## ---- 0) Install & Load packages ----

pkgs <- c("readxl", "dplyr", "MatchIt", "tableone")

to\_install <- pkgs[!pkgs %in% rownames(installed.packages())]

if (length(to\_install) > 0) install.packages(to\_install)

library(readxl)

library(dplyr)

library(MatchIt)

library(tableone)

## ---- 1) Read data ----

my\_data <- read\_excel("D:/my\_data.xlsx")

df <- as.data.frame(my\_data)

print(names(df))

print(head(df))

## ---- 2) Define treatment variable ----

df <- df %>%

mutate(

treatment = case\_when(

treatment %in% c(1, "1", "treated", "Treatment") ~ 1,

TRUE ~ 0

),

treatment = factor(treatment, levels = c(0, 1),

labels = c("Control", "Treatment"))

)

## ---- 3) Define covariates ----

catVars <- c("Age", "Sex", "Smoking", "Alcohol", "Hypertension",

"Diabetes\_mellitus", "BMI", "Tumor\_length",

"Tumor\_location", "TNM")

vars\_in\_data <- intersect(catVars, names(df))

validCatVars <- vars\_in\_data[sapply(df[vars\_in\_data], function(x) {

length(unique(x[!is.na(x)])) > 1

})]

df[validCatVars] <- lapply(df[validCatVars], as.factor)

## ---- 4) Pre-matching balance ----

cat("\n===== Pre-matching sample sizes =====\n")

print(table(df$treatment))

tab1\_pre <- CreateTableOne(

vars = vars\_in\_data,

strata = "treatment",

data = df,

factorVars = intersect(validCatVars, vars\_in\_data),

test = FALSE

)

cat("\n===== Table 1 (Pre-matching) with SMD =====\n")

print(tab1\_pre, showAllLevels = TRUE, quote = TRUE, noSpaces = TRUE, smd = TRUE)

tab1\_pre\_df <- print(tab1\_pre, showAllLevels = TRUE, quote = TRUE, noSpaces = TRUE,

smd = TRUE, printToggle = FALSE)

write.csv(as.data.frame(tab1\_pre\_df), "table1\_pre\_matching\_with\_SMD.csv", row.names = FALSE)

## ---- 5) Propensity Score Matching ----

m.formula <- as.formula(paste("treatment ~", paste(vars\_in\_data, collapse = " + ")))

m.out <- matchit(

formula = m.formula,

data = df,

method = "nearest",

caliper = 0.1,

ratio = 1

)

m.data <- match.data(m.out)

## ---- 6) Post-matching balance ----

cat("\n===== Post-matching sample sizes =====\n")

print(table(m.data$treatment))

tab1\_post <- CreateTableOne(

vars = vars\_in\_data,

strata = "treatment",

data = m.data,

factorVars = intersect(validCatVars, vars\_in\_data),

test = FALSE

)

cat("\n===== Table 1 (Post-matching) with SMD =====\n")

print(tab1\_post, showAllLevels = TRUE, quote = TRUE, noSpaces = TRUE, smd = TRUE)

tab1\_post\_df <- print(tab1\_post, showAllLevels = TRUE, quote = TRUE, noSpaces = TRUE,

smd = TRUE, printToggle = FALSE)

write.csv(as.data.frame(tab1\_post\_df), "table1\_post\_matching\_with\_SMD.csv", row.names = FALSE)

## ---- 7) Save matched dataset ----

write.csv(m.data, "match.csv", row.names = FALSE)

cat("\nOutputs saved:\n",

"- table1\_pre\_matching\_with\_SMD.csv\n",

"- table1\_post\_matching\_with\_SMD.csv\n",

"- match.csv (matched dataset)\n")