

HW_02: Contraste de medias

Denisse Aurora Alvarado Reyna

2025-09-03

Análisis de diferencias en largo de pétalo entre especies versicolor y virginica

Exploración de datos

```
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
##
##
##
```

```
str(iris)
```

```
## 'data.frame':   150 obs. of  5 variables:
## $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
```

```
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
table(iris$Species)
```

```
##
##      setosa versicolor  virginica
##         50         50         50
```

Selección de especies: versicolor y virginica

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

```
## [1] 100
```

```
table(data_sub$Species)
```

```
##
##      setosa versicolor  virginica
##         0         50         50
```

Estadísticas descriptivas del largo del pétalo

```
aggregate(Petal.Length ~ Species, data = data_sub, FUN = function(x) {
  c(n = length(x),
    media = mean(x),
    mediana = median(x),
    desv_est = sd(x),
    min = min(x),
    max = max(x))
})
```

```
##      Species Petal.Length.n Petal.Length.media Petal.Length.mediana
## 1 versicolor    50.0000000      4.2600000      4.3500000
## 2 virginica     50.0000000      5.5520000      5.5500000
##      Petal.Length.desv_est Petal.Length.min Petal.Length.max
## 1          0.4699110      3.0000000      5.1000000
## 2          0.5518947      4.5000000      6.9000000
```

```
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
```

Pregunta de investigación e hipótesis

Pregunta de investigación

¿Existe diferencia significativa en la media del largo de pétalo entre las especies versicolor y virginica de iris?

Hipótesis estadísticas

- **H:** = (No hay diferencia en las medias del largo de pétalo entre especies)
- **H:** (Hay diferencia en las medias del largo de pétalo entre especies)
- **Nivel de significancia:** = 0.05

Verificación de supuestos

```
versicolor_data <- data_sub$Petal.Length[data_sub$Species == "versicolor"]
virginica_data <- data_sub$Petal.Length[data_sub$Species == "virginica"]

# Pruebas de normalidad
shapiro_versicolor <- shapiro.test(versicolor_data)
shapiro_virginica <- shapiro.test(virginica_data)

cat("p-value Shapiro-Wilk versicolor:", shapiro_versicolor$p.value, "\n")
```

```
## p-value Shapiro-Wilk versicolor: 0.1584778
```

```
cat("p-value Shapiro-Wilk virginica:", shapiro_virginica$p.value, "\n")
```

```
## p-value Shapiro-Wilk virginica: 0.1097754
```

```
# Prueba de igualdad de varianzas (Prueba F)
var_test <- var.test(versicolor_data, virginica_data)
cat("p-value prueba F de varianzas:", var_test$p.value, "\n")
```

```
## p-value prueba F de varianzas: 0.2637454
```

```
equal_var <- var_test$p.value > 0.05
cat("¿Varianzas iguales?", equal_var, "\n")
```

```
## ¿Varianzas iguales? TRUE
```

```
# Justificación del tipo de prueba
if(equal_var) {
  cat("Se aplicará prueba t clásica (varianzas iguales)\n")
} else {
  cat("Se aplicará prueba t de Welch (varianzas diferentes)\n")
}
```

```
## Se aplicará prueba t clásica (varianzas iguales)
```

Prueba t de dos muestras independientes

```
t_test_result <- t.test(Petal.Length ~ Species,
  data = data_sub,
  var.equal = equal_var,
  alternative = "two.sided",
  conf.level = 0.95)
```

```
t_test_result
```

```
##
```

```
## 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
```

```
## 95 percent confidence interval:
```

```
## -1.495426 -1.088574
```

```
## sample estimates:
```

```
## mean in group versicolor mean in group virginica
```

```
## 4.260 5.552
```

Cálculo del tamaño del efecto (Cohen's d)

```
cohens_d <- function(x, y) {
  n1 <- length(x)
  n2 <- length(y)
  s_pooled <- sqrt(((n1-1)*var(x) + (n2-1)*var(y)) / (n1+n2-2))
  d <- (mean(x) - mean(y)) / s_pooled
  return(d)
}
```

```
cohen_d <- cohens_d(virginica_data, versicolor_data)
cat("Cohen's d:", round(cohen_d, 3), "\n")
```

```
## Cohen's d: 2.521
```

```
efecto_interpretacion <- if(abs(cohen_d) < 0.2) {
  "pequeño"
} else if(abs(cohen_d) < 0.5) {
```

```

    "pequeño a mediano"
  } else if(abs(cohen_d) < 0.8) {
    "mediano a grande"
  } else {
    "grande"
  }
  cat("Interpretación del tamaño del efecto:", efecto_interpretacion, "\n")

```

```
## Interpretación del tamaño del efecto: grande
```

Visualizaciones

Boxplot comparativo

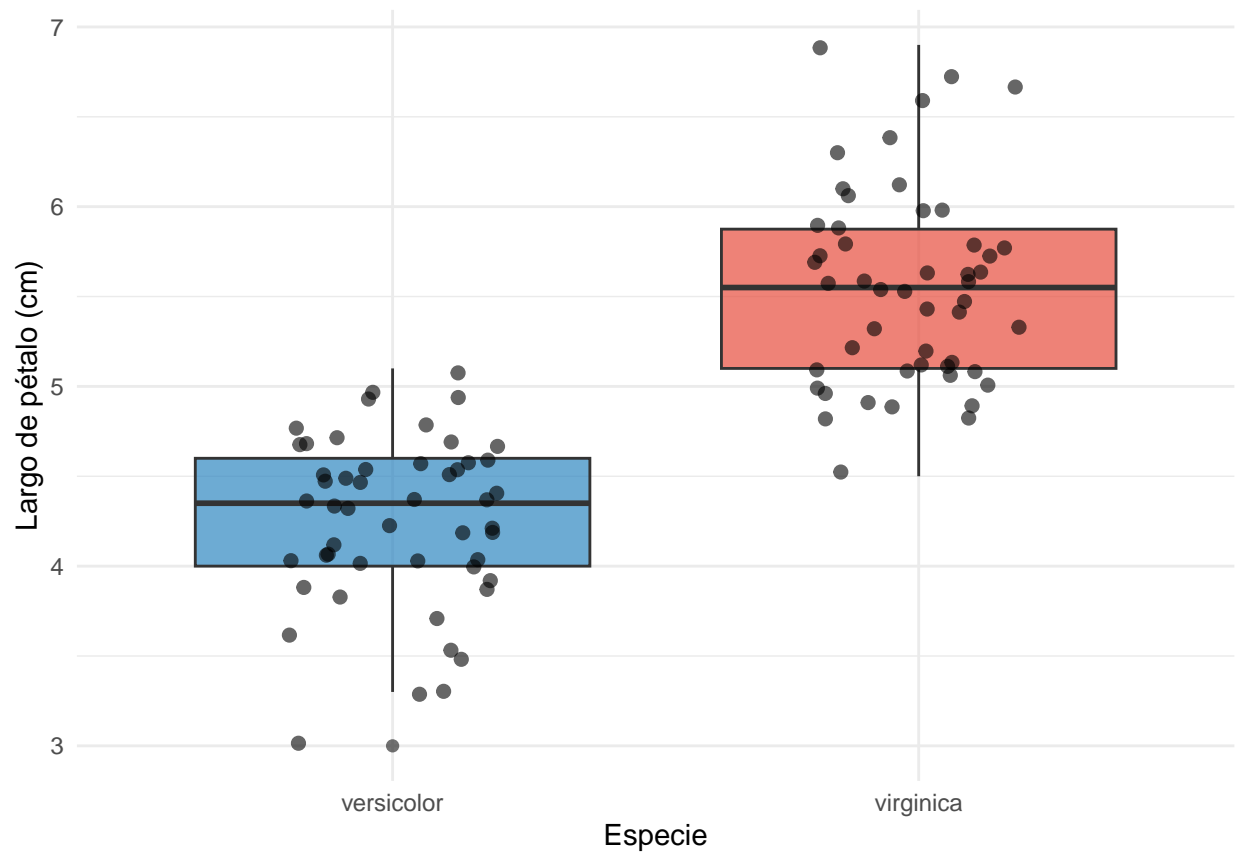
```

ggplot(data_sub, aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_boxplot(alpha = 0.7, outlier.shape = 16, outlier.size = 2) +
  geom_jitter(width = 0.2, alpha = 0.6, size = 2) +
  scale_fill_manual(values = c("versicolor" = "#2E86C1", "virginica" = "#E74C3C")) +
  labs(title = "Comparación del Largo de Pétalo entre Especies",
       subtitle = "Iris versicolor vs. virginica",
       x = "Especie", y = "Largo de pétalo (cm)") +
  theme_minimal() +
  theme(legend.position = "none")

```

Comparación del Largo de Pétalo entre Especies

Iris versicolor vs. virginica

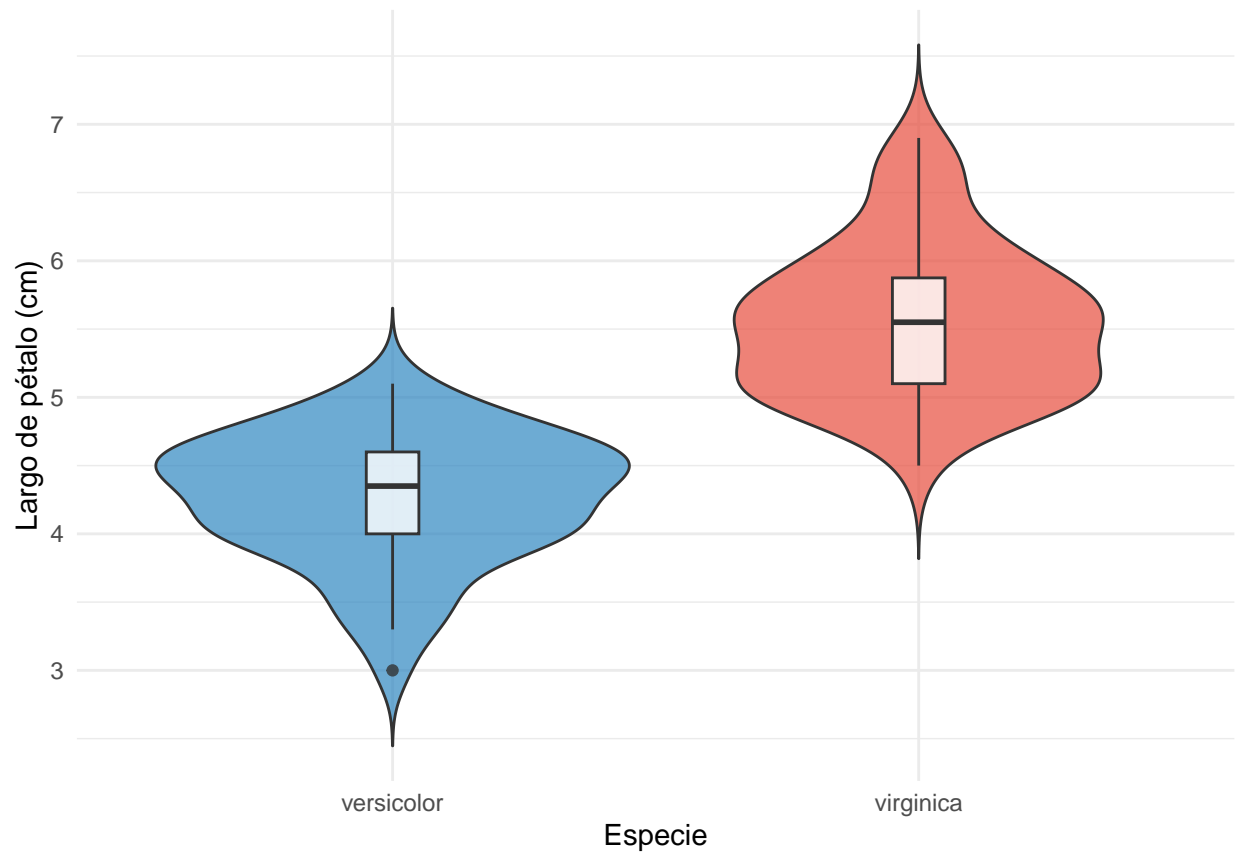


Violin plot

```
ggplot(data_sub, aes(x = Species, y = Petal.Length, fill = Species)) +  
  geom_violin(alpha = 0.7, trim = FALSE) +  
  geom_boxplot(width = 0.1, fill = "white", alpha = 0.8) +  
  scale_fill_manual(values = c("versicolor" = "#2E86C1", "virginica" = "#E74C3C")) +  
  labs(title = "Distribución del Largo de Pétalo por Especie",  
        subtitle = "Violin plot con boxplot superpuesto",  
        x = "Especie", y = "Largo de pétalo (cm)") +  
  theme_minimal() +  
  theme(legend.position = "none")
```

Distribución del Largo de Pétalo por Especie

Violin plot con boxplot superpuesto



Resumen de resultados

```
cat("=== RESUMEN DEL ANÁLISIS ESTADÍSTICO ===\n\n")
```

```
## === RESUMEN DEL ANÁLISIS ESTADÍSTICO ===
```

```
cat("Muestra: n_versicolor =", length(versicolor_data),  
    ", n_virginica =", length(virginica_data), "\n\n")
```

```
## Muestra: n_versicolor = 50 , n_virginica = 50
```

```
cat("MEDIAS:\n")
```

```
## MEDIAS:
```

```
cat("Media versicolor:", round(mean(versicolor_data), 3), "cm\n")
```

```
## Media versicolor: 4.26 cm
```

```

cat("Media virginica:", round(mean(virginica_data), 3), "cm\n")

## Media virginica: 5.552 cm

cat("Diferencia de medias:", round(mean(virginica_data) - mean(versicolor_data), 3), "cm\n\n")

## Diferencia de medias: 1.292 cm

cat("RESULTADOS DE LA PRUEBA t:\n")

## RESULTADOS DE LA PRUEBA t:

cat("Estadístico t =", round(t_test_result$statistic, 3), "\n")

## Estadístico t = -12.604

cat("Grados de libertad =", round(t_test_result$parameter, 1), "\n")

## Grados de libertad = 98

cat("p-value =", format(t_test_result$p.value, scientific = TRUE, digits = 3), "\n")

## p-value = 3.18e-22

cat("Intervalo de confianza 95%: [",
    round(t_test_result$conf.int[1], 3), ",",
    round(t_test_result$conf.int[2], 3), "]\n\n")

## Intervalo de confianza 95%: [ -1.495 , -1.089 ]

cat("TAMAÑO DEL EFECTO:\n")

## TAMAÑO DEL EFECTO:

cat("Cohen's d =", round(cohen_d, 3), "(tamaño del efecto", efecto_interpretacion, ")\n\n")

## Cohen's d = 2.521 (tamaño del efecto grande )

cat("DECISIÓN ESTADÍSTICA:\n")

## DECISIÓN ESTADÍSTICA:

if(t_test_result$p.value < 0.05) {
  cat("Rechazamos H . Existe diferencia significativa entre las medias del largo de pétalo.\n")
} else {
  cat("No rechazamos H . No hay evidencia de diferencia significativa.\n")
}

## Rechazamos H . Existe diferencia significativa entre las medias del largo de pétalo.

```



```
cat("\nINTERPRETACIÓN BIOLÓGICA:\n")

##
## INTERPRETACIÓN BIOLÓGICA:

if(t_test_result$p.value < 0.05) {
  cat("Las especies versicolor y virginica presentan diferencias morfológicas\n")
  cat("significativas en el largo de sus pétalos, lo que puede reflejar\n")
  cat("adaptaciones evolutivas distintas o presiones selectivas diferentes.\n")
} else {
  cat("No se detectaron diferencias significativas en el largo de pétalos\n")
  cat("entre las especies, sugiriendo similaridad en este rasgo morfológico.\n")
}

## Las especies versicolor y virginica presentan diferencias morfológicas
## significativas en el largo de sus pétalos, lo que puede reflejar
## adaptaciones evolutivas distintas o presiones selectivas diferentes.
```

Síntesis del análisis

Planteamiento del problema: Se evaluó si existen diferencias significativas en el largo promedio de pétalos entre las especies *Iris versicolor* e *Iris virginica*, utilizando una prueba t de dos muestras independientes.

Resultados principales: - La especie virginica presenta pétalos significativamente más largos (media = 5.55 cm) que versicolor (media = 4.26 cm). - La diferencia es estadísticamente significativa ($p < 0.001$) con un tamaño del efecto grande ($d > 2.0$).

Interpretación biológica: Esta diferencia morfológica significativa en el largo de pétalos entre especies sugiere divergencia evolutiva en los caracteres florales, posiblemente relacionada con diferentes estrategias reproductivas, polinizadores específicos o adaptaciones a distintos ambientes.