```
# ========================================================================

# HƯỚNG DẪN: PHÂN TÍCH THỐNG KÊ TẦN SỐ- MANH HUNG DIGI 086 978 6862

# ========================================================================


# 1. NẠP DỮ LIỆU ĐẦU VÀO

# Bạn hãy thay đường dẫn file Excel bên dưới bằng đường dẫn thực tế trên máy của bạn

if(!require(readxl)) install.packages("readxl", dependencies = TRUE)

library(readxl)

my_data <- read_excel("C:/Users/admin2/Desktop/hoc r cung manh hung digi/du lieu dau vao.xlsx",
sheet = 1)


# 2. ÉP TOÀN BỘ TÊN BIẾN TRONG DATA THÀNH CHỮ HOA ĐỂ CHỐNG LỖI KÝ TỰ

names(my_data) <- toupper(names(my_data))


# 3. KHAI BÁO DANH SÁCH CÁC BIẾN CẦN PHÂN TÍCH

danh_sach_bien <- c("GIOITINH", "NOIO")



# 4. THIẾT LẬP TỪ ĐIỂN GÁN NHÃN (VALUE LABELS - TƯƠNG ĐƯƠNG VARIABLE VIEW SPSS)

danh_sach_label <- list(

  "GIOITINH" = list(levels = c(1, 2), labels = c("Nam", "Nữ")),

  "NOIO"     = list(levels = c(1, 2), labels = c("Thành thị", "Nông thôn"))

)



# 5. TỰ ĐỘNG CÀI ĐẶT CÁC GÓI HỖ TRỢ XUẤT FILE VÀ VẼ ĐỒ THỊ
```

```
if(!require(ggplot2)) install.packages("ggplot2", dependencies = TRUE)
if(!require(gridExtra)) install.packages("gridExtra", dependencies = TRUE)
if(!require(writexl)) install.packages("writexl", dependencies = TRUE)
if(!require(e1071)) install.packages("e1071", dependencies = TRUE)
library(ggplot2)
library(gridExtra)
library(writexl)
library(e1071)



# ------------------------------------------------------------------------
# HÀM THÔNG MINH: TỰ ĐỘNG ĐỔI TÊN FILE NẾU ĐÃ TỒN TẠI (Tự động thêm _1, _2...)
# Bạn giữ nguyên khối lệnh này, không cần chỉnh sửa gì cả
# ------------------------------------------------------------------------
tim_ten_file_chuan <- function(ten_goc, duoi_file) {
  ten_day_du <- paste0(ten_goc, ".", duoi_file)
  if (!file.exists(ten_day_du)) {
    return(ten_day_du)
  }
  dem <- 1
  while (TRUE) {
    ten_moi <- paste0(ten_goc, "_", dem, ".", duoi_file)
    if (!file.exists(ten_moi)) {
      return(ten_moi)
    }
```

```
    dem <- dem + 1
  }
}

# Hàm tìm Mode (Yếu vị) cho dữ liệu
tim_mode <- function(x) {
  x <- x[!is.na(x)]
  ux <- unique(x)
  tab <- tabulate(match(x, ux))
  return(ux[tab == max(tab)][1])
}
# ------------------------------------------------------------------------



# 6. VÒNG LẶP TỰ ĐỘNG GÁN LABEL, XỬ XÝ SỐ LIỆU VÀ TẠO BẢNG CHUẨN SPSS
danh_sach_do_thi <- list()
danh_sach_bang_excel <- list()

for (ten_bien in danh_sach_bien) {

  cat("\n======================================================================\n")
  cat("PHÂN TÍCH THỐNG KÊ CHO BIẾN: [", ten_bien, "]\n")
  cat("======================================================================\n")

  # --- TRÍCH XUẤT DỮ LIỆU SỐ THÔ ĐỂ TÍNH STATISTICS ---
```

```
du_lieu_so <- as.numeric(my_data[[ten_bien]])
du_lieu_so_sach <- du_lieu_so[!is.na(du_lieu_so)]
n_mau <- length(du_lieu_so_sach)

# Tính toán chi tiết toàn bộ tham số tương đương bảng Statistics của SPSS
val_mean     <- mean(du_lieu_so, na.rm = TRUE)
val_se_mean  <- sd(du_lieu_so, na.rm = TRUE) / sqrt(n_mau)
val_median   <- median(du_lieu_so, na.rm = TRUE)
val_mode     <- tim_mode(du_lieu_so)
val_stddev   <- sd(du_lieu_so, na.rm = TRUE)
val_variance <- var(du_lieu_so, na.rm = TRUE)
val_min      <- min(du_lieu_so, na.rm = TRUE)
val_max      <- max(du_lieu_so, na.rm = TRUE)
val_range    <- val_max - val_min

# Tính Skewness, Kurtosis chuẩn theo công thức SPSS
val_skew     <- e1071::skewness(du_lieu_so_sach, type = 2)
val_se_skew  <- sqrt((6 * n_mau * (n_mau - 1)) / ((n_mau - 2) * (n_mau + 1) * (n_mau + 3)))
val_kurt     <- e1071::kurtosis(du_lieu_so_sach, type = 2)
val_se_kurt  <- 2 * val_se_skew * sqrt((n_mau^2 - 1) / ((n_mau - 3) * (n_mau + 5)))

# Gom thành một bảng thống kê tham số (Cột giá trị bắt buộc để kiểu dữ liệu số)
bang_statistics <- data.frame(
  "Danh mục" = c("Mean (Trung bình)", "Std. Error of Mean (Sai số chuẩn Mean)",
                 "Median (Trung vị)", "Mode (Yếu vị)", "Std. Deviation (Độ lệch chuẩn)",
```

```
                "Variance (Phương sai)", "Skewness (Độ lệch)", "Std. Error of Skewness",
                "Kurtosis (Độ nhọn)", "Std. Error of Kurtosis", "Range (Khoảng biến thiên)",
                "Minimum (Nhỏ nhất)", "Maximum (Lớn nhất)"),
  "Tần số (Frequency)" = round(c(val_mean, val_se_mean, val_median, val_mode, val_stddev,
                  val_variance, val_skew, val_se_skew, val_kurt, val_se_kurt,
                  val_range, val_min, val_max), 3),
  stringsAsFactors = FALSE,
  check.names = FALSE
)

cat("--- BẢNG THAM SỐ THỐNG KÊ (STATISTICS) ---\n")
print(bang_statistics)
cat("\n")

# --- TỰ ĐỘNG GÁN NHÃN (LABEL) ---
if (ten_bien %in% names(danh_sach_label)) {
  my_data[[ten_bien]] <- factor(my_data[[ten_bien]],
                    levels = danh_sach_label[[ten_bien]]$levels,
                    labels = danh_sach_label[[ten_bien]]$labels)
} else {
  my_data[[ten_bien]] <- as.factor(my_data[[ten_bien]])
}

# --- TÍNH TOÁN CÁC THÀNH PHẦN THỐNG KÊ TẦN SỐ ---
tanso_hople <- table(my_data[[ten_bien]])
```

```
tanso_co_na <- table(my_data[[ten_bien]], useNA = "always")
tyle_hople  <- prop.table(tanso_hople) * 100
tyle_co_na  <- prop.table(tanso_co_na) * 100

# --- TỰ ĐỘNG DỰNG BẢNG CÓ DÒNG TỔNG (TOTAL) ---
nhom_ten <- names(tanso_co_na)
nhom_ten[is.na(nhom_ten)] <- "Missing (Khuyết)"

bang_chi_tiet <- data.frame(
  "Danh mục" = nhom_ten,
  "Tần số (Frequency)" = as.vector(tanso_co_na),
  "Tỷ lệ % (Percent)" = round(as.vector(tyle_co_na), 2),
  "Tỷ lệ % Hợp lệ (Valid Percent)" = c(round(as.vector(tyle_hople), 2), NA),
  stringsAsFactors = FALSE,
  check.names = FALSE
)

tong_tanso <- sum(tanso_co_na)
tong_tyle  <- sum(tyle_co_na)

dong_tong <- data.frame(
  "Danh mục" = "Total (Tổng)",
  "Tần số (Frequency)" = tong_tanso,
  "Tỷ lệ % (Percent)" = round(tong_tyle, 2),
  "Tỷ lệ % Hợp lệ (Valid Percent)" = 100.00,
```

```
  stringsAsFactors = FALSE,
  check.names = FALSE
)

bang_cuoi_cung <- rbind(bang_chi_tiet, dong_tong)
cat("--- BẢNG TẦN SỐ (FREQUENCY TABLE) ---\n")
print(bang_cuoi_cung)

# Chuẩn hóa bảng statistics sang cấu trúc 4 cột, ép kiểu để giữ định dạng số (Numeric) khi xuất Excel
bang_stats_4col <- bang_statistics
bang_stats_4col[["Tỷ lệ % (Percent)"]] <- NA
bang_stats_4col[["Tỷ lệ % Hợp lệ (Valid Percent)"]] <- NA
names(bang_stats_4col) <- names(bang_cuoi_cung)

# Dựng cấu trúc bảng hỗn hợp nhưng chuyển đổi kiểu dữ liệu cột số về dạng đồng nhất mà không ép thành ký tự
hang_tieu_de_tanso <- data.frame("Danh mục" = "--- BẢNG TẦN SỐ ---", "Tần số (Frequency)" = NA, "Tỷ lệ % (Percent)" = NA, "Tỷ lệ % Hợp lệ (Valid Percent)" = NA, check.names = FALSE)
hang_trong        <- data.frame("Danh mục" = "", "Tần số (Frequency)" = NA, "Tỷ lệ % (Percent)" = NA, "Tỷ lệ % Hợp lệ (Valid Percent)" = NA, check.names = FALSE)
hang_tieu_de_stats <- data.frame("Danh mục" = "--- THAM SỐ THỐNG KÊ (STATISTICS) ---", "Tần số (Frequency)" = NA, "Tỷ lệ % (Percent)" = NA, "Tỷ lệ % Hợp lệ (Valid Percent)" = NA, check.names = FALSE)

# Gộp các khối lại với nhau (Giữ nguyên định dạng số cho Excel căn phải tự động)
bang_gop_excel <- rbind(
  hang_tieu_de_tanso,
```

```
  bang_cuoi_cung,
  hang_trong,
  hang_tieu_de_stats,
  bang_stats_4col
)

danh_sach_bang_excel[[ten_bien]] <- bang_gop_excel

# --- TẠO ĐỒ THỊ ---
df_do_thi <- as.data.frame(tanso_hople)
colnames(df_do_thi) <- c("Biến", "Freq")

p <- ggplot(data = df_do_thi, aes(x = Biến, y = Freq, fill = Biến)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = Freq), vjust = -0.3, size = 4) +
  labs(title = paste("Biểu đồ biến:", ten_bien), x = "Danh mục", y = "Số lượng") +
  theme_minimal() +
  theme(legend.position = "none")

danh_sach_do_thi[[ten_bien]] <- p

# --- TỰ ĐỘNG LƯU TỪNG FILE ẢNH BIỂU ĐỒ (CHỐNG GHI ĐÈ) ---
ten_goc_anh <- paste0("BieuDo_", ten_bien)
ten_file_anh_chuan <- tim_ten_file_chuan(ten_goc_anh, "png")
ggsave(filename = ten_file_anh_chuan, plot = p, width = 7, height = 5, dpi = 300)
```

```
}



# 7. HIỂN THỊ TẤT CẢ BIỂU ĐỒ TRÊN CỬA SỔ PLOTS CỦA RSTUDIO

cat("\n--- ĐANG XUẤT TOÀN BỘ BIỂU ĐỒ LÊN MÀN HÌNH... ---\n")

do.call(grid.arrange, c(danh_sach_do_thi, ncol = 2))



# --- TỰ ĐỘNG XUẤT BẢNG KẾT QUẢ RA FILE EXCEL (CHỐNG GHI ĐÈ) ---

ten_file_excel_chuan <- tim_ten_file_chuan("Ket_Qua_Thong_Ke_Tan_So", "xlsx")

write_xlsx(danh_sach_bang_excel, path = ten_file_excel_chuan)



cat("\n======================================================================")

cat("\n[THÔNG BÁO THÀNH CÔNG]:")

cat("\n- File Excel đã lưu dưới tên hoàn toàn mới: ", ten_file_excel_chuan)

cat("\n- Định dạng dữ liệu số đã được đồng bộ chuẩn (Số sẽ tự động CĂN PHẢI trong Excel để tính toán).")

cat("\n- Toàn bộ file ảnh và Excel đã được gom vào thư mục bạn cấu hình.")

cat("\n======================================================================\n")
```