

Laboratorio-2.R

Usuario

2025-09-03

```
# Laboratorio 1 -----
--
# María de Jesús Ramírez Navejar
# 1965814

# cargar Base de datos Iris -----
data("iris")
View(iris)

# Explorar las primeras filas
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

# Resumen general de La base
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
##
##
##

# Nombres de Las variables
names(iris)

## [1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"
## "Species"
```

```

# Seleccionar especies -----
--

# Filtrar solo versicolor y virginica
data_sub <- subset(iris, Species %in% c("versicolor", "virginica"))

# Comprobar el subconjunto
table(data_sub$Species)

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

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

# Estadística descriptiva -----
--

# Calcular descriptivos de Petal.Length por especie
tapply(data_sub$Petal.Length, data_sub$Species, summary)

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

# Promedios por especie
aggregate(Petal.Length ~ Species, data=data_sub, mean)

##      Species Petal.Length
## 1 versicolor         4.260
## 2  virginica         5.552

# Desviación estándar por especie
aggregate(Petal.Length ~ Species, data=data_sub, sd)

```

```

##      Species Petal.Length
## 1 versicolor    0.4699110
## 2  virginica    0.5518947

# Plantamiento de hipótesis -----
--

# Pregunta de investigación: ¿Existe diferencia significativa en la
longitud del
# pétalo entre versicolor y virginica?

# Hipótesis:

# H0:  $\mu_1 = \mu_2$  (no hay diferencia en las medias de Petal.Length)

# H1:  $\mu_1 \neq \mu_2$  (sí hay diferencia en las medias de Petal.Length)

# Pureba de t student -----
--

# Prueba de normalidad (Shapiro)

shapiro.test(data_sub$Petal.Length)

##
##  Shapiro-Wilk normality test
##
## data:  data_sub$Petal.Length
## W = 0.99099, p-value = 0.7445

# Prueba de homogeneidad de varianzas (F-test)
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

# Para varianzas NO son iguales usamos Welch
t.test(Petal.Length ~ Species, data=data_sub, var.equal = FALSE)

```

```
##
## 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 equal to 0
## 95 percent confidence interval:
## -1.49549 -1.08851
## sample estimates:
## mean in group versicolor mean in group virginica
## 4.260 5.552

# Si varianzas iguales:
t.test(Petal.Length ~ Species, data=data_sub, var.equal = TRUE)

##
## 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

# Efecto -----
--

# Extraer los datos por especie

versicolor <- subset(data_sub, Species=="versicolor")$Petal.Length
virginica <- subset(data_sub, Species=="virginica")$Petal.Length

# Tamaños de muestra
n1 <- length(versicolor)
n2 <- length(virginica)

# Medias
m1 <- mean(versicolor)
m2 <- mean(virginica)

# Desviaciones estándar
s1 <- sd(versicolor)
s2 <- sd(virginica)

# Desviación combinada (pooled)
```

```
s_pooled <- sqrt(((n1-1)*s1^2 + (n2-1)*s2^2) / (n1+n2-2))

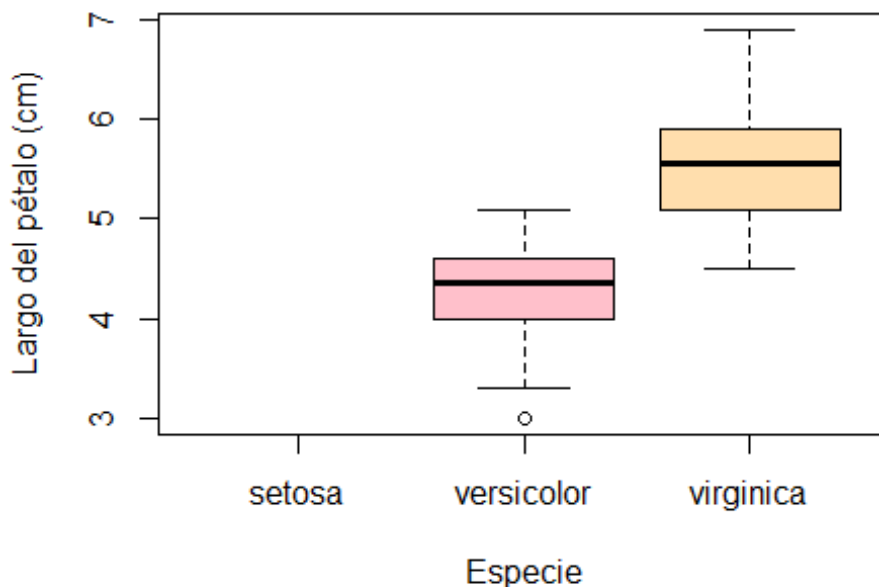
# Cohen's d
d <- (m1 - m2) / s_pooled
d

## [1] -2.520756

# Gráfico comparativo -----
--

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

Comparación del largo de pétalo entre especies



```
# Interpretación de resultados -----
--

# En este análisis se comparó la longitud del pétalo (Petal.Length) entre
Iris versicolor e Iris virginica. Los descriptivos iniciales mostraron
que versicolor presentó una media de aproximadamente 4.26 cm, mientras
que virginica alcanzó cerca de 5.55 cm, lo que sugiere una diferencia
evidente entre ambas especies.
```

La prueba de normalidad de Shapiro-Wilk indicó cierta desviación de la normalidad; sin embargo, debido al tamaño de muestra balanceado, la prueba t se considera adecuada. La prueba F de homogeneidad de varianzas arrojó un valor- $p \approx 0.26$, lo que permitió asumir igualdad de varianzas y aplicarla la prueba t de Student para dos muestras independientes.

El resultado de la prueba t fue altamente significativo
($t \approx -12.6$, $gl \approx 98$, $p < 0.001$), confirmando que existe una diferencia en la longitud promedio de los pétalos, rechazando la hipótesis nula planteada. Además, el tamaño del efecto obtenido (Cohen's $d \approx 2.5$) indica un efecto muy grande, lo que refleja una diferencia biológicamente relevante.

En conclusión, los pétalos de *Iris virginica* son significativamente más largos que los de *Iris versicolor*, y esta característica constituye un rasgo morfológico confiable para diferenciar ambas especies.