 0.012,  $F(6, 348) = 0.348$ ,  $p = 0.911$ ) also did not reach statistical significance, suggesting that the effect of different instructional methods did not depend on the student's gender. The differences in performance between the Treatment groups were consistent for male and female students (i.e. if MethodB performed better for males, this is also true for females).



## MANCOVA

Re-run the MANOVA, but now include Pre-Test scores (Pre1, Pre2, Pre3) as covariates.

```
mancova_result <- manova(
  cbind(Post_Math, Post_Reading, Post_Science) ~ Treatment*Gender + Pre_Math+Pre_Reading+ Pre_Science,
  data = df_no_outliers
)
summary(mancova_result)
```

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
## Treatment	2	0.120758	3.6628	6	342	0.00154 **
## Gender	1	0.029397	1.7163	3	170	0.16552
## Pre_Math	1	0.291192	23.2797	3	170	1.128e-12 ***
## Pre_Reading	1	0.239556	17.8512	3	170	4.049e-10 ***
## Pre_Science	1	0.145715	9.6656	3	170	6.327e-06 ***
## Treatment:Gender	2	0.011310	0.3242	6	342	0.92416
## Residuals	172					
## ---						
## Signif. codes:	0	'***'	0.001	'**'	0.01	'*' 0.05 '.' 0.1 ' ' 1

A multivariate analysis of covariance (MANCOVA) was conducted to assess the effect of teaching method, gender, and their interaction on students' post-test scores in Math, Reading, and Science, after adjusting for pre-test performance. The results confirm that there was still a statistically significant multivariate effect of teaching method on post-test scores, after adjusting for pre-test scores (Pillai's Trace = 0.120,  $F(6, 348) = 3.698$ ,  $p = 0.001$ ). The MANOVA results also hold for Gender and its interaction with Treatment as these effects did not reach statistical significance in this MANCOVA as well. Furthermore, the covariates (Pre\_Math, Pre\_Reading, and Pre\_Science) were all statistically significant predictors of post-test outcomes (all  $p < 0.05$ ), indicating that students' pre-test scores strongly influenced their post-test scores.

## Follow-up ANCOVA

```
summary.aov(mancova_result)
```

```
## Response Post_Math :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment    2  123.2   61.62  1.7954  0.16916
## Gender        1   37.2   37.19  1.0835  0.29938
## Pre_Math      1 2212.4 2212.37 64.4609 1.484e-13 ***
## Pre_Reading   1   19.1   19.08  0.5560  0.45689
## Pre_Science   1  100.5  100.47  2.9275  0.08889 .
## Treatment:Gender  2   21.6   10.81  0.3150  0.73018
## Residuals    172 5903.2   34.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response Post_Reading :
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment    2  237.4  118.72  4.3264  0.01468 *
## Gender        1   28.8   28.85  1.0514  0.30663
## Pre_Math      1  584.5  584.55 21.3025 7.642e-06 ***
## Pre_Reading   1 1413.9 1413.90 51.5264 2.030e-11 ***
## Pre_Science   1  124.5  124.46  4.5357  0.03462 *
```

```

## Treatment:Gender    2    44.8    22.42    0.8171    0.44341
## Residuals          172 4719.7    27.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response Post_Science :
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment      2   567.9   283.96   7.5177 0.0007413 ***
## Gender          1    70.3    70.30   1.8612 0.1742675
## Pre_Math        1   173.9   173.89   4.6036 0.0333074 *
## Pre_Reading      1    45.2    45.17   1.1960 0.2756564
## Pre_Science      1 1097.0 1096.99 29.0426 2.305e-07 ***
## Treatment:Gender  2    19.1     9.56   0.2531 0.7766896
## Residuals       172 6496.8    37.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Follow-up univariate ANCOVAs were conducted to examine the effect of Treatment on each Post-Test subject score individually, while controlling for Pre-Test scores in Math, Reading, and Science. The results indicate that teaching method had a statistically significant effect on Post\_Reading scores ( $F(2,172) = 4.326$ ,  $p = 0.015$ ) and Post\_Science scores ( $F(2,172) = 7.512$ ,  $p = 0.0007$ ), but not on Post\_Math scores ( $F(2,172) = 1.795$ ,  $p = 0.169$ ). These results indicate that the teaching method significantly influenced students' performance in Reading and Science, but not in Math, after adjusting for prior scores. Gender and its interaction with treatment also did not significantly affect student outcomes. Among the covariates, Pre\_Math was a significant predictor for all three post-test scores, while Pre\_Reading was a significant predictor for Post\_Reading score only, and Pre\_Science significantly predicted Post\_Science score. This suggests that students' initial performance, particularly in the same subject, played a key role in their post-test results.

## Summary

Based on the analyses conducted, it's clear that the instructional methods have a statistically significant effect on student outcomes. The initial one-way MANOVA result ( $p = 0.008$ ) confirms a significant multivariate effect of instructional method on the combined post-test scores in Math, Reading, and Science. The follow-up univariate ANOVAs further reveal that this difference is specifically significant for Post-Reading ( $p = 0.0496$ ) and Post-Science ( $p = 0.0021$ ) scores, but not for Post-Math scores ( $p > 0.05$ ). The profile plots visually support this, showing distinct, non-parallel lines for each group.

Gender does not appear to modify the outcomes. The Two-Way MANOVA showed no statistically significant main effect for Gender ( $p = 0.225$ ), meaning male and female students had similar overall performance. Furthermore, the interaction between Treatment and Gender was also not significant ( $p = 0.911$ ), indicating that the effects of the teaching methods was consistent for both male and female students.

The significant effects of the teaching methods hold even after adjusting for pre-test scores. The MANCOVA showed a significant multivariate effect of Treatment ( $p = 0.001$ ) after controlling for students' pre-test scores. The follow-up ANCOVAs confirmed that the teaching methods had a significant impact on Post-Reading ( $p=0.015$ ) and Post-Science ( $p=0.0007$ ) scores, but not on Post-Math scores ( $p=0.165$ ), when accounting for initial performance. Moreover, the students' pre-test scores were revealed to be strong predictors of their post-test outcomes, especially within the same subject areas.

These findings suggest that the instructional strategies do not seem to have similar effects across all subject areas. Teaching methods effective in Reading and Science may not necessarily be as effective in Math. Educators must note that teaching strategies may need to be tailored to specific subjects accordingly. Moreover, differences in performance between genders were not statistically significant, indicating no need for teaching methods to be adjusted based on gender. Additionally, since pre-test scores were highly predictive of students' post-test scores, incorporating diagnostic assessments early in the academic year would be beneficial to help evaluate students' initial skill levels and adjust teaching strategies as necessary. Identifying learning gaps early on allows teachers to provide

additional support where needed, ensuring that the learning needs of students are satisfied thereby maximizing their potential.

Conducting separate univariate ANOVAs only would lead us to miss important findings about the true effect of the instructional methods. Since performance in the three subjects are likely correlated, analyzing them separately ignores the underlying relationships between these variables. Furthermore, the MANOVA revealed significant overall effect of the teaching method. Proceeding to follow-up ANOVAs allowed us to examine the treatment effects on individual subjects. However, if only univariate ANOVAs had been conducted, we might have missed that the instructional methods had a statistically significant multivariate effect, even if not all subjects strongly differed under different treatments. Additionally, conducting the MANCOVA provided further insight on the importance of pre-test scores. It confirmed that even after adjusting the pre-test influence, the teaching methods still significantly affected outcomes. It was also found that initial performance was a great predictor of their post-test performance in the same subject.