Case

Hebert Frangiotti

2024-02-06

# Starting the Data

```{r} library(readxl)

file\_path <- “C:\Users\Admin\Downloads\Data\_performance.xlsx” explicit\_group\_data <- read\_excel(file\_path, sheet = “Explicit group”) implicit\_group\_data <- read\_excel(file\_path, sheet = “Implicit group”)

# For the explicit\_group\_data dataframe

colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Score in the selection test”] <- “Score sel.” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Score reading test before”] <- “read. bef” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Score reading test after”] <- “read. aft” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Number of aspects observed before”] <- “aspects bef” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Number of aspects observed after”] <- “aspects aft” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Oral score before”] <- “Oral bef” colnames(explicit\_group\_data)[colnames(explicit\_group\_data) == “Oral score after”] <- “Oral aft”

# For the implicit\_group\_data dataframe

colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Score in the selection test”] <- “Score sel.” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Score reading test before”] <- “read. bef” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Score reading test after”] <- “read. aft” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Number of aspects observed before”] <- “aspects bef” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Number of aspects observed after”] <- “aspects aft” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Oral score before”] <- “Oral bef” colnames(implicit\_group\_data)[colnames(implicit\_group\_data) == “Oral score after”] <- “Oral aft”

# Adding group indicators

explicit\_group <- rep(“Explicit”, nrow(explicit\_group\_data)) implicit\_group <- rep(“Implicit”, nrow(implicit\_group\_data))

explicit\_group\_datagroup <- implicit\_group

# Combining the data

combined\_data <- rbind(explicit\_group\_data, implicit\_group\_data)

# Selecting variables of interest

interest\_data <- combined\_data[, c(“Score sel.”, “read. bef”, “read. aft”, “Oral bef”, “Oral aft”)]

library(ggplot2) library(reshape2) library(GGally)

# Correlation matrix

correlation\_matrix <- cor(interest\_data) correlation\_melted <- melt(correlation\_matrix)

# Heatmap

ggplot(correlation\_melted, aes(Var1, Var2, fill = value)) + geom\_tile(color = “white”) + scale\_fill\_gradient(low = “blue”, high = “red”) + labs(title = “Heatmap of Variable Correlations”, x = “Variable”, y = “Variable”) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, vjust = 1, hjust = 1))

ggpairs(interest\_data)

# The selection of interest\_data was based on the relevance of variables related to student results before and after the course.

library(dplyr)

# Interest variables for each group

databox\_implicit <- implicit\_group\_data[, c(“read. bef”, “read. aft”, “Oral bef”, “Oral aft”)] databox\_explicit <- explicit\_group\_data[, c(“read. bef”, “read. aft”, “Oral bef”, “Oral aft”)]

# Melting data

data\_melted\_implicit <- melt(databox\_implicit) data\_melted\_implicit$Group <- "Implicit" data\_melted\_explicit <- melt(databox\_explicit) data\_melted\_explicit$Group <- “Explicit”

# Combining melted data

combined\_melted\_data <- rbind(data\_melted\_implicit, data\_melted\_explicit)

# Boxplots

ggplot(combined\_melted\_data %>% filter(variable %in% c(“read. bef”, “read. aft”)), aes(x = variable, y = value, fill = Group)) + geom\_boxplot() + labs(title = “Boxplot of Reading Before and After”, x = “Variable”, y = “Value”, fill = “Group”) + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

ggplot(combined\_melted\_data %>% filter(!variable %in% c(“read. bef”, “read. aft”)), aes(x = variable, y = value, fill = Group)) + geom\_boxplot() + labs(title = “Boxplot of Other Variables Before and After”, x = “Variable”, y = “Value”, fill = “Group”) + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Observations indicate outliers affecting normality, such as Rodolpho, a photographer, and Elena, a pedagogue. Reanalysis without Rodolpho shows the impact of outliers on averages.

# Removing outlier

combined\_melted\_data\_filtered <- combined\_melted\_data %>% filter(!(variable %in% c(“read. bef”, “read. aft”) & Group == “Explicit” & value == 33))

# Boxplot without outlier

ggplot(combined\_melted\_data\_filtered %>% filter(variable %in% c(“read. bef”, “read. aft”)), aes(x = variable, y = value, fill = Group)) + geom\_boxplot() + labs(title = “Boxplot of Reading Before and After”, x = “Variable”, y = “Value”, fill = “Group”) + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Comparing results, the mean for read. bef increased after outlier removal. This suggests outliers negatively affected the mean scores for the explicit group, while no significant impact was observed for the implicit group.

library(car)

# Linear regression models

model1\_explicit <- lm(Oral aft ~ Age + Study time (years), data = explicit\_group\_data) model2\_explicit <- lm(read. aft ~ Age + Study time (years), data = explicit\_group\_data) model3\_implicit <- lm(Oral aft ~ Age + Study time (years), data = implicit\_group\_data) model4\_implicit <- lm(read. aft ~ Age + Study time (years), data = implicit\_group\_data)

# Summaries

summary(model1\_explicit) summary(model2\_explicit) summary(model3\_implicit) summary(model4\_implicit)

# Removing outlier for explicit group

explicit\_group\_data\_filtered <- explicit\_group\_data[explicit\_group\_data$read. aft != 33, ]

# New regression models without outlier

model1\_explicit\_filtered <- lm(Oral aft ~ Age + Study time (years), data = explicit\_group\_data\_filtered) model2\_explicit\_filtered <- lm(read. aft ~ Age + Study time (years), data = explicit\_group\_data\_filtered)

# Summaries without outlier

summary(model1\_explicit\_filtered) summary(model2\_explicit\_filtered)

# Shapiro-Wilk test for normality of residuals

shapiro1 <- shapiro.test(model1\_explicitresiduals) shapiro3 <- shapiro.test(model3\_implicitresiduals) shapirofiltered1 <- shapiro.test(model1\_explicit\_filtered$residuals)

# Summaries of residuals

summary(rstandard(model1\_explicit)) summary(rstandard(model2\_explicit)) summary(rstandard(model3\_implicit)) summary(rstandard(model4\_implicit)) summary(rstandard(model1\_explicit\_filtered)) summary(rstandard(model2\_explicit\_filtered))

par(mfrow = c(2, 2))

# Plot diagnostics

plot(model1\_explicit) text(x = 0, y = par(“usr”)[4], labels = paste(“Shapiro p-value:”, round(shapiro1$p.value, 4)), adj = c(0, 1), col = “blue”)

plot(model2\_explicit) text(x = 0, y = par(“usr”)[4], labels = paste(“Shapiro p-value:”, round(shapiro2$p.value, 4)), adj = c(0, 1), col = “blue”)

plot(model3\_implicit) text(x = 0, y = par(“usr”)[4], labels = paste(“Shapiro p-value:”, round(shapiro3$p.value, 4)), adj = c(0, 1), col = “blue”)

plot(model4\_implicit) text(x = 0, y = par(“usr”)[4], labels = paste(“Shapiro p-value:”, round(shapiro4$p.value, 4)), adj = c(0, 1), col = “blue”)

plot(model1\_explicit\_filtered) text(x = 0, y = par(“usr”)[4], labels = paste(“Shapiro p-value:”, round(shapirofiltered1$p.value, 4)), adj = c(0, 1), col = “blue”) # Perform t-tests result\_t\_after <- t.test(Oral aft ~ group, data = combined\_data) result\_t\_after2 <- t.test(read. aft ~ group, data = combined\_data)

# T-tests with filtered data

result\_t\_after\_filtered1 <- t.test(Oral aft ~ group, data = combined\_data\_filtered) result\_t\_after\_filtered2 <- t.test(read. aft ~ group, data = combined\_data\_filtered)

# Summaries

summary(result\_t\_after) summary(result\_t\_after2) summary(result\_t\_after\_filtered1) summary(result\_t\_after\_filtered2)

# Oral aft vs. Group

p-value: 0.0364861, indicating a statistically significant difference in Oral aft scores between groups. Read. aft vs. Group p-value: 0.6900625, indicating no significant difference in Read. aft scores between groups. Filtered Data Analysis Oral aft vs. Group p-value: 0.02973489, indicating a significant difference in Oral aft scores between groups. Read. aft vs. Group p-value: 0.7270604, indicating no significant difference in Read. aft scores between groups. Summary and Implications Significant Differences: Oral aft scores differ significantly between groups. No Differences in Read. aft: No significant differences in Read. aft scores between groups.

# Paired t-tests

result\_paired\_t\_test\_PAIRE1 <- t.test(combined\_data$`Oral aft`, combined\_data$Oral bef, paired = TRUE, alternative = “greater”) result\_paired\_t\_test\_PAIRE2 <- t.test(combined\_data$`read. aft`, combined\_data$read. bef, paired = TRUE, alternative = “greater”)

# Summaries

print(result\_paired\_t\_test\_PAIRE1) cat(“P-value for Oral aft vs. Oral bef:”, result\_paired\_t\_test\_PAIRE1$p.value, “”)

print(result\_paired\_t\_test\_PAIRE2) cat(“P-value for read. aft vs. read. bef:”, result\_paired\_t\_test\_PAIRE2$p.value, “”)

# Paired t-tests with filtered data

result\_paired\_t\_test\_PAIRE\_filtered1 <- t.test(combined\_data\_filtered$`Oral aft`, combined\_data\_filtered$Oral bef, paired = TRUE, alternative = “greater”) result\_paired\_t\_test\_PAIRE\_filtered2 <- t.test(combined\_data\_filtered$`read. aft`, combined\_data\_filtered$read. bef, paired = TRUE, alternative = “greater”)

# Summaries

print(result\_paired\_t\_test\_PAIRE\_filtered1) cat(“P-value for filtered Oral aft vs. Oral bef:”, result\_paired\_t\_test\_PAIRE\_filtered1$p.value, “”)

print(result\_paired\_t\_test\_PAIRE\_filtered2) cat(“P-value for filtered read. aft vs. read. bef:”, result\_paired\_t\_test\_PAIRE\_filtered2$p.value, “”)

# Professional group indicators

combined\_dataProfession %in% c(“Literature student”, “english translator”, “reviewer”, “pedagogue”, “Italian teacher”), “Related to Letters”, “Not Related to Letters”) combined\_data\_filteredProfession %in% c(“Literature student”, “english translator”, “reviewer”, “pedagogue”, “Italian teacher”), “Related to Letters”, “Not Related to Letters”)

# Regression analysis

regression\_model\_multiple1 <- lm(Oral aft ~ Professional\_Group, data = combined\_data) summary(regression\_model\_multiple1) boxplot(Oral aft ~ Professional\_Group, data = combined\_data, main = “Oral aft by Professional Group”, xlab = “Professional Group”, ylab = “Oral aft”, col = c(“lightblue”, “lightgreen”))

regression\_model\_multiple\_filtered1 <- lm(Oral aft ~ Professional\_Group, data = combined\_data\_filtered) summary(regression\_model\_multiple\_filtered1) boxplot(Oral aft ~ Professional\_Group, data = combined\_data\_filtered, main = “Oral aft by Professional Group”, xlab = “Professional Group”, ylab = “Oral aft”, col = c(“lightblue”, “lightgreen”))

# Replicate to Read.aft

regression\_model\_multiple1\_read <- lm(read. aft ~ Professional\_Group, data = combined\_data) summary(regression\_model\_multiple1\_read) boxplot(read. aft ~ Professional\_Group, data = combined\_data, main = “Read. aft by Professional Group”, xlab = “Professional Group”, ylab = “Read. aft”, col = c(“lightblue”, “lightgreen”))

regression\_model\_multiple\_filtered1\_read <- lm(read. aft ~ Professional\_Group, data = combined\_data\_filtered) summary(regression\_model\_multiple\_filtered1\_read) boxplot(read. aft ~ Professional\_Group, data = combined\_data\_filtered, main = “Read. aft by Professional Group”, xlab = “Professional Group”, ylab = “Read. aft”, col = c(“lightblue”, “lightgreen”))

# T-Test for Professional Group

result\_t\_Professional\_Group1 <- t.test(Oral aft ~ Professional\_Group, data = combined\_data) result\_t\_Professional\_Group\_filtered1 <- t.test(Oral aft ~ Professional\_Group, data = combined\_data\_filtered) cat(“P-value for Oral aft vs. Professional Group:”, result\_t\_Professional\_Group1$p.value, "\n") cat("P-value for filtered Oral aft vs. Professional Group:", result\_t\_Professional\_Group\_filtered1$p.value, “”)

result\_t\_Professional\_Group2 <- t.test(read. aft ~ Professional\_Group, data = combined\_data) result\_t\_Professional\_Group\_filtered2 <- t.test(read. aft ~ Professional\_Group, data = combined\_data\_filtered) cat(“P-value for read. aft vs. Professional Group:”, result\_t\_Professional\_Group2$p.value, "\n") cat("P-value for filtered read. aft vs. Professional Group:", result\_t\_Professional\_Group\_filtered2$p.value, “”)

# Linear regression for combined data

regression\_model\_group <- lm(Oral aft ~ group, data = combined\_data) summary(regression\_model\_group)

regression\_model\_group\_filtered1 <- lm(Oral aft ~ group, data = combined\_data\_filtered) summary(regression\_model\_group\_filtered1)

# Replicate for read. aft

regression\_model\_group2 <- lm(read. aft ~ group, data = combined\_data) summary(regression\_model\_group2)

regression\_model\_group\_filtered2 <- lm(read. aft ~ group, data = combined\_data\_filtered) summary(regression\_model\_group\_filtered2)

# Regression with additional predictors

regression\_model\_Age\_Study <- lm(Oral aft ~ Age + Study time (years), data = combined\_data) summary(regression\_model\_Age\_Study)

regression\_model\_Age\_Study\_filtered <- lm(Oral aft ~ Age + Study time (years), data = combined\_data\_filtered) summary(regression\_model\_Age\_Study\_filtered)

regression\_model\_Age\_Study3 <- lm(read. aft ~ Age + Study time (years), data = combined\_data) summary(regression\_model\_Age\_Study3)

regression\_model\_Age\_Study\_filtered3 <- lm(read. aft ~ Age + Study time (years), data = combined\_data\_filtered) summary(regression\_model\_Age\_Study\_filtered3)

# Correlation matrix with additional predictor

combined\_data\_filteredHave you traveled to Italy? == “Yes”, 1, 0)

cor\_matrix <- cor(combined\_data\_filtered[, c(“Oral aft”, “read. aft”, “Age”, “Study time (years)”, “traveled\_to\_italy\_numeric”)]) cor\_matrix\_melted <- melt(cor\_matrix)

ggplot(cor\_matrix\_melted, aes(Var2, Var1, fill = value)) + geom\_tile(color = “white”) + geom\_text(aes(label = ifelse(abs(value) > 0.5, ifelse(value > 0, “Positive”, “Negative”), ““)), size = 3, color =”black”) + scale\_fill\_gradient2(low = “blue”, high = “red”, name = “Correlation”) + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + labs(x = “Variables”, y = “Variables”, title = “Correlation Heatmap”) + coord\_fixed()

# Regression for travel to Italy

regression\_model\_ht1 <- lm(Oral aft ~ Have you traveled to Italy?, data = combined\_data) summary(regression\_model\_ht1)

regression\_model\_ht2 <- lm(read. aft ~ Have you traveled to Italy?, data = combined\_data) summary(regression\_model\_ht2)

# Study context indicators

combined\_dataStudy context == “university”, “University”, “Not University”) combined\_data\_filteredStudy context == “university”, “University”, “Not University”)

# Regression analysis for study context

regression\_model\_Study1 <- lm(Oral aft ~ Study\_context\_category, data = combined\_data) summary(regression\_model\_Study1)

regression\_model\_Study2 <- lm(read. aft ~ Study\_context\_category, data = combined\_data) summary(regression\_model\_Study2)

regression\_model\_Study\_filtered <- lm(Oral aft ~ Study\_context\_category, data = combined\_data\_filtered) summary(regression\_model\_Study\_filtered)

regression\_model\_Study\_filtered <- lm(read. aft ~ Study\_context\_category, data = combined\_data\_filtered) summary(regression\_model\_Study\_filtered)

#The analyses performed on the combined and filtered datasets provide valuable insights into the impact of different variables on student performance. Significant differences were found in Oral aft scores between groups, while Read. aft scores showed no significant differences. Professional background and study context were not significant predictors of performance in most cases. Further research with larger sample sizes and additional variables could help to better understand these relationships.