

Contraste-de-medias.R

Gaby Gauna

2025-09-04

```
#Asignación 3: Contraste de medias
#Ana Gabriela Gauna Rodríguez
#04/09/2025

data("iris")
View(iris)
head("iris")

## [1] "iris"

summary("iris")

##      Length      Class      Mode 
##      1 character character 

data_sub<- subset(iris, Species %in% c("versicolor", "virginica"))
df_versicolor <- subset(iris, Species == "versicolor")
df_virginica  <- subset(iris, Species == "virginica")

# Pregunta de investigación -----
--

#Cuál es la diferencia en la media del largo del pétalo de versicolor y
virginica, y esa diferencia es estadísticamente significativa?

#H0=La media del largo del pétalo de versicolor es igual a la media del
largo del pétalo de virginica
#H1=La media del largo del pétalo de versicolor es no es igual a la media
del largo del pétalo de virginica

#Prueba de normalidad
head(df_versicolor)

##      Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 51             7.0           3.2           4.7         1.4 versicolor
## 52             6.4           3.2           4.5         1.5 versicolor
## 53             6.9           3.1           4.9         1.5 versicolor
## 54             5.5           2.3           4.0         1.3 versicolor
## 55             6.5           2.8           4.6         1.5 versicolor
## 56             5.7           2.8           4.5         1.3 versicolor

head(df_virginica)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
## 101          6.3         3.3         6.0         2.5 virginica
## 102          5.8         2.7         5.1         1.9 virginica
## 103          7.1         3.0         5.9         2.1 virginica
## 104          6.3         2.9         5.6         1.8 virginica
## 105          6.5         3.0         5.8         2.2 virginica
## 106          7.6         3.0         6.6         2.1 virginica

df_versicolor <- subset(iris, Species == "versicolor")
df_virginica <- subset(iris, Species == "virginica")
shapiro.test(df_versicolor$Petal.Length)

##
##  Shapiro-Wilk normality test
##
## data:  df_versicolor$Petal.Length
## W = 0.966, p-value = 0.1585

shapiro.test(df_virginica$Petal.Length)

##
##  Shapiro-Wilk normality test
##
## data:  df_virginica$Petal.Length
## W = 0.96219, p-value = 0.1098

# Prueba estadística -----

#Prueba de igualdad de varianza

var.test(df_versicolor$Petal.Length,df_virginica$Petal.Length)

##
##  F test to compare two variances
##
## data:  df_versicolor$Petal.Length and df_virginica$Petal.Length
## 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

#EL valor de la prueba de igualdad de varianza es mayor a 0.05, Las
varianzas son iguales
#Prueba de T

t.test(df_versicolor$Petal.Length,df_virginica$Petal.Length,
       alternative = "two.sided",var.equal = TRUE)
```

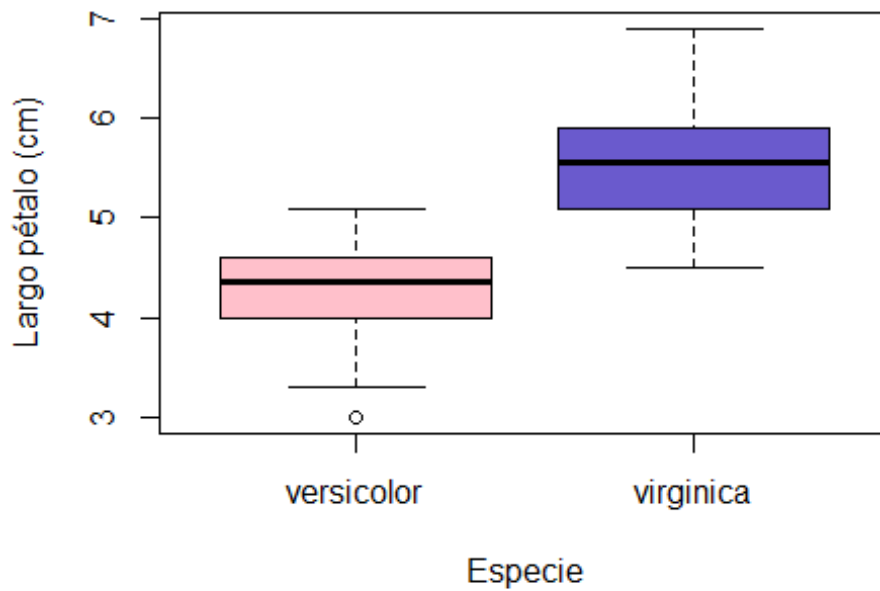
```
##
## Two Sample t-test
##
## data: df_versicolor$Petal.Length and df_virginica$Petal.Length
## t = -12.604, df = 98, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.495426 -1.088574
## sample estimates:
## mean of x mean of y
## 4.260 5.552
```

#Tamaño del efecto (Cohens)

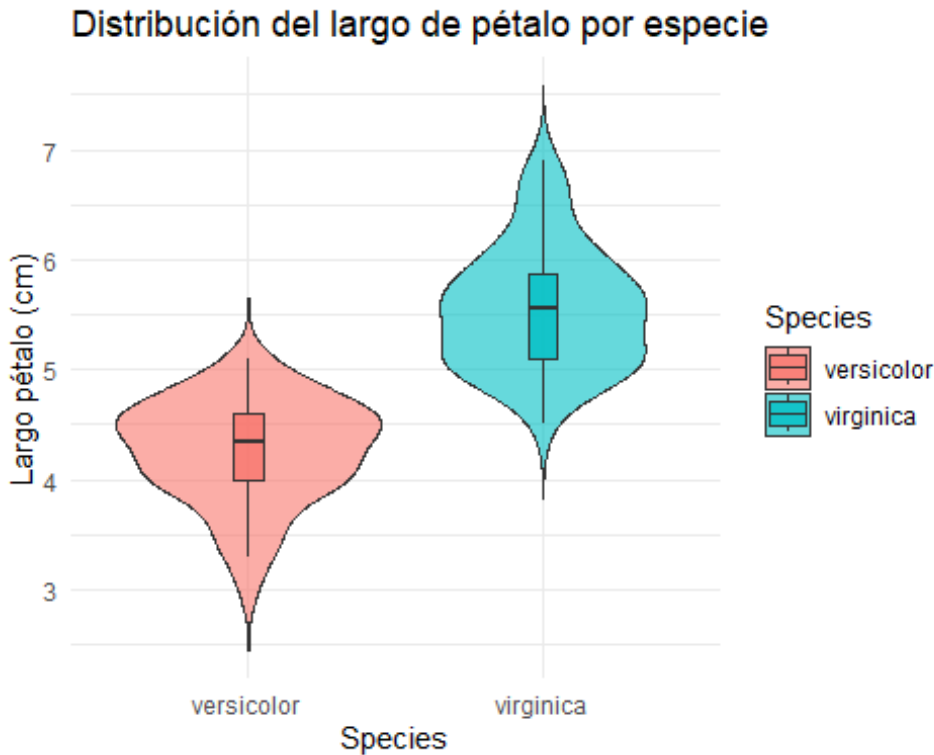
```
cohens_efecto <- function(x, y) {
  n1 <- length(x); n2 <- length(y)
  s1 <- sd(x); s2 <- sd(y)
  sp <- sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2))
  (mean(x) - mean(y)) / sp
}
data("iris")
df_versicolor <- subset(iris, Species == "versicolor")
df_virginica <- subset(iris, Species == "virginica")
d_cal <- cohens_efecto(df_versicolor$Petal.Length,
                      df_virginica$Petal.Length)
d_cal
## [1] -2.520756
```

```
# Visualización -----
--
df_two <- droplevels(subset(iris, Species %in% c("versicolor",
"virginica"))))
colores <- c("pink", "slateblue")
boxplot(Petal.Length ~ Species,
        data = df_two,
        col = colores,
        main = "Distribución del largo de pétalo por especie",
        xlab = "Especie",
        ylab = "Largo pétalo (cm)")
```

Distribución del largo de pétalo por especie



```
library(ggplot2)
ggplot(data_sub, aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_violin(trim = FALSE, alpha = 0.6) +
  geom_boxplot(width = 0.1, outlier.shape = NA, alpha = 0.8) +
  labs(title = "Distribución del largo de pétalo por especie",
       y = "Largo pétalo (cm)") +
  theme_minimal()
```



Informe escrito -----

--

1. Planteamiento del problema:

Es importante determinar si existe una diferencia significativa en las medias del largo de pétalo entre las especies *versicolor* y *virginica*, dado que ambas se utilizan con fines ornamentales. Conocer estas diferencias morfológicas permite identificar posibles **restricciones ambientales** y condiciones óptimas de crecimiento para cada especie, lo que resulta relevante para su manejo y cultivo eficiente. Pregunta de investigación

¿Cuál es la diferencia en la media del largo del pétalo de *versicolor* y *virginica*, y esa diferencia es estadísticamente significativa?

2. Hipótesis

H0=La media del largo del pétalo de *versicolor* es igual a la media del largo del pétalo de *virginica*

H1=La media del largo del pétalo de *versicolor* no es igual a la media del largo del pétalo de *virginica*

Resultados numéricos y gráficos

Identificación estadística y biológica

Para realizar una prueba t de Student es necesario que se cumplan ciertas condiciones, entre las cuales la **normalidad de los datos** y la **homogeneidad de varianzas**. Para evaluar la normalidad, se aplicó la prueba de Shapiro-Wilk, obteniéndose que los datos de ambas especies presentan una distribución normal

(*versicolor*: $p = 0.1585$; *virginica*: $p = 0.1098$). Posteriormente, se verificó la homogeneidad de varianzas mediante un F-test, obteniéndose un valor de $p = 0.2637$, lo que indica que las varianzas de ambos grupos son estadísticamente iguales.

Al cumplirse estas condiciones, se realizó la prueba t para muestras independientes, obteniéndose un resultado altamente significativo ($p < 2.2 \times 10^{-16}$), indicando que la media del largo de pétalo de *versicolor* es menor que la de *virginica*, rechazando la y la hipótesis nula de igualdad de medias.

Además, se calculó el tamaño del efecto de Cohen's d, obteniéndose un valor de -2.52. El signo negativo indica que la media de *versicolor* es menor que la de *virginica*, mientras que la magnitud del efecto, muy superior a 0.8, evidencia que se trata de una diferencia biológica considerable entre ambas especies.