

Asignación-3.R

Usuario

2025-10-15

```
# Asignación 3: Contraste de medias/Base de datos Iris
# Semana 2 (04/09/2025)
# Andrea Michelle Luna Vasconcelos 1950889

# Base de datos Iris -----
--

data("iris")
View(iris)
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
##
##
##

knitr::kable(iris[1:10, ], caption = "Cuadro 1: Muestra de las
características
de las tres especies de iris")
```

Cuadro 1: Muestra de las características de las tres especies de iris

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa

Distribución de la longitud de pétalo por especie

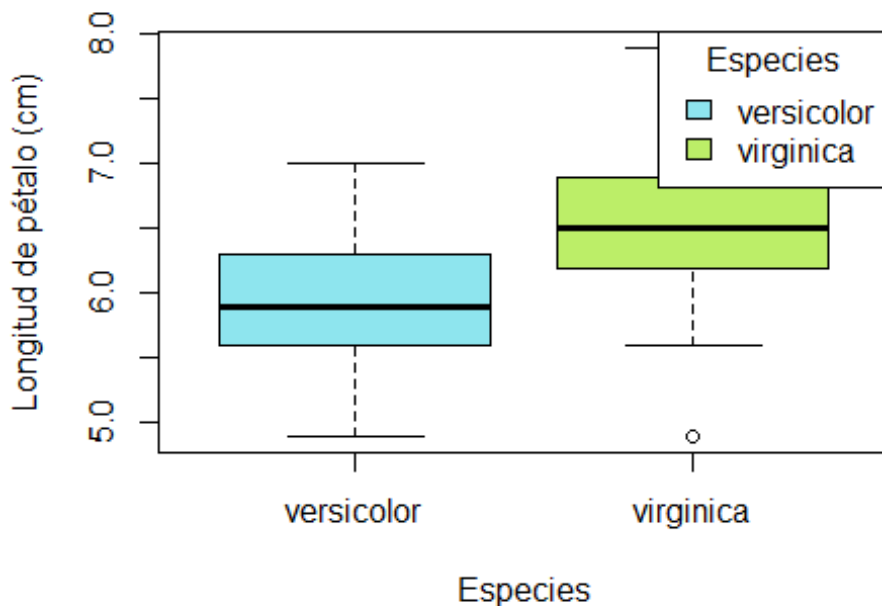
[illegible]

```

dat_iris$Species <- droplevels(dat_iris$Species) # Para quitar "setosa"
de la gráfica de la base de datos original.
boxplot(dat_iris$Sepal.Length ~ dat_iris$Species,
        xlab = "Especies",
        ylab = "Longitud de pétalo (cm)",
        main = "Distribución de la longitud de pétalo de dos especies",
        col = c("cadetblue2", "darkolivegreen2"))
legend("topright",
      legend = levels(dat_iris$Species),
      fill = c("cadetblue2", "darkolivegreen2"),
      title = "Especies")

```

Distribución de la longitud de pétalo de dos especi



Prueba de T -----
--

```
head(dat_iris)
```

```

##      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(dat_iris)
```

```
## Sepal.Length Sepal.Width Petal.Length Petal.Width
## Min. :4.900 Min. :2.000 Min. :3.000 Min. :1.000
## 1st Qu.:5.800 1st Qu.:2.700 1st Qu.:4.375 1st Qu.:1.300
## Median :6.300 Median :2.900 Median :4.900 Median :1.600
## Mean :6.262 Mean :2.872 Mean :4.906 Mean :1.676
## 3rd Qu.:6.700 3rd Qu.:3.025 3rd Qu.:5.525 3rd Qu.:2.000
## Max. :7.900 Max. :3.800 Max. :6.900 Max. :2.500
## Species
## versicolor:50
## virginica :50
##
##
##
##
```

Pregunta de investigación: ¿Varía La Longitud de pétalo entre La especie versicolor y virginica?

H0 = No existe diferencia significativa en La Longitud de pétalos entre versicolor y virginica.

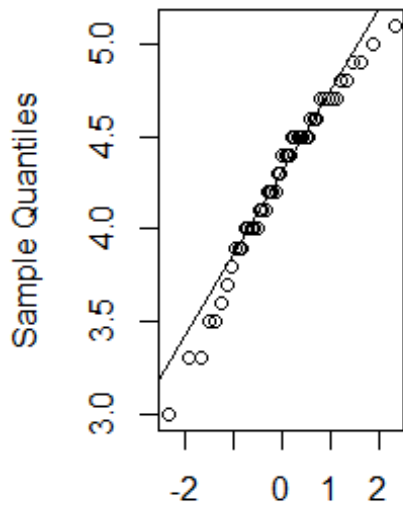
H1 = Sí existe una diferencia significativa en La Longitud de pétalos entre versicolor y virginica.

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

```
# QQ-plot -----
--
```

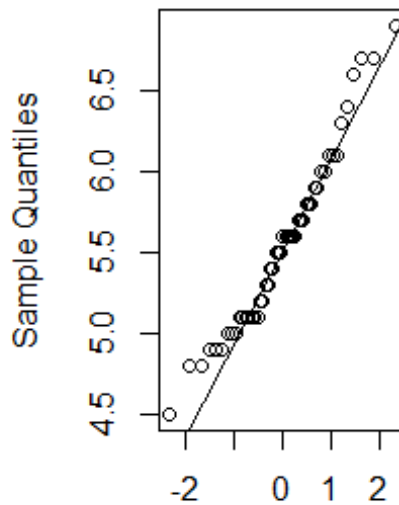
```
par(mfrow=c(1,2))
qqnorm(versicolor); qqline(versicolor)
qqnorm(virginica); qqline(virginica)
```

Normal Q-Q Plot



Theoretical Quantiles

Normal Q-Q Plot



Theoretical Quantiles

```
par(mfrow=c(1,2))
```

```
# Normalidad -----
```

```
--
```

```
# Normalidad ☒
```

```
shapiro.test(versicolor) # p-value = 0.1585
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: versicolor
```

```
## W = 0.966, p-value = 0.1585
```

```
shapiro.test(virginica) # p-value = 0.1098
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: virginica
```

```
## W = 0.96219, p-value = 0.1098
```

```
# Varianza -----
```

```
--
```

```
# Varianza ☒
```

```
var.test(versicolor, virginica) # df = 49, p-value = 0.2637
```

```
##  
## F test to compare two variances  
##  
## data: versicolor and virginica  
## 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
```

```
# Homogeneidad -----  
--
```

```
# Homogeneidad 
```

```
var.test(versicolor, virginica, alternative = "two.sided")
```

```
##  
## F test to compare two variances  
##  
## data: versicolor and virginica  
## 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 Student -----  
--
```

```
# Prueba t de Student
```

```
t.test(versicolor, virginica,  
       alternative = "two.sided",  
       var.equal = T)
```



```
##  
## Two Sample t-test  
##  
## data: versicolor and virginica  
## 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

# p-value < 2.2e-16 = 0.00000000000000022

# **Se rechaza la H0 y se acepta la H1.**

# Cohen's effect -----
--

# Cohen's effect  

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
}

d1_cal <- cohens_efecto(versicolor, virginica)
d1_cal

## [1] -2.520756

# Cohen's effect = -2.520756 Lo que nos dice que es un efecto muy grande.

# Visualización -----
--

library(ggplot2)
dat_sub <- subset(dat_iris, Species %in% c("versicolor", "virginica"))

ggplot(dat_sub, aes(x = Species, y = Petal.Length, fill = Species)) +
  geom_violin(trim = FALSE, color = "black", alpha = 0.6) +
  geom_boxplot(width = 0.1, fill = "white", outlier.shape = NA) +
  stat_summary(fun = mean, geom = "point", shape = 23, size = 3, fill =
"red") +
  theme_minimal() +
  labs(title = "Distribución de la longitud de pétalos por especie",
       x = "Especie",
       y = "Longitud de pétalo (cm)") +
  theme(legend.position = "none")
```

Distribución de la longitud de pétalos por especie

