Biểu đồ scatter plot và boxplot cho nhóm Nam hút thuốc vs Nam không hút thuốc

Minh

2024-02-28

options(warn = -1)

## Load bộ số liệu và các packages

library(tidyverse)  
library(gtsummary)  
library(rstatix)  
library(readxl)  
library(boot)  
library(forestplot)  
d <- read\_excel("C:/Users/Admin/OneDrive/NC\_Quangpho/Data\_regardless\_sleep\_18\_03.xlsx")  
View(d)  
  
d$BMI\_25\_30 <- ifelse(d$BMI\_index < 25, 1,   
 ifelse(d$BMI\_index >= 25 & d$BMI\_index < 30, 2, 3))  
d$Age\_stage <- ifelse(d$Age <= 30, 1,   
 ifelse(d$Age >= 30 & d$Age < 40, 2, ifelse(d$Age >= 40 & d$Age < 50, 3,4)))  
d$Gender <- ifelse(d$Gender == 0, "female", ifelse(d$Gender == 1, "male", NA))  
d$Race <- d$`6. Race`  
d <- d %>%  
 mutate(Age\_stage = case\_when(  
 Age\_stage == 1 ~ "Age <30",  
 Age\_stage == 2 ~ "Age [30-40)",  
 Age\_stage == 3 ~ "Age [40-50)",  
 Age\_stage == 4 ~ "Age >=50",  
 TRUE ~ NA\_character\_ # This line handles any unexpected values  
 ))  
d$Hypertension <- ifelse(d$Hypertension == 0, "No", ifelse(d$Hypertension == 1, "Yes", NA))  
d$`Stress index` <- ifelse(d$`Stress index` == 0, "No", ifelse(d$`Stress index` == 1, "Yes", NA))  
d$Obesity <- ifelse(d$Obesity == 0, "No", ifelse(d$Obesity == 1, "Yes", NA))  
d <- d %>%  
 mutate(BMI\_25\_30 = case\_when(  
 BMI\_25\_30 == 1 ~ "BMI <25",  
 BMI\_25\_30 == 2 ~ "BMI [25-30)",  
 BMI\_25\_30 == 3 ~ "BMI >=30",  
 TRUE ~ NA\_character\_ # This line handles any unexpected values  
 ))  
d$Smoking <- ifelse(d$Smoking == "Do not smoke", "Non smoker", ifelse(d$Smoking == "Cigarettes, pipe tobacco...", "Smoker", NA))  
d <- d %>% filter(Gender == "male")

## So sánh các chỉ số trung bình của nhóm Smoker vs Non smoker bằng Wilcox test

Tạo hàm tính 95%CI

library(dplyr)  
  
calculate\_95CI <- function(data, variable) {  
 ci\_lower <- quantile(data[[variable]], probs = 0.025, na.rm = TRUE)  
 ci\_upper <- quantile(data[[variable]], probs = 0.975, na.rm = TRUE)  
   
 return(c(CI\_lower = ci\_lower, CI\_upper = ci\_upper))  
}

Tạo hàm so sánh trung bình bằng Wilcox.test nếu subgroup có 2 level

library(dplyr)  
  
compare\_means <- function(data, subgroup, variable) {  
 # Split data into subgroups  
 subgroup\_levels <- unique(data[[subgroup]])  
 results <- data.frame(Subgroup = subgroup\_levels, n = NA, Mean = NA, CI\_lower = NA, CI\_upper = NA)  
   
 # Loop through each subgroup to calculate mean, n, and 95% CI  
 for (i in 1:length(subgroup\_levels)) {  
 subgroup\_data <- data[data[[subgroup]] == subgroup\_levels[i], ]  
 n <- nrow(subgroup\_data)  
 mean\_value <- mean(subgroup\_data[[variable]], na.rm = TRUE)  
 CI\_lower <- quantile(subgroup\_data[[variable]], probs = 0.025, na.rm = TRUE)  
 CI\_upper <- quantile(data[[variable]], probs = 0.975, na.rm = TRUE)  
   
 results[i, "n"] <- n  
 results[i, "Mean"] <- mean\_value  
 results[i, "CI\_lower"] <- CI\_lower  
 results[i, "CI\_upper"] <- CI\_upper  
 }  
   
 # Statistical comparison  
 if (length(subgroup\_levels) > 2) {  
 results$P\_Value <- NA # Not applicable for more than two subgroups  
 } else if (length(subgroup\_levels) == 2) {  
 # If exactly two subgroups, perform Wilcoxon signed-rank test and add p-values to the results  
 test\_result <- wilcox.test(data[[variable]] ~ data[[subgroup]], data = data)  
 results$P\_Value <- rep(test\_result$p.value, nrow(results))  
 } else {  
 # If only one subgroup, no comparison is needed  
 results$P\_Value <- NA  
 message("Only one subgroup detected. No comparison needed.")  
 }  
   
 return(results)  
}

List\_var <- c("M\_mean", "Kv100\_mean", "δ\_mean", "T\_mean", "A365\_mean", "A460\_mean", "Anadn\_mean", "POM\_mean", "A-E\_mean", "A-N\_mean", "A-M\_mean", "A-R\_mean", "A-C\_mean", "F-E\_mean", "F-N\_mean", "F-M\_mean", "F-R\_mean", "F-C\_mean")

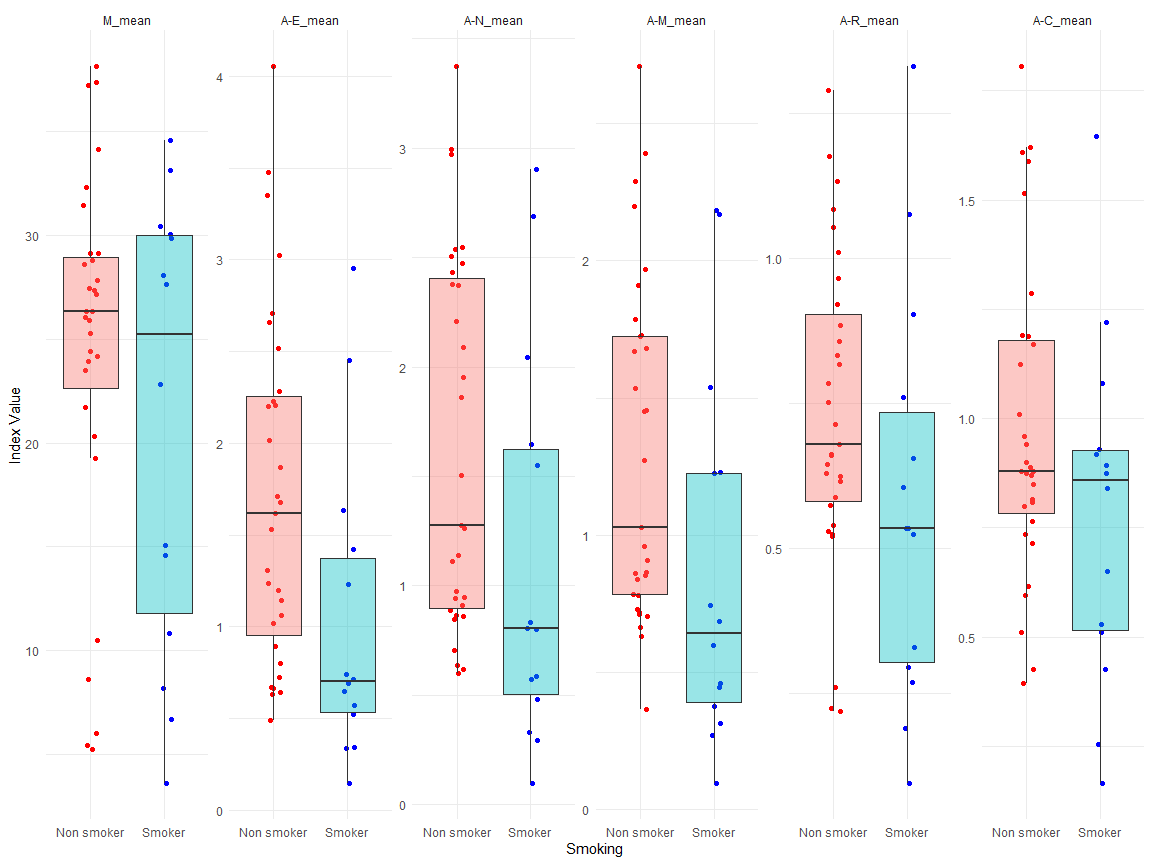
# Initialize an empty list to store results  
results\_list <- list()  
  
# Loop through each variable and apply compare\_means  
for (var in List\_var) {  
 result <- compare\_means(data = d, subgroup = "Smoking", variable = var)  
 result$Variable <- var # Add a new column for the variable name  
 results\_list[[var]] <- result  
}  
  
# Combine individual data frames into a final results table  
final\_results <- do.call(rbind, results\_list)  
  
# Reordering columns to make 'Variable' the first column  
final\_results <- final\_results[, c("Variable", setdiff(names(final\_results), "Variable"))]  
  
# Reset row names to remove any potential duplicates  
rownames(final\_results) <- NULL  
  
# View the final results  
final\_results %>% knitr::kable()

| Variable | Subgroup | n | Mean | CI\_lower | CI\_upper | P\_Value |
| --- | --- | --- | --- | --- | --- | --- |
| M\_mean | Non smoker | 31 | 24.4826613 | 5.372500 | 37.353250 | 0.6893053 |
| M\_mean | Smoker | 14 | 21.1096667 | 4.602750 | 37.353250 | 0.6893053 |
| Kv100\_mean | Non smoker | 31 | 20.6321237 | 6.808125 | 40.097500 | 0.0795145 |
| Kv100\_mean | Smoker | 14 | 14.4326905 | 7.875125 | 40.097500 | 0.0795145 |
| δ\_mean | Non smoker | 31 | 4.1933333 | 1.848750 | 7.860000 | 0.0269689 |
| δ\_mean | Smoker | 14 | 2.9509881 | 0.551750 | 7.860000 | 0.0269689 |
| T\_mean | Non smoker | 31 | 31.5779301 | 24.133750 | 35.666000 | 0.5362906 |
| T\_mean | Smoker | 14 | 29.9534405 | 21.912875 | 35.666000 | 0.5362906 |
| A365\_mean | Non smoker | 31 | 87.4220430 | 19.625000 | 155.800000 | 0.7313942 |
| A365\_mean | Smoker | 14 | 85.8571429 | 2.812500 | 155.800000 | 0.7313942 |
| A460\_mean | Non smoker | 31 | 57.2473118 | 14.875000 | 106.950000 | 0.3451713 |
| A460\_mean | Smoker | 14 | 65.9904762 | 13.650000 | 106.950000 | 0.3451713 |
| Anadn\_mean | Non smoker | 31 | 0.7467742 | 0.436250 | 4.177500 | 0.0954619 |
| Anadn\_mean | Smoker | 14 | 1.4245952 | 0.409375 | 4.177500 | 0.0954619 |
| POM\_mean | Non smoker | 31 | 9.5858333 | 2.242500 | 18.402500 | 0.0839258 |
| POM\_mean | Smoker | 14 | 7.8422738 | 1.097625 | 18.402500 | 0.0839258 |
| A-E\_mean | Non smoker | 31 | 1.7288978 | 0.595000 | 3.462500 | 0.0107407 |
| A-E\_mean | Smoker | 14 | 1.0282619 | 0.206750 | 3.462500 | 0.0107407 |
| A-N\_mean | Non smoker | 31 | 1.6372043 | 0.611250 | 2.992833 | 0.0179084 |
| A-N\_mean | Smoker | 14 | 1.1155238 | 0.158375 | 2.992833 | 0.0179084 |
| A-M\_mean | Non smoker | 31 | 1.2960215 | 0.563750 | 2.382250 | 0.0166848 |
| A-M\_mean | Smoker | 14 | 0.8810714 | 0.151875 | 2.382250 | 0.0166848 |
| A-R\_mean | Non smoker | 31 | 0.7346237 | 0.223750 | 1.278500 | 0.0861069 |
| A-R\_mean | Smoker | 14 | 0.5788095 | 0.125875 | 1.278500 | 0.0861069 |
| A-C\_mean | Non smoker | 31 | 0.9783602 | 0.417500 | 1.642500 | 0.2111059 |
| A-C\_mean | Smoker | 14 | 0.7811071 | 0.194250 | 1.642500 | 0.2111059 |
| F-E\_mean | Non smoker | 31 | 0.0147742 | 0.010000 | 0.020000 | 0.3505025 |
| F-E\_mean | Smoker | 14 | 0.0156702 | 0.012500 | 0.020000 | 0.3505025 |
| F-N\_mean | Non smoker | 31 | 0.0340860 | 0.021750 | 0.047375 | 0.5155402 |
| F-N\_mean | Smoker | 14 | 0.0353881 | 0.021000 | 0.047375 | 0.5155402 |
| F-M\_mean | Non smoker | 31 | 0.0765242 | 0.061000 | 0.134350 | 0.1074320 |
| F-M\_mean | Smoker | 14 | 0.0912810 | 0.061000 | 0.134350 | 0.1074320 |
| F-R\_mean | Non smoker | 31 | 0.3239946 | 0.206000 | 0.507400 | 0.4312457 |
| F-R\_mean | Smoker | 14 | 0.3495262 | 0.206000 | 0.507400 | 0.4312457 |
| F-C\_mean | Non smoker | 31 | 1.1142957 | 0.851000 | 1.486000 | 0.2152527 |
| F-C\_mean | Smoker | 14 | 1.2020857 | 1.016466 | 1.486000 | 0.2152527 |

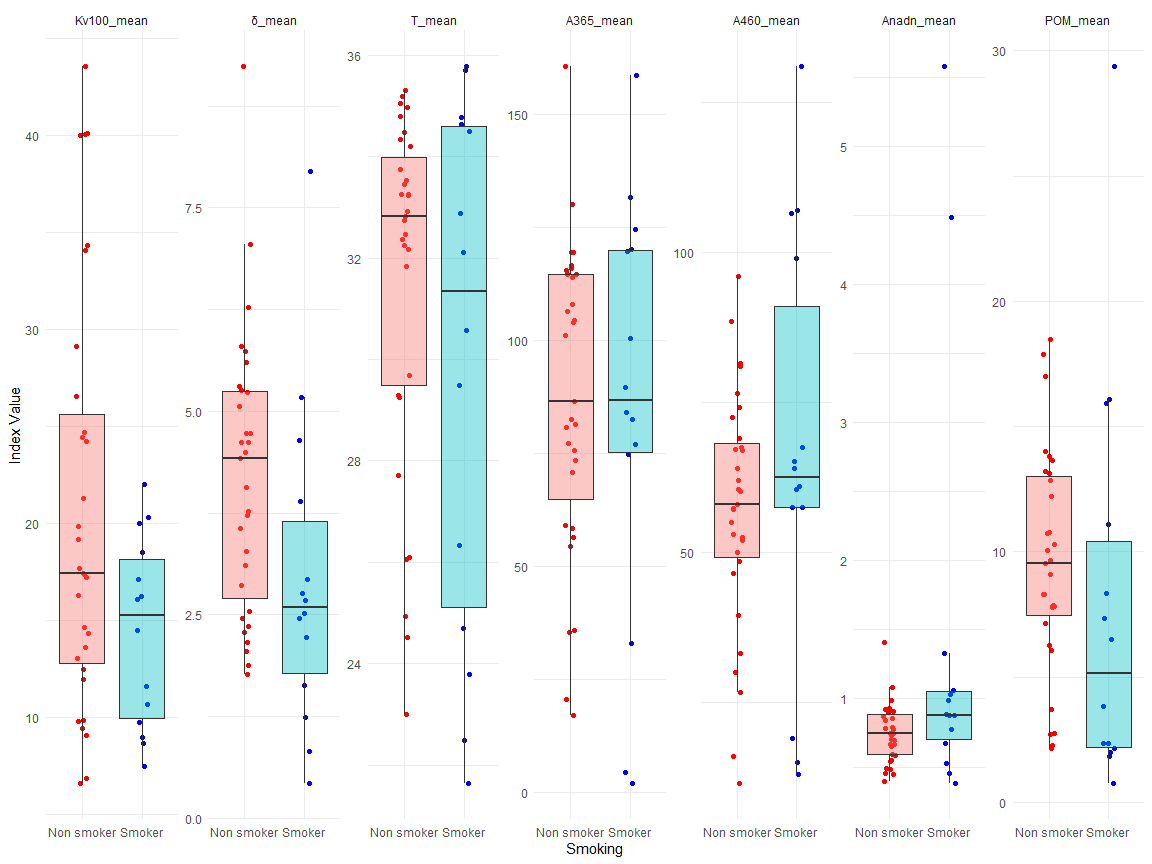
## Vẽ đồ thị scatter plot và boxplot

### Subgroup Smoking

long\_data\_Smoking1 <- reshape2::melt(d, id.vars = "Smoking", measure.vars = c("M\_mean", "A-E\_mean", "A-N\_mean", "A-M\_mean", "A-R\_mean", "A-C\_mean"))  
  
# Plotting  
ggplot(long\_data\_Smoking1, aes(x = Smoking, y = value, fill = Smoking)) +  
 geom\_jitter(aes(colour = Smoking), position = position\_jitterdodge(jitter.width = 0.2, dodge.width = 0.75)) +  
 geom\_boxplot(outlier.shape = NA, position = position\_dodge(width = 0.1), alpha = 0.4) +  
 facet\_wrap(~variable, scales = "free\_y", nrow = 1) +  
 theme\_minimal() +  
 theme(legend.position = "none") + # Hide the legend if not necessary  
 ylab("Index Value") +  
 xlab("Smoking") +  
 scale\_color\_manual(values = c("Smoker" = "blue", "Non smoker" = "red"))



long\_data\_Smoking2 <- reshape2::melt(d, id.vars = "Smoking", measure.vars = c("Kv100\_mean", "δ\_mean", "T\_mean", "A365\_mean", "A460\_mean", "Anadn\_mean", "POM\_mean"))  
  
# Plotting  
ggplot(long\_data\_Smoking2, aes(x = Smoking, y = value, fill = Smoking)) +  
 geom\_jitter(aes(colour = Smoking), position = position\_jitterdodge(jitter.width = 0.2, dodge.width = 0.75)) +  
 geom\_boxplot(outlier.shape = NA, position = position\_dodge(width = 0.1), alpha = 0.4) +  
 facet\_wrap(~variable, scales = "free\_y", nrow = 1) +  
 theme\_minimal() +  
 theme(legend.position = "none") + # Hide the legend if not necessary  
 ylab("Index Value") +  
 xlab("Smoking") +  
 scale\_color\_manual(values = c("Smoker" = "blue", "Non smoker" = "red"))



long\_data\_Smoking3 <- reshape2::melt(d, id.vars = "Smoking", measure.vars = c("F-E\_mean", "F-N\_mean", "F-M\_mean", "F-R\_mean", "F-C\_mean"))  
  
# Plotting  
ggplot(long\_data\_Smoking3, aes(x = Smoking, y = value, fill = Smoking)) +  
 geom\_jitter(aes(colour = Smoking), position = position\_jitterdodge(jitter.width = 0.2, dodge.width = 0.75)) +  
 geom\_boxplot(outlier.shape = NA, position = position\_dodge(width = 0.1), alpha = 0.4) +  
 facet\_wrap(~variable, scales = "free\_y", nrow = 1) +  
 theme\_minimal() +  
 theme(legend.position = "none") + # Hide the legend if not necessary  
 ylab("Index Value") +  
 xlab("Smoking") +  
 scale\_color\_manual(values = c("Smoker" = "blue", "Non smoker" = "red"))

