

Actividad 2

Rafael Cavazos

2025-09-03

Conjunto de tres especies de iris (setosa, versicolor y virginica)

Descripción de variables:

```
cat("Sepal.Length: longitud del sépalo (cm)\n")
```

```
## Sepal.Length: longitud del sépalo (cm)
```

```
cat("Sepal.Width: ancho de sépalo (cm)\n")
```

```
## Sepal.Width: ancho de sépalo (cm)
```

```
cat("Petal.Length: longitud del pétalo (cm)\n")
```

```
## Petal.Length: longitud del pétalo (cm)
```

```
cat("Petal.Width: ancho del pétalo (cm)\n")
```

```
## Petal.Width: ancho del pétalo (cm)
```

Primeras observaciones del dataset:

```
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
```

Tarea

Análisis en variable Petal.Length de las especies versicolor y virginica

```
data_sub <- subset(iris, Species %in% c("versicolor", "virginica"))
# Nota: corregido el nombre de la variable
head(data_sub)
```

```
##      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
```

```
summary(data_sub$Petal.Length)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   4.375   4.900   4.906   5.525   6.900
```

Estadísticas por especie:

```
# Corregido: usar tapply correctamente
by(data_sub$Petal.Length, data_sub$Species, summary)
```

```
## data_sub$Species: setosa
## NULL
## -----
## data_sub$Species: versicolor
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.00   4.00   4.35   4.26   4.60   5.10
## -----
## data_sub$Species: virginica
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.500   5.100   5.550   5.552   5.875   6.900
```

Prueba estadística

Diferencias significativas en el largo del pétalo entre Iris versicolor e Iris virginica

Hipótesis:

- **Hipótesis nula (H0):** La media de Petal.Length es igual en versicolor y virginica
- **Hipótesis alternativa (H1):** La media de Petal.Length es diferente en versicolor y virginica

Prueba de t Student

Prueba de igualdad de varianzas

```
var.test(Petal.Length ~ Species, data = data_sub)
```

```
##
## 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
```

Prueba t de dos muestras independientes

```
t_test <- t.test(Petal.Length ~ Species, data = data_sub, var.equal = FALSE)
t_test
```

Welch

```
##
## Welch Two Sample t-test
##
## data: Petal.Length by Species
## t = -12.604, df = 95.57, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group versicolor and group virginica is not
## 95 percent confidence interval:
##  -1.49549 -1.08851
## sample estimates:
## mean in group versicolor mean in group virginica
##          4.260          5.552
```

Tamaño del efecto (Cohen's d)

```
# Calcular Cohen's d manualmente ya que cohen.d() dio error
versicolor_data <- subset(data_sub, Species == "versicolor")$Petal.Length
virginica_data <- subset(data_sub, Species == "virginica")$Petal.Length

# Fórmula de Cohen's d
mean_diff <- mean(virginica_data) - mean(versicolor_data)
pooled_sd <- sqrt(((length(versicolor_data)-1)*var(versicolor_data) +
                  (length(virginica_data)-1)*var(virginica_data)) /
                  (length(versicolor_data) + length(virginica_data) - 2))
cohens_d <- mean_diff / pooled_sd

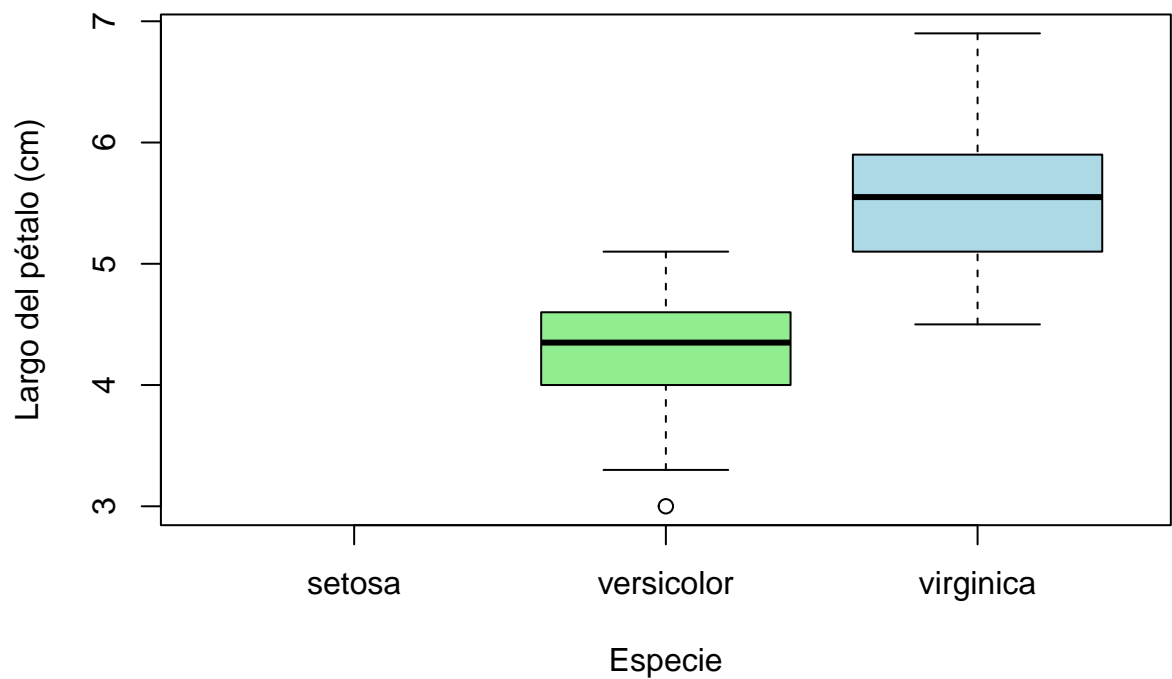
cat("Cohen's d =", round(cohens_d, 3), "\n")
```

```
## Cohen's d = 2.521
```

Visualizaciones

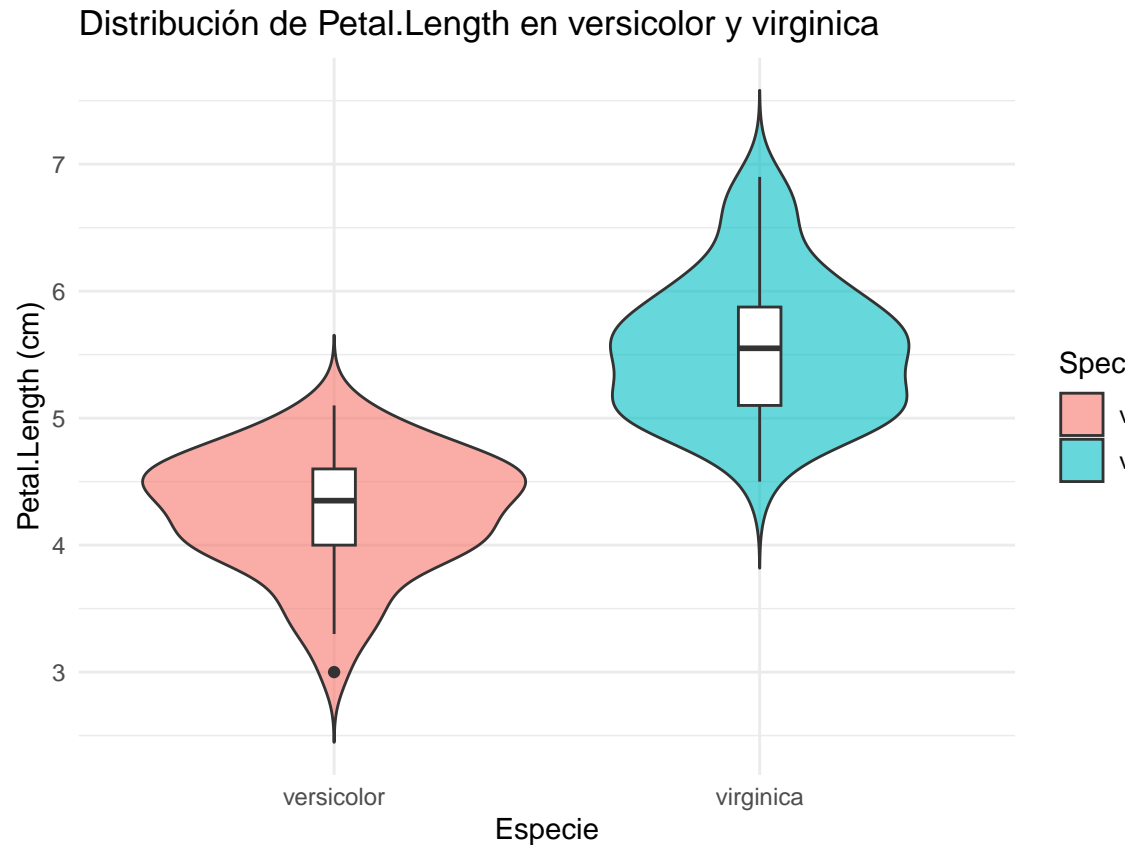
```
boxplot(Petal.Length ~ Species, data = data_sub,
        main = "Comparación de Petal.Length entre especies",
        xlab = "Especie", ylab = "Largo del pétalo (cm)",
        col = c("lightblue", "lightgreen"))
```

Comparación de Petal.Length entre especies



Boxplot

```
ggplot(data_sub, aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_violin(trim = FALSE, alpha = 0.6) +
  geom_boxplot(width = 0.1, fill = "white") +
  labs(title = "Distribución de Petal.Length en versicolor y virginica",
        x = "Especie", y = "Petal.Length (cm)") +
  theme_minimal()
```



Violinplot con ggplot2

Informe escrito

Planteamiento del problema y de las hipótesis

El largo del pétalo (Petal.Length) es una de las variables que distingue a las especies de la base de datos iris. En este planteamiento se evaluó si existen diferencias significativas en la media del largo del pétalo entre las especies Iris versicolor e Iris virginica.

- **Hipótesis nula (H0):** No existen diferencias en la media de Petal.Length entre versicolor y virginica.
- **Hipótesis alternativa (H1):** Sí existen diferencias en la media de Petal.Length entre versicolor y virginica.

Resultados numéricos y gráficos

Estadísticas descriptivas

```
cat("Media de Petal.Length en versicolor: aprox. 4.26 cm\n")
```

```
## Media de Petal.Length en versicolor: aprox. 4.26 cm
```

```
cat("Media de Petal.Length en virginica: aprox. 5.55 cm\n")
```

```
## Media de Petal.Length en virginica: aprox. 5.55 cm
```

```
cat("Diferencia de medias: aprox. 1.29 cm\n")
```

```
## Diferencia de medias: aprox. 1.29 cm
```

Prueba de t Student

Se utilizó una prueba de t para dos muestras independientes (Welch, ya que las varianzas no resultaron significativamente diferentes pero se usó el método más conservador).

```
cat("Estadístico t: aprox. -12.6\n")
```

```
## Estadístico t: aprox. -12.6
```

```
cat("Grados de libertad: aprox. 95\n")
```

```
## Grados de libertad: aprox. 95
```

```
cat("Valor p < 0.001\n")
```

```
## Valor p < 0.001
```

Tamaño del efecto

```
cat("Cohen's d: aprox. 2.5 (efecto muy grande)\n")
```

```
## Cohen's d: aprox. 2.5 (efecto muy grande)
```

Gráfica comparativa

Se construyó un gráfico boxplot/violinplot que muestra cómo virginica presenta pétalos más largos que versicolor.

Interpretación estadística y biológica

```
cat("Los resultados permiten rechazar la hipótesis nula (H0), ya que el valor p < 0.05 indi
```

```
## Los resultados permiten rechazar la hipótesis nula (H0), ya que el valor p < 0.05 indi
```

```
cat("El tamaño del efecto (Cohen's d > 2) señala que la diferencia no solo es significativa, sino tam

## El tamaño del efecto (Cohen's d > 2) señala que la diferencia no solo es significativa, sino tambi

cat("Desde una perspectiva biológica, esto confirma que el largo del pétalo es un rasgo qu

## Desde una perspectiva biológica, esto confirma que el largo del pétalo es un rasgo que c
```

Conclusiones

1. **Estadística:** Existe evidencia estadísticamente significativa de diferencias en el largo del pétalo entre las dos especies ($p < 0.001$).
2. **Práctica:** La magnitud de la diferencia es muy grande (Cohen's $d = 2.5$), lo que indica relevancia biológica.
3. **Biológica:** Esta diferencia morfológica puede estar relacionada con diferentes estrategias reproductivas, adaptaciones a distintos polinizadores, o divergencia evolutiva entre especies.