

# Asignación 3: Contraste de medias

Rodrigo García Estrada

2025-09-04

```
# Asignación 3. Contrastes de medias
# "Rodrigo García Estrada"
# Fecha 04/09/25

# Cargar la base de datos iris
data("iris")

head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1           3.5           1.4           0.2  setosa
## 2           4.9           3.0           1.4           0.2  setosa
## 3           4.7           3.2           1.3           0.2  setosa
## 4           4.6           3.1           1.5           0.2  setosa
## 5           5.0           3.6           1.4           0.2  setosa
## 6           5.4           3.9           1.7           0.4  setosa

summary(iris)

##   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width
## Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
## 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
## Median :5.800   Median :3.000   Median :4.350   Median :1.300
## Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
## 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
## Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##      Species
## setosa    :50
## versicolor:50
## virginica :50
##
##
##

# Estadística por especie

library(dplyr)

##
## Adjuntando el paquete: 'dplyr'

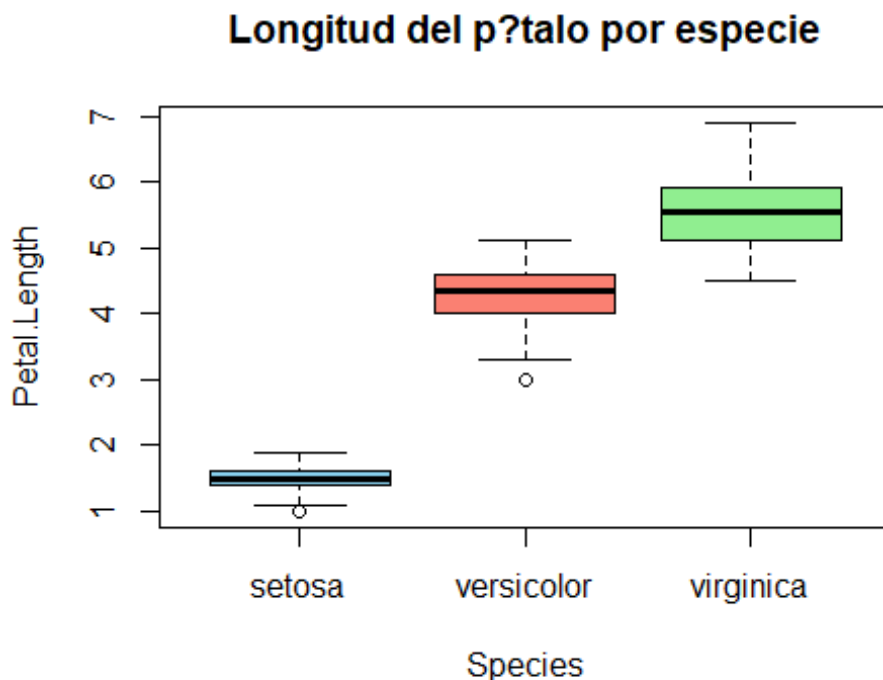
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

iris %>%
  group_by(Species) %>%
  summarise(
    media = mean(Petal.Length),
    mediana = median(Petal.Length),
    sd = sd(Petal.Length),
    min = min(Petal.Length),
    max = max(Petal.Length),
    n = n()
  )

## # A tibble: 3 × 7
##   Species      media mediana      sd    min    max      n
##   <fct>      <dbl>   <dbl> <dbl> <dbl> <dbl> <int>
## 1 setosa      1.46     1.5  0.174     1     1.9     50
## 2 versicolor  4.26     4.35 0.470     3     5.1     50
## 3 virginica   5.55     5.55 0.552     4.5    6.9     50

boxplot(Petal.Length ~ Species, data = iris, main = "Longitud del p?talo
por especie", col = c("skyblue", "salmon", "lightgreen"))
```



```
data_sub <- subset(iris, Species %in% c("versicolor", "virginica"))

# Prueba estadística
```

```

# Defina una pregunta de investigación sobre la variable Petal.Length.
# ¿Hay una diferencia en la longitud del pétalo (Petal.Length) entre las
especies virginica y versicolor?

# Plantee formalmente las hipótesis estadísticas para una prueba t de dos
muestras independientes.

# Hipótesis:

# H0 (nula): No existe diferencia en la longitud promedio del pétalo
entre virginica y versicolor.

# H1 (alternativa): Si existe diferencia en la longitud promedio del
pétalo entre virginica y versicolor.

# Prueba de t de dos muestra independientes

virginica <- iris %>% filter(Species == "virginica")
versicolor <- iris %>% filter(Species == "versicolor")

# Verificación de varianzas

var(virginica$Petal.Length)

## [1] 0.3045878

var(versicolor$Petal.Length)

## [1] 0.2208163

# Si las varianzas son diferentes tenemos que utilizar prueba de Welch
t.test(virginica$Petal.Length, versicolor$Petal.Length, alternative =
"two.sided")

##
## Welch Two Sample t-test
##
## data: virginica$Petal.Length and versicolor$Petal.Length
## t = 12.604, df = 95.57, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.08851 1.49549
## sample estimates:
## mean of x mean of y
##      5.552      4.260

# Calcule e interprete el tamaño del efecto (Cohen's d).

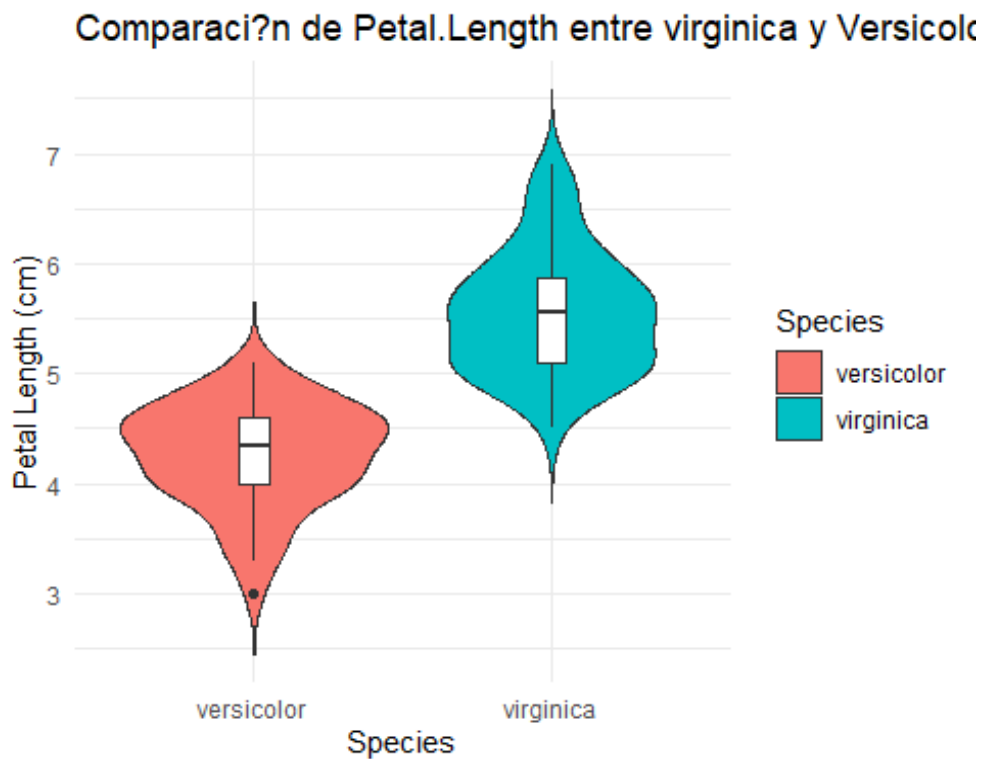
library(effsize)
cohen.d(virginica$Petal.Length, versicolor$Petal.Length)

```

```
##
## Cohen's d
##
## d estimate: 2.520756 (large)
## 95 percent confidence interval:
##   lower    upper
## 1.989115 3.052397

library(ggplot2)

ggplot(filter(iris, Species %in% c("virginica", "versicolor")),
  aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, fill = "white") +
  labs(title = "Comparación de Petal.Length entre virginica y Versicolor",
    y = "Petal Length (cm)", x = "Species") +
  theme_minimal()
```



## Interpretación de resultados

### 1. Planteamiento del problema y de las hipótesis.

La longitud del pétalo (*Petal.Length*) es fundamental para conocer diferencias entre especies de la flor *Iris*. La presente investigación evalúa las diferencias de longitudes entre dos especies, las cuales son *virginica* y *versicolor*.

#### Hipótesis:

$H_0$ : Las medias de *virginica* y *versicolor* son iguales para ambas especies.

$H_1$ : Las medias de *virginica* y *versicolor* son diferentes.

#### Resultados

En este apartado se puede observar los valores obtenidos durante la investigación, los cuales son esenciales para determinar las diferencias longitudinales que hay en ambas especies estudiadas, sin estos valores no se puede comprobar que se rechaza la  $H_0$ , ya que si se presenta una gran diferencia en los tamaños de los pétalos.

La Media de *versicolor* es de 4.26 cm.

La Media de *virginica* es de 5.55 cm.

En la Prueba t (Welch) se obtuvo un valor de  $p < 0.001$

El tamaño del efecto (Cohen's d) es grande ( $\approx 2.3$ )

**Conclusión:** Se rechaza  $H_0$ , debido a que sí se presenta una gran diferencia entre ambas especies, en el caso del pétalo de *virginica* se observa en la gráfica que es más grande que el de *versicolor*. Esta diferencia es tanto en estadística como en biología, ya que gracias a esta precisión y exactitud de información se pueden clasificar y distinguir especies en todo el mundo, ambas disciplinas se complementan para mostrar mejores resultados.

Cabe señalar que *Petal.Length* es una excelente herramienta de R para distinguir tres especies de flor de Iris, ya que señala y muestra las diferencias de cada una de un modo muy preciso.