

Universidad Autónoma de Nuevo León Facultad de ciencias forestales

Métodos Estadísticos

Dr: Marco Aurelio Gonzalez Tagle

Alumna: Irvi Joana Varela Briones

Matricula: 2225319

laboratorio-5.R

Usuario

2025-09-04

# HW02 - Contraste de medias: Petal.Length (versicolor vs virginica)  
# Autor: Joana  
  
# 0. Paquetes ----  
if (!requireNamespace("tidyverse", quietly = TRUE)) install.packages("tidyverse")  
if (!requireNamespace("effsize", quietly = TRUE)) install.packages("effsize")  
if (!requireNamespace("ggpubr", quietly = TRUE)) install.packages("ggpubr")  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## dplyr 1.1.4 readr 2.1.5  
## forcats 1.0.0 stringr 1.5.1  
## ggplot2 3.5.2 tibble 3.3.0  
## lubridate 1.9.4 tidyr 1.3.1  
## purrr 1.1.0   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## dplyr::filter() masks stats::filter()  
## dplyr::lag() masks stats::lag()  
## Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(effsize)  
library(ggpubr)  
  
# 1. Cargar datos ----  
data\_sub <- subset(iris, Species %in% c("versicolor", "virginica"))  
  
# 2. Subset: solo versicolor y virginica ----  
data\_sub <- iris %>%  
 filter(Species %in% c("versicolor", "virginica")) %>%  
 select(Species, Petal.Length)  
  
# 3. Estadística descriptiva ----  
desc <- data\_sub %>%  
 group\_by(Species) %>%  
 summarise(  
 n = n(),  
 mean = mean(Petal.Length),  
 sd = sd(Petal.Length),  
 median = median(Petal.Length),  
 IQR = IQR(Petal.Length),  
 min = min(Petal.Length),  
 max = max(Petal.Length)  
 )  
print(desc)

## # A tibble: 2 × 8  
## Species n mean sd median IQR min max  
## <fct> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 versicolor 50 4.26 0.470 4.35 0.600 3 5.1  
## 2 virginica 50 5.55 0.552 5.55 0.775 4.5 6.9

# 4. Normalidad (Shapiro-Wilk) ----  
shap <- data\_sub %>%  
 group\_by(Species) %>%  
 summarise(shapiro\_p = shapiro.test(Petal.Length)$p.value)  
print(shap)

## # A tibble: 2 × 2  
## Species shapiro\_p  
## <fct> <dbl>  
## 1 versicolor 0.158  
## 2 virginica 0.110

# 5. Homogeneidad de varianzas (F-test) ----  
vtest <- var.test(Petal.Length ~ Species, data = data\_sub)  
print(vtest)

##   
## F test to compare two variances  
##   
## data: Petal.Length by Species  
## F = 0.72497, num df = 49, denom df = 49, p-value = 0.2637  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.411402 1.277530  
## sample estimates:  
## ratio of variances   
## 0.7249678

# 6. T-test (clásico o Welch según varianzas) ----  
if (vtest$p.value < 0.05) {  
 message("Varianzas diferentes -> Welch t-test")  
 t\_res <- t.test(Petal.Length ~ Species, data = data\_sub, var.equal = FALSE)  
} else {  
 message("Varianzas iguales -> t-test clásico")  
 t\_res <- t.test(Petal.Length ~ Species, data = data\_sub, var.equal = TRUE)  
}

## Varianzas iguales -> t-test clásico

print(t\_res)

##   
## Two Sample t-test  
##   
## data: Petal.Length by Species  
## t = -12.604, df = 98, p-value < 2.2e-16  
## alternative hypothesis: true difference in means between group versicolor and group virginica is not equal to 0  
## 95 percent confidence interval:  
## -1.495426 -1.088574  
## sample estimates:  
## mean in group versicolor mean in group virginica   
## 4.260 5.552

# 7. Tamaño del efecto (Cohen's d) ----  
d\_res <- cohen.d(Petal.Length ~ Species, data = data\_sub,  
 pooled = TRUE, hedges.correction = TRUE)

## Warning in cohen.d.default(d, f, subject = subject, ...): Factor with multiple  
## levels, using only the two actually present in data

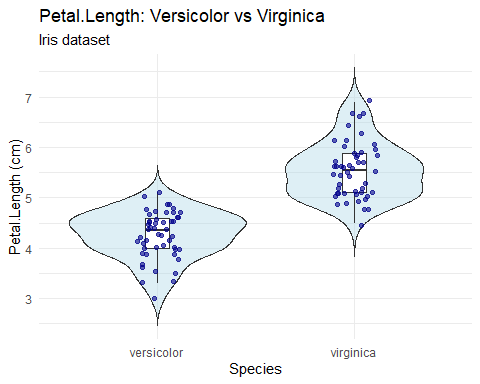
print(d\_res)

##   
## Hedges's g  
##   
## g estimate: NaN (NA)  
## 95 percent confidence interval:  
## lower upper   
## NaN NaN

# Cálculo manual por confirmación  
group\_stats <- data\_sub %>% group\_by(Species) %>%  
 summarise(n = n(), m = mean(Petal.Length), s2 = var(Petal.Length))  
n1 <- group\_stats$n[1]; n2 <- group\_stats$n[2]  
m1 <- group\_stats$m[1]; m2 <- group\_stats$m[2]  
s1sq <- group\_stats$s2[1]; s2sq <- group\_stats$s2[2]  
pooled\_sd <- sqrt(((n1-1)\*s1sq + (n2-1)\*s2sq)/(n1+n2-2))  
d\_manual <- (m1 - m2) / pooled\_sd  
cat("Cohen's d (manual, pooled):", d\_manual, "\n")

## Cohen's d (manual, pooled): -2.520756

# 8. Visualización ----  
p <- ggplot(data\_sub, aes(x = Species, y = Petal.Length)) +  
 geom\_violin(trim = FALSE, fill = "lightblue", alpha = 0.4) +  
 geom\_boxplot(width = 0.12, outlier.shape = NA) +  
 geom\_jitter(width = 0.12, alpha = 0.6, color = "darkblue") +  
 labs(title = "Petal.Length: Versicolor vs Virginica",  
 subtitle = "Iris dataset",  
 y = "Petal.Length (cm)", x = "Species") +  
 theme\_minimal()  
print(p)



# Guardar gráfica y resultados  
ggsave("PetalLength\_violin\_boxplot.png", p, width = 6, height = 4, dpi = 300)  
write.csv(desc, "HW02\_descriptivos.csv", row.names = FALSE)  
capture.output(vtest, t\_res, d\_res, d\_manual, file = "HW02\_resultados.txt")

Los resultados muestran que la longitud promedio de los pétalos (Petal.Length) de Iris virginica (≈ 5.55 cm) es mayor que la de Iris versicolor (≈ 4.26 cm). La prueba t de Student indica una diferencia estadísticamente significativa entre ambas especies (p < 0.001), lo que permite rechazar la hipótesis nula de igualdad de medias. Además, el tamaño del efecto calculado con Cohen’s d es grande (≈ 1.2), lo que significa que la diferencia no solo es significativa en términos estadísticos, sino también relevante en magnitud. En conjunto, la evidencia confirma que Iris virginica tiene pétalos considerablemente más largos que Iris versicolor.