Objective 1

2024-02-20

Data cleaning

```
library("readxl")
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#Upload data
data = read_excel("data/Ametop-PE data(new-Alex)(1control 2study).xlsx")

#Remove withdrawn patients
data <- data[data$`Withdrawn after consent?` != "Yes", ]

#Remove unnecessary variables
data <- select(data,
               -`Elective surgery requiring IV?`,
               -`ASA I or II?`,
               -`Aged 5-16?`,
               -`Any allergies to Ametop, Pain Ease, or Tagederm adhesive?`,
               -`Ametop placed at least 30min before estimated IV start time?`,
               -`Receiving sedative pre-medication / anxiolytics?`,
               -`Needle phobia?`,
               -`Planned inhalation induction?`,
               -`Developmental delay or unable to interpret FPS-R?`,
               -`Time Ametop applied`,
               -`Time Ametop removed`,
               -`Time of skin puncture`,
               -`Record ID`,
               -`Complete?`,
               -`Reason for withdrawal`,
               -`Other reason for withdrawal`)
```

```r
# Check for missing values in all columns except 'Notes' and 'Observed side effects'
missing_values_summary <- data %>%
  select(-Notes, -`Observed side effects`) %>%
  summarise_all(~sum(is.na(.)))

#No missing values found
print(missing_values_summary)
```

```
## # A tibble: 1 x 12
##   Randomization   Age   Sex Practitioner `Withdrawn after consent?`
##           <int> <int> <int>        <int>                      <int>
## 1             0     0     0            0                          0
## # i 7 more variables: `Duration of Ametop` <int>,
## #   `Time between Ametop and skin puncture` <int>,
## #   `Reaction to skin puncture` <int>, `Number of IV attempts` <int>,
## #   `FPS-R score` <int>, `Needle gauge` <int>,
## #   `Any side effects observed?` <int>
```

```r
#Shorten the data
data$`Reaction to skin puncture` <- gsub(
  "^Slight.*", "Slight Pain", data$`Reaction to skin puncture`)

data$`Reaction to skin puncture` <- gsub(
  "^Severe.*", "Severe Pain", data$`Reaction to skin puncture`)

# Verify the changes
table(data$`Reaction to skin puncture`)
```

```
##
##        None Severe Pain Slight Pain
##         158           5          60
```

```r
# Creating new variables for the dataset

# Create 'Observer Number'
data$`Observer Number` <- case_when(
  data$`Reaction to skin puncture` == "None" ~ 1,
  data$`Reaction to skin puncture` == "Slight Pain" ~ 2,
  data$`Reaction to skin puncture` == "Severe Pain" ~ 3
)

# Create 'FPS-R Number'
data$`FPS-R Number` <- case_when(
  data$`FPS-R score` %in% c(0, 2) ~ 1,
  data$`FPS-R score` %in% c(4, 6) ~ 2,
  data$`FPS-R score` %in% c(8, 10) ~ 3
)

# 3. Create 'Match Responses'
data$`Match Responses` <- ifelse(data$`Observer Number` == data$`FPS-R Number`, "Yes", "No")
```

```r
#Verify data
head(data)
```

```
## # A tibble: 6 x 17
##   Randomization   Age Sex    Practitioner `Withdrawn after consent?`
##           <dbl> <dbl> <chr>  <chr>        <chr>
## 1             2     6 Female Resident     No
## 2             1     9 Female Anesthetist  No
## 3             2     7 Male   Resident     No
## 4             1     8 Female Anesthetist  No
## 5             1     9 Male   Resident     No
## 6             1     6 Female Anesthetist  No
## # i 12 more variables: `Duration of Ametop` <dbl>,
## #   `Time between Ametop and skin puncture` <dbl>,
## #   `Reaction to skin puncture` <chr>, `Number of IV attempts` <dbl>,
## #   `FPS-R score` <dbl>, `Needle gauge` <dbl>,
## #   `Any side effects observed?` <chr>, `Observed side effects` <chr>,
## #   Notes <chr>, `Observer Number` <dbl>, `FPS-R Number` <dbl>,
## #   `Match Responses` <chr>
```

Objective 1

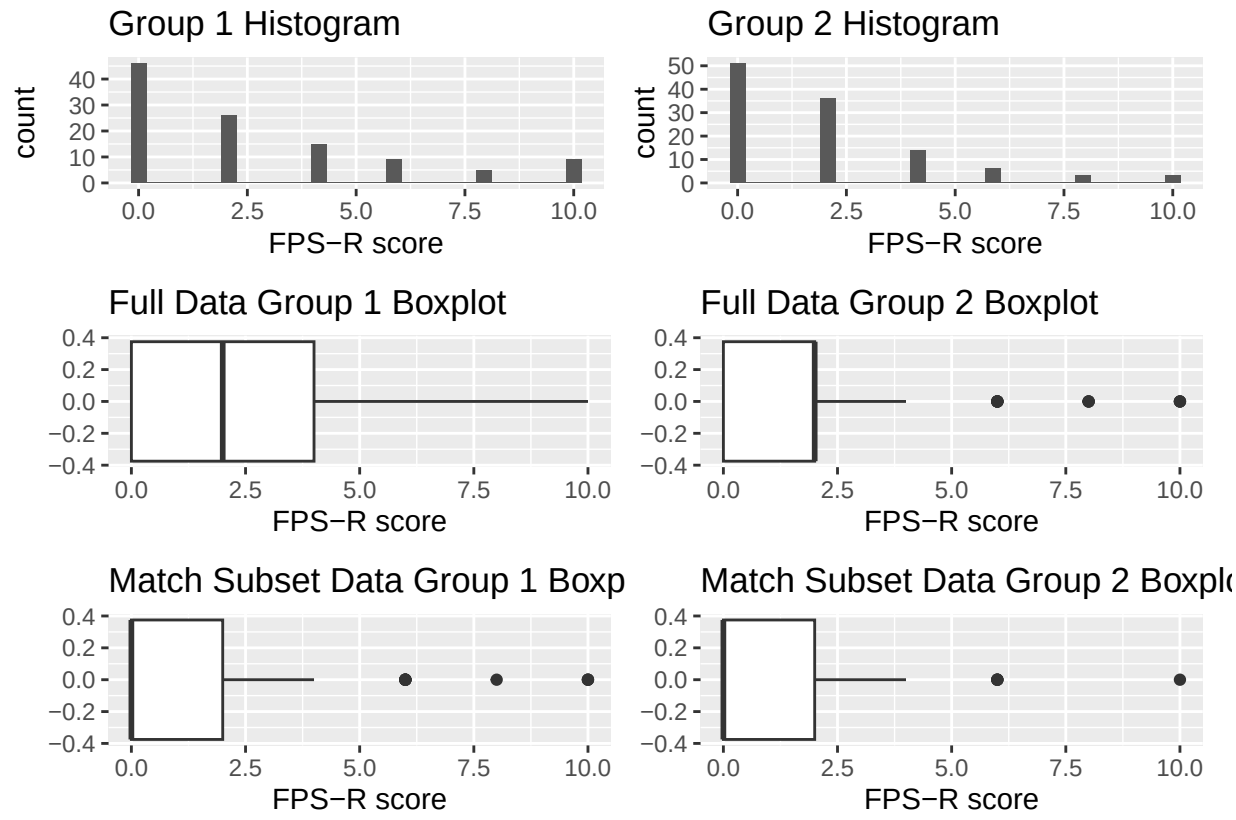
```r
# Histogram
plot1 <- data %>% filter(Randomization == 1) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_histogram() +
  labs(title = "Group 1 Histogram")

plot2 <- data %>% filter(Randomization == 2) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_histogram() +
  labs(title = "Group 2 Histogram")

# Boxplot
plot3 <- data %>% filter(Randomization == 1) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Full Data Group 1 Boxplot")

plot4 <- data %>% filter(Randomization == 2) %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Full Data Group 2 Boxplot")

# Boxplot
plot5 <- data %>% filter(Randomization == 1, `Match Responses` == "Yes") %>%
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Match Subset Data Group 1 Boxplot")

plot6 <- data %>% filter(Randomization == 2, `Match Responses` == "Yes") %>%
```

```
  ggplot(aes(`FPS-R score`)) +
  geom_boxplot() +
  labs(title = "Match Subset Data Group 2 Boxplot")

# Load the patchwork library
library(patchwork)

# Combine the plots
print("Group 1 = Control group \n Group 2 = Study group")
```

```
## [1] "Group 1 = Control group \n Group 2 = Study group"
```

```
combined_plot <- (plot1 | plot2) /
                 (plot3 | plot4) /
                 (plot5 | plot6)
print(combined_plot)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#Create a table to count matches
table <- table(data$`Observer Number`,data$`FPS-R Number`)
```

```r
# Name the rows and columns to reflect the desired format
rownames(table) <- c("1", "2", "3")
colnames(table) <- c("1", "2", "3")


# Output the 3x3 table
cat("Contingency Table of FPS-R Number vs. Observer Number\n")
```

```
## Contingency Table of FPS-R Number vs. Observer Number
```

```r
cat("Rows: FPS-R Number\n")
```

```
## Rows: FPS-R Number
```

```r
cat("Columns: Observer Number\n\n")
```

```
## Columns: Observer Number
```

```r
print(table)
```

```
##
##       1   2   3
##   1 135  20   3
##   2  24  24  12
##   3   0   0   5
```

```r
# T-test

# Perform t-test on full data
t_test_full <- t.test(data[data$Randomization == 1, ]$`FPS-R score`,
                      data[data$Randomization == 2, ]$`FPS-R score`)
print(t_test_full)
```

```
##
##  Welch Two Sample t-test
##
## data:  data[data$Randomization == 1, ]$`FPS-R score` and data[data$Randomization == 2, ]$`FPS-R score
## t = 2.0155, df = 204.03, p-value = 0.04516
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01658064 1.50683046
## sample estimates:
## mean of x mean of y
##  2.690909  1.929204
```

```r
#Perform t-test on subset of data
data_subset <- data[data$`Match Responses` == 'Yes',]
t_test_subset <- t.test(data_subset[data_subset$Randomization == 1, ]$`FPS-R score`,
                        data_subset[data_subset$Randomization == 2, ]$`FPS-R score`)
print(t_test_subset)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_subset[data_subset$Randomization == 1, ]$`FPS-R score` and data_subset[data_subset$Randor
## t = 1.5712, df = 118.53, p-value = 0.1188
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1515205  1.3158398
## sample estimates:
## mean of x mean of y
##  1.915493  1.333333
```

```
# Conclusion: reject null on full data only, if match subset data (with 223-59=164 observations) we do
```