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

# www.manhhungdigi.com 086 978 6862

# 024_HoanChinh

# HƯỚNG DẪN: ONE-WAY ANOVA MULTI-VARIABLE CHUẨN SPSS

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


# 1. NẠP DỮ LIỆU ĐẦU VÀO

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 CẤU HÌNH BIẾN PHÂN TÍCH TẠI ĐÂY <<<

bien_phan_nhom <- "TUOI"

danh_sach_phu_thuoc <- c("HL1", "HL2", "HL3")



# 4. TỰ ĐỘNG CÀI ĐẶT CÁC GÓI HỖ TRỢ TOÁN VÀ GHI FILE NÂNG CAO

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

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

if(!require(PMCMRplus)) install.packages("PMCMRplus", dependencies = TRUE) # Games-Howell chuẩn SPSS

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

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


library(car)
library(emmeans)
library(PMCMRplus)
library(openxlsx)
library(plyr)


# ------------------------------------------------------------------------
# HÀM THÔNG MINH: TỰ ĐỘNG ĐỔI TÊN FILE NẾU ĐÃ TỒN TẠI (Tự động thêm _1, _2...)
# Giữ nguyên khối lệnh này, không cần chỉnh sửa
# ------------------------------------------------------------------------
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
  }
```

```
}
# ------------------------------------------------------------------------

# 5. KHỞI TẠO FILE EXCEL VÀ SHEET DUY NHẤT TRONG BỘ NHỚ
wb <- createWorkbook()
ten_sheet_duy_nhat <- "OneWay_ANOVA_Full_PostHoc"
addWorksheet(wb, sheetName = ten_sheet_duy_nhat)
dong_hien_tai <- 1

list_descriptives <- list()
list_levene_4types <- list()
list_anova_table <- list()
list_welch_table <- list()
list_posthoc_pairs <- list()

# 6. VÒNG LẶP XỬ LÝ SỐ LIỆU CHO TỪNG BIẾN PHỤ THUỘC
for (bien_phu_thuoc in danh_sach_phu_thuoc) {

  # Lọc dữ liệu khuyết thiếu cục bộ từng biến (/MISSING=ANALYSIS)
  data_sub <- na.omit(as.data.frame(my_data[, c(bien_phan_nhom, bien_phu_thuoc)]))
  data_sub[[bien_phan_nhom]] <- as.factor(data_sub[[bien_phan_nhom]])
  data_sub[[bien_phu_thuoc]]  <- as.numeric(data_sub[[bien_phu_thuoc]])

  if (length(unique(data_sub[[bien_phan_nhom]])) < 2) next
```

```
# --- BẢNG 1: DESCRIPTIVES (Thống kê mô tả tất cả các nhóm + Dòng Total) ---
stats_nhom <- ddply(data_sub, bien_phan_nhom, function(df) {
  x <- df[[bien_phu_thuoc]]
  n <- length(x)
  m <- mean(x)
  s <- sd(x)
  se <- s / sqrt(n)
  ci_margin <- qt(0.975, df = n - 1) * se
  data.frame(
    N = n, Mean = m, `Std. Deviation` = s, `Std. Error` = se,
    `95% CI Lower` = m - ci_margin, `95% CI Upper` = m + ci_margin,
    check.names = FALSE
  )
})

x_all <- data_sub[[bien_phu_thuoc]]
n_all <- length(x_all)
m_all <- mean(x_all)
s_all <- sd(x_all)
se_all <- s_all / sqrt(n_all)
ci_m_all <- qt(0.975, df = n_all - 1) * se_all

row_total <- data.frame(
  V1 = "Total", N = n_all, Mean = m_all, `Std. Deviation` = s_all, `Std. Error` = se_all,
  `95% CI Lower` = m_all - ci_m_all, `95% CI Upper` = m_all + ci_m_all, check.names = FALSE
```

```
  )
  colnames(row_total)[1] <- bien_phan_nhom
  stats_des <- rbind(stats_nhom, row_total)

  stats_des$Mean <- round(stats_des$Mean, 4)
  stats_des$`Std. Deviation` <- round(stats_des$`Std. Deviation`, 4)
  stats_des$`Std. Error` <- round(stats_des$`Std. Error`, 4)
  stats_des$`95% CI Lower` <- round(stats_des$`95% CI Lower`, 4)
  stats_des$`95% CI Upper` <- round(stats_des$`95% CI Upper`, 4)

  df_des_final <- data.frame("Biến phụ thuộc" = c(bien_phu_thuoc, rep("", nrow(stats_des)-1)),
stats_des, check.names = FALSE)
  list_descriptives[[bien_phu_thuoc]] <- df_des_final


  # --- BẢNG 2: TEST OF HOMOGENEITY OF VARIANCES (ĐỦ 4 LOẠI LEVENE KHỚP SPSS) ---
  f_chuoi <- paste(bien_phu_thuoc, "~", bien_phan_nhom)
  lev_mean <- car::leveneTest(as.formula(f_chuoi), data = data_sub, center = mean)
  df1_val <- lev_mean$Df[1]; df2_val <- lev_mean$Df[2]

  w_test <- oneway.test(as.formula(f_chuoi), data = data_sub, var.equal = FALSE)
  df2_adj <- round(w_test$parameter[2], 3)

  lev_median <- car::leveneTest(as.formula(f_chuoi), data = data_sub, center = median)
  lev_trim <- car::leveneTest(as.formula(f_chuoi), data = data_sub, center = function(x) mean(x, trim =
0.05))
```

```
  bang_levene <- data.frame(
    "Biến" = c(bien_phu_thuoc, "", "", ""),
    "Kiểm định phương sai Levene" = c("Based on Mean", "Based on Median", "Based on Median and with adjusted df", "Based on trimmed mean"),
    "Levene Statistic" = as.numeric(round(c(lev_mean$`F value`[1], lev_median$`F value`[1], lev_median$`F value`[1], lev_trim$`F value`[1]), 3)),
    "df1" = as.numeric(c(df1_val, df1_val, df1_val, df1_val)),
    "df2" = as.numeric(c(df2_val, df2_val, df2_adj, df2_val)),
    "Sig." = as.numeric(round(c(lev_mean$`Pr(>F)`[1], lev_median$`Pr(>F)`[1], lev_median$`Pr(>F)`[1], lev_trim$`Pr(>F)`[1]), 3)),
    check.names = FALSE
  )
  list_levene_4types[[bien_phu_thuoc]] <- bang_levene


  # --- BẢNG 3: ANOVA TABLE ---
  anova_fit <- aov(as.formula(f_chuoi), data = data_sub)
  sum_anova <- summary(anova_fit)[[1]]

  ss_between <- sum_anova$`Sum Sq`[1]; df_between <- sum_anova$Df[1]; ms_between = sum_anova$`Mean Sq`[1]
  ss_within  <- sum_anova$`Sum Sq`[2]; df_within  <- sum_anova$Df[2]; ms_within  = sum_anova$`Mean Sq`[2]
  ss_total   <- ss_between + ss_within; df_total   <- df_between + df_within

  bang_anova <- data.frame(
```

```
  "Biến" = c(bien_phu_thuoc, "", ""),
  "Thành phần" = c("Between Groups", "Within Groups", "Total"),
  "Sum of Squares" = as.numeric(round(c(ss_between, ss_within, ss_total), 3)),
  "df" = as.numeric(c(df_between, df_within, df_total)),
  "Mean Square" = c(as.numeric(round(ms_between, 3)), as.numeric(round(ms_within, 3)), NA),
  "F" = c(as.numeric(round(sum_anova$`F value`[1], 3)), NA, NA),
  "Sig." = c(as.numeric(round(sum_anova$`Pr(>F)`[1], 3)), NA, NA),
  check.names = FALSE
)
list_anova_table[[bien_phu_thuoc]] <- bang_anova



# --- BẢNG 4: ROBUST TESTS OF EQUALITY OF MEANS (WELCH TEST) ---
bang_welch <- data.frame(
  "Biến" = bien_phu_thuoc,
  "Kiểm định" = "Welch",
  "Statistic" = as.numeric(round(w_test$statistic, 3)),
  "df1" = as.numeric(w_test$parameter[1]),
  "df2" = as.numeric(round(w_test$parameter[2], 3)),
  "Sig." = as.numeric(round(w_test$p.value, 3)),
  check.names = FALSE
)
list_welch_table[[bien_phu_thuoc]] <- bang_welch
```

```
# --- BẢNG 5: POST-HOC MULTIPLE COMPARISONS (BẢNG SO SÁNH CẶP ĐAN XEN (I) VÀ (J)) ---
# 5.1. Trích xuất cặp Tukey HSD (Đẳng phương sai)
emm <- emmeans(anova_fit, specs = bien_phan_nhom)
pairs_tukey <- as.data.frame(contrast(emm, method = "pairwise", adjust = "tukey"))
labels_pairs <- strsplit(as.character(pairs_tukey$contrast), " - ")
nhom_I_tukey <- sapply(labels_pairs, function(x) x[1])
nhom_J_tukey <- sapply(labels_pairs, function(x) x[2])

df_tk1 <- data.frame(
  "Phép thử" = "Tukey HSD", "(I) Nhóm" = nhom_I_tukey, "(J) Nhóm" = nhom_J_tukey,
  "Mean Difference (I-J)" = round(pairs_tukey$estimate, 4), "Std. Error" = round(pairs_tukey$SE, 4), "Sig." = round(pairs_tukey$p.value, 4),
  "95% CI Lower" = round(pairs_tukey$estimate - (qt(0.975, pairs_tukey$df) * pairs_tukey$SE), 4),
  "95% CI Upper" = round(pairs_tukey$estimate + (qt(0.975, pairs_tukey$df) * pairs_tukey$SE), 4), check.names = FALSE, stringsAsFactors = FALSE
)
df_tk2 <- data.frame(
  "Phép thử" = "Tukey HSD", "(I) Nhóm" = nhom_J_tukey, "(J) Nhóm" = nhom_I_tukey,
  "Mean Difference (I-J)" = round(-pairs_tukey$estimate, 4), "Std. Error" = round(pairs_tukey$SE, 4), "Sig." = round(pairs_tukey$p.value, 4),
  "95% CI Lower" = round(-pairs_tukey$estimate - (qt(0.975, pairs_tukey$df) * pairs_tukey$SE), 4),
  "95% CI Upper" = round(-pairs_tukey$estimate + (qt(0.975, pairs_tukey$df) * pairs_tukey$SE), 4), check.names = FALSE, stringsAsFactors = FALSE
)
df_tk_comb <- rbind(df_tk1, df_tk2)
df_tk_comb <- df_tk_comb[order(df_tk_comb$`(I) Nhóm`, df_tk_comb$`(J) Nhóm`), ]
```

```
# 5.2. Trích xuất cặp Games-Howell bằng PMCMRplus (Không đẳng phương sai)
gh_fit <- PMCMRplus::gamesHowellTest(as.formula(f_chuoi), data = data_sub)
gh_table <- as.data.frame(summary(gh_fit)$p.value)
nhom_names <- rownames(gh_table)

gh_l1 <- list(); gh_l2 <- list(); idx_gh <- 1
for(i in 1:nrow(gh_table)) {
  for(j in 1:ncol(gh_table)) {
    if(!is.na(gh_table[i, j])) {
      i_name <- nhom_names[i]; j_name <- colnames(gh_table)[j]; p_val <- gh_table[i, j]
      mean_i <- mean(data_sub[data_sub[[bien_phan_nhom]] == i_name, bien_phu_thuoc], na.rm=TRUE)
      mean_j <- mean(data_sub[data_sub[[bien_phan_nhom]] == j_name, bien_phu_thuoc], na.rm=TRUE)
      diff_val <- mean_i - mean_j

      gh_l1[[idx_gh]] <- data.frame("Phép thử" = "Games-Howell", "(I) Nhóm" = i_name, "(J) Nhóm" = j_name, "Mean Difference (I-J)" = round(diff_val, 4), "Std. Error" = NA, "Sig." = round(p_val, 3), "95% CI Lower" = NA, "95% CI Upper" = NA, check.names = FALSE, stringsAsFactors = FALSE)
      gh_l2[[idx_gh]] <- data.frame("Phép thử" = "Games-Howell", "(I) Nhóm" = j_name, "(J) Nhóm" = i_name, "Mean Difference (I-J)" = round(-diff_val, 4), "Std. Error" = NA, "Sig." = round(p_val, 3), "95% CI Lower" = NA, "95% CI Upper" = NA, check.names = FALSE, stringsAsFactors = FALSE)
      idx_gh <- idx_gh + 1
    }
  }
}
df_gh_comb <- rbind(do.call(rbind, gh_l1), do.call(rbind, gh_l2))
df_gh_comb <- df_gh_comb[order(df_gh_comb$`(I) Nhóm`, df_gh_comb$`(J) Nhóm`), ]
```

```
  # Gộp 2 phép thử so sánh cặp

  df_var_pairs <- rbind(df_tk_comb, df_gh_comb)

  df_final_pairs <- data.frame("Biến phụ thuộc" = c(bien_phu_thuoc, rep("", nrow(df_var_pairs)-1)), df_var_pairs, check.names = FALSE, stringsAsFactors = FALSE)

  list_posthoc_pairs[[bien_phu_thuoc]] <- df_final_pairs

}


# 7. GỘP CÁC BIẾN THEO CHIỀU DỌC CHO TỪNG LOẠI BẢNG TỔNG HỢP

full_descriptives  <- ldply(list_descriptives, data.frame); full_descriptives$.id <- NULL

full_levene_4types <- ldply(list_levene_4types, data.frame); full_levene_4types$.id <- NULL

full_anova_table   <- ldply(list_anova_table, data.frame); full_anova_table$.id <- NULL

full_welch_table   <- ldply(list_welch_table, data.frame); full_welch_table$.id <- NULL

full_multiple_comparisons <- ldply(list_posthoc_pairs, data.frame); full_multiple_comparisons$.id <- NULL



# 8. HIỂN THỊ TRỰC TIẾP RA MÀN HÌNH CONSOLE RSTUDIO

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

cat("ONE-WAY ANOVA TỔNG HỢP FULL MÓN CHUẨN SPSS\n")

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

cat("BẢNG 1: DESCRIPTIVES STATISTICS\n"); print(full_descriptives, row.names = FALSE)

cat("\nBẢNG 2: TEST OF HOMOGENEITY OF VARIANCES (LEVENE 4 DÒNG)\n"); print(full_levene_4types, row.names = FALSE, na.print = "")

cat("\nBẢNG 3: ANOVA TABLE\n"); print(full_anova_table, row.names = FALSE, na.print = "")

cat("\nBẢNG 4: ROBUST TESTS OF EQUALITY OF MEANS (WELCH)\n"); print(full_welch_table, row.names = FALSE)
```

```
cat("\nBẢNG 5: POST-HOC MULTIPLE COMPARISONS (SO SÁNH CẶP ĐÔI MỘT)\n");
print(full_multiple_comparisons, row.names = FALSE, na.print = ".")
```

9. ĐẨY TOÀN BỘ KẾT QUẢ VÀO 1 SHEET EXCEL TỔNG HỢP (SỐ CĂN PHẢI, TRỐNG NA)

```
writeData(wb, sheet = ten_sheet_duy_nhat, "Descriptives", startRow = dong_hien_tai, startCol = 1)

writeData(wb, sheet = ten_sheet_duy_nhat, full_descriptives, startRow = dong_hien_tai + 1, startCol = 1,
keepNA = FALSE)

dong_hien_tai <- dong_hien_tai + nrow(full_descriptives) + 4


writeData(wb, sheet = ten_sheet_duy_nhat, "Test of Homogeneity of Variances (Levene 4 Lines)",
startRow = dong_hien_tai, startCol = 1)

writeData(wb, sheet = ten_sheet_duy_nhat, full_levene_4types, startRow = dong_hien_tai + 1, startCol =
1, keepNA = FALSE)

dong_hien_tai <- dong_hien_tai + nrow(full_levene_4types) + 4


writeData(wb, sheet = ten_sheet_duy_nhat, "ANOVA", startRow = dong_hien_tai, startCol = 1)

writeData(wb, sheet = ten_sheet_duy_nhat, full_anova_table, startRow = dong_hien_tai + 1, startCol =
1, keepNA = FALSE)

dong_hien_tai <- dong_hien_tai + nrow(full_anova_table) + 4


writeData(wb, sheet = ten_sheet_duy_nhat, "Robust Tests of Equality of Means (Welch)", startRow =
dong_hien_tai, startCol = 1)

writeData(wb, sheet = ten_sheet_duy_nhat, full_welch_table, startRow = dong_hien_tai + 1, startCol = 1,
keepNA = FALSE)

dong_hien_tai <- dong_hien_tai + nrow(full_welch_table) + 4


writeData(wb, sheet = ten_sheet_duy_nhat, "Post Hoc Tests - Multiple Comparisons (So sanh cap)",
startRow = dong_hien_tai, startCol = 1)
```

```
writeData(wb, sheet = ten_sheet_duy_nhat, full_multiple_comparisons, startRow = dong_hien_tai + 1, startCol = 1, keepNA = FALSE)



# 10. TỰ ĐỘNG LƯU FILE EXCEL TỔNG HỢP (CHỐNG GHI ĐÈ THÔNG MINH)

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

saveWorkbook(wb, file = ten_file_excel_chuan, overwrite = TRUE)


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

cat("\n[THÔNG BÁO THÀNH CÔNG]:")

cat("\n- Đã khôi phục và đồng bộ trọn gói hệ thống One-Way ANOVA 5 bảng full món.")

cat("\n- Kết hợp hoàn hảo cả Levene 4 dòng cũ và bảng so sánh cặp (Multiple Comparisons) mới.")

cat("\n- File Excel lưu trữ an toàn tại sheet duy nhất, số căn phải chuẩn đét: ", ten_file_excel_chuan)

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