

# HW\_2\_Iris.R

Ramon

2025-10-09

```
##HW_02

##30/8/2025
##Ramón Copado García
##Matrícula 1059439

##Objetivo
#El objetivo de esta práctica es que el estudiante se familiarice con el entorno
#de R y RStudio, explorando una de las bases de datos más utilizadas en
#estadística (iris), con el fin de:
# +Describir y comprender la estructura de un conjunto de datos reales.
# +Aplicar pruebas estadísticas básicas (prueba t de dos muestras) para
#   contrastar hipótesis sobre medias poblacionales.
# +Interpretar los resultados tanto en términos estadísticos (valores de p,
#   intervalos de confianza, tamaño del efecto) como en términos biológicos
#   (diferencias entre especies de iris).
# +Desarrollar habilidades prácticas en la escritura de código reproducible en
#   R y en la presentación de resultados mediante reportes en formato PDF.

##Base de datos Iris

#Importar datos de Github

url<-"https://gist.githubusercontent.com/netj/8836201/raw/6f9306ad21398ea43cba4f7d537619d0e07d5ae3/iris.csv"
url2<-paste0("https://gist.githubusercontent.com/netj/8836201/raw/",
             "6f9306ad21398ea43cba4f7d537619d0e07d5ae3/iris.csv")

iris<-read.csv(url,header=T)

iris<-read.csv(url2,header=T)

View (iris)

#Tambien se puede utilizar, data("iris"), y trabajar sobre la base de datos

# Ejercicio
#En la base iris, las especies versicolor y virginica suelen diferir en sus
#rasgos florales. Nos interesa evaluar si el largo del pétalo (Petal.Length)
#presenta diferencias en su media poblacional entre estas dos especies.
head(iris) #Primeras 6 filas
```

```
##      sepal.length sepal.width petal.length petal.width variety
## 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) #Resumen estadístico
```

```
##      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
##      variety
## Length:150
## Class :character
## Mode  :character
##
##
##
```

```
dim(iris) #Dimensiones de filas y columnas
```

```
## [1] 150   5
```

```
names(iris) #Revisar los nombre de las columnas
```

```
## [1] "sepal.length" "sepal.width" "petal.length" "petal.width" "variety"
```

```
str(iris) #Información sobre dimensiones, variables, el tipo de dato y valores
```

```
## '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 ...
## $ variety      : chr  "Setosa" "Setosa" "Setosa" "Setosa" ...
```

```
df<-iris[3] #data Frame de la variable a medir
```

```
by(iris[3],iris$variety, summary) #Resumen estadístico de la Variable a trabajar
```

```
## iris$variety: Setosa
##      petal.length
## Min.      :1.000
## 1st Qu.:1.400
```

```
## Median :1.500
## Mean   :1.462
## 3rd Qu.:1.575
## Max.   :1.900
## -----
## iris$variety: Versicolor
##   petal.length
##   Min.   :3.00
##   1st Qu.:4.00
##   Median :4.35
##   Mean   :4.26
##   3rd Qu.:4.60
##   Max.   :5.10
## -----
## iris$variety: Virginica
##   petal.length
##   Min.   :4.500
##   1st Qu.:5.100
##   Median :5.550
##   Mean   :5.552
##   3rd Qu.:5.875
##   Max.   :6.900
```

```
###Solo informativo y visualizar como sub ejercicio
tapply(iris$petal.length, iris$variety, mean)
```

```
##      Setosa Versicolor  Virginica
##      1.462      4.260      5.552
```

```
tapply(iris$petal.length, iris$variety, sd)
```

```
##      Setosa Versicolor  Virginica
## 0.1736640 0.4699110 0.5518947
```

```
tapply(iris$petal.length, iris$variety, var)
```

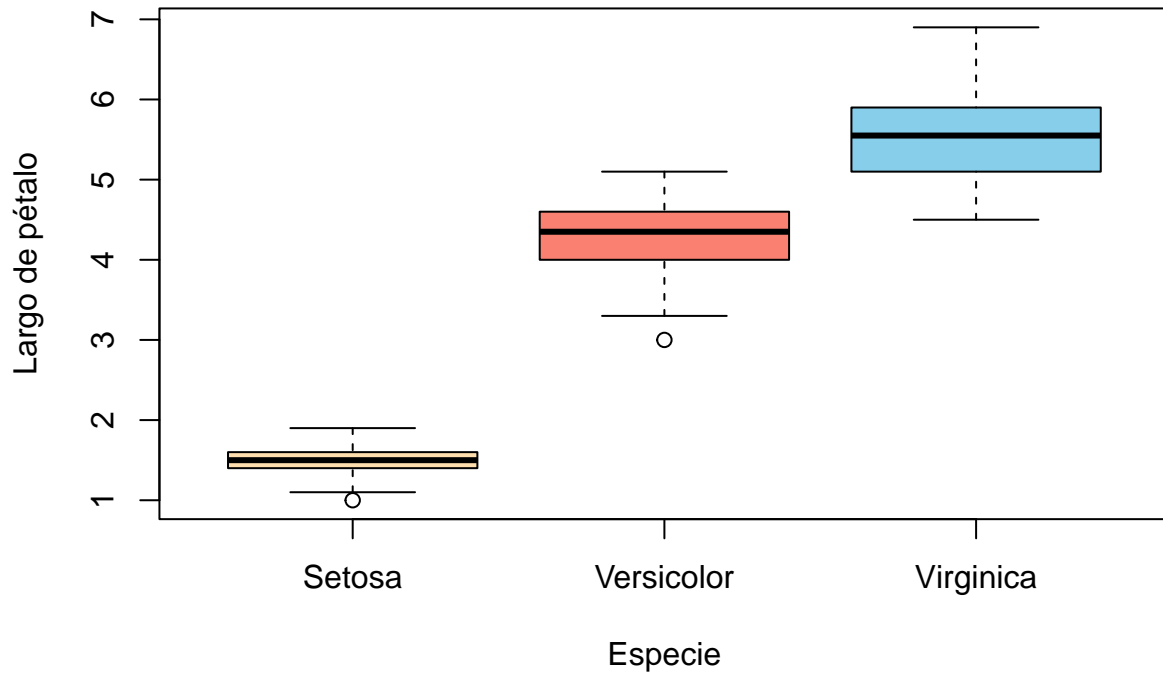
```
##      Setosa Versicolor  Virginica
## 0.03015918 0.22081633 0.30458776
```

```
colores <-c ("navajowhite", "salmon", "skyblue")
```

```
# Crear un boxplot iris
```

```
boxplot (iris$petal.length ~ iris$variety, col = colores,
        main = "Distribución del largo de pétalo por especie",
        xlab = "Especie",
        ylab = "Largo de pétalo")
```

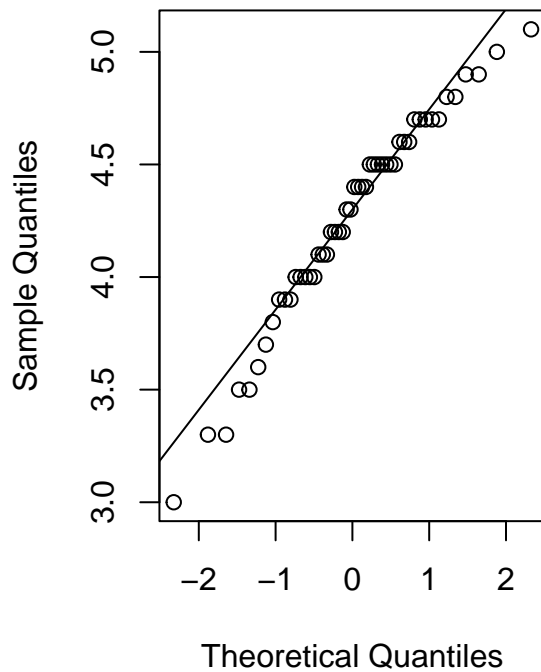
## Distribución del largo de pétalo por especie



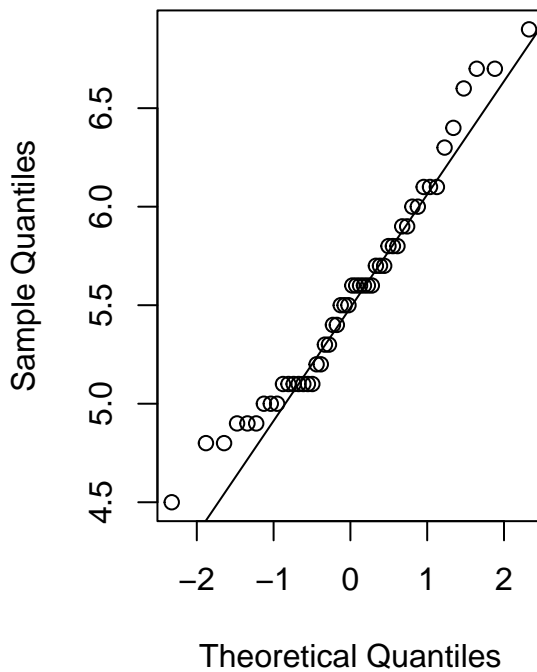
```
df_Versicolor <- subset(iris, variety == "Versicolor")
df_Virginica <- subset(iris, variety == "Virginica")

par(mfrow=c(1,2))
qqnorm(df_Versicolor$petal.length); qqline(df_Versicolor$petal.length)
qqnorm(df_Virginica$petal.length); qqline(df_Virginica$petal.length)
```

Normal Q-Q Plot



Normal Q-Q Plot



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

*#Lo anterior es solo para visualizar las 150 muestras.*

**##Ejercicio**

*#Datos a trabajar*

*#A partir de la base de datos iris disponible en R, realice lo siguiente:*

*#+Selección de especies: elija las especies versicolor y virginica de la base  
#y enfoque su análisis en la variable Petal.Length.*

```
data_sub <-subset(iris, variety %in% c("Versicolor","Virginica"))
table(data_sub$variety)
```

```
##
```

```
## Versicolor  Virginica
```

```
##          50          50
```

```
variety<-("Versicolor, Virginica")
```

*#Vamos a introducir el operador%in% para realizar el subset. En R, el*

*#operador%in% se utiliza para preguntar si un valor pertenece a un conjunto  
# de valores. Devuelve un vector lógico (TRUE o FALSE) indicando si cada  
#elemento de la izquierda está contenido dentro del vector de la derecha.*

```
data_sub <-subset(iris, variety %in% c("Versicolor","Virginica"))
table(data_sub$variety)
```

```
##
## Versicolor  Virginica
##           50          50
```

```
variety<-("Versicolor,Virginica")
```

```
data_sub <-subset(iris, variety %in% c("Versicolor","Virginica"))
table(data_sub$variety)
```

```
##
## Versicolor  Virginica
##           50          50
```

```
data_sub
```

```
##      sepal.length sepal.width petal.length petal.width  variety
## 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
## 57           6.3         3.3         4.7         1.6 Versicolor
## 58           4.9         2.4         3.3         1.0 Versicolor
## 59           6.6         2.9         4.6         1.3 Versicolor
## 60           5.2         2.7         3.9         1.4 Versicolor
## 61           5.0         2.0         3.5         1.0 Versicolor
## 62           5.9         3.0         4.2         1.5 Versicolor
## 63           6.0         2.2         4.0         1.0 Versicolor
## 64           6.1         2.9         4.7         1.4 Versicolor
## 65           5.6         2.9         3.6         1.3 Versicolor
## 66           6.7         3.1         4.4         1.4 Versicolor
## 67           5.6         3.0         4.5         1.5 Versicolor
## 68           5.8         2.7         4.1         1.0 Versicolor
## 69           6.2         2.2         4.5         1.5 Versicolor
## 70           5.6         2.5         3.9         1.1 Versicolor
## 71           5.9         3.2         4.8         1.8 Versicolor
## 72           6.1         2.8         4.0         1.3 Versicolor
## 73           6.3         2.5         4.9         1.5 Versicolor
## 74           6.1         2.8         4.7         1.2 Versicolor
## 75           6.4         2.9         4.3         1.3 Versicolor
## 76           6.6         3.0         4.4         1.4 Versicolor
## 77           6.8         2.8         4.8         1.4 Versicolor
## 78           6.7         3.0         5.0         1.7 Versicolor
## 79           6.0         2.9         4.5         1.5 Versicolor
```

## 80	5.7	2.6	3.5	1.0 Versicolor
## 81	5.5	2.4	3.8	1.1 Versicolor
## 82	5.5	2.4	3.7	1.0 Versicolor
## 83	5.8	2.7	3.9	1.2 Versicolor
## 84	6.0	2.7	5.1	1.6 Versicolor
## 85	5.4	3.0	4.5	1.5 Versicolor
## 86	6.0	3.4	4.5	1.6 Versicolor
## 87	6.7	3.1	4.7	1.5 Versicolor
## 88	6.3	2.3	4.4	1.3 Versicolor
## 89	5.6	3.0	4.1	1.3 Versicolor
## 90	5.5	2.5	4.0	1.3 Versicolor
## 91	5.5	2.6	4.4	1.2 Versicolor
## 92	6.1	3.0	4.6	1.4 Versicolor
## 93	5.8	2.6	4.0	1.2 Versicolor
## 94	5.0	2.3	3.3	1.0 Versicolor
## 95	5.6	2.7	4.2	1.3 Versicolor
## 96	5.7	3.0	4.2	1.2 Versicolor
## 97	5.7	2.9	4.2	1.3 Versicolor
## 98	6.2	2.9	4.3	1.3 Versicolor
## 99	5.1	2.5	3.0	1.1 Versicolor
## 100	5.7	2.8	4.1	1.3 Versicolor
## 101	6.3	3.3	6.0	2.5 Virginica
## 102	5.8	2.7	5.1	1.9 Virginica
## 103	7.1	3.0	5.9	2.1 Virginica
## 104	6.3	2.9	5.6	1.8 Virginica
## 105	6.5	3.0	5.8	2.2 Virginica
## 106	7.6	3.0	6.6	2.1 Virginica
## 107	4.9	2.5	4.5	1.7 Virginica
## 108	7.3	2.9	6.3	1.8 Virginica
## 109	6.7	2.5	5.8	1.8 Virginica
## 110	7.2	3.6	6.1	2.5 Virginica
## 111	6.5	3.2	5.1	2.0 Virginica
## 112	6.4	2.7	5.3	1.9 Virginica
## 113	6.8	3.0	5.5	2.1 Virginica
## 114	5.7	2.5	5.0	2.0 Virginica
## 115	5.8	2.8	5.1	2.4 Virginica
## 116	6.4	3.2	5.3	2.3 Virginica
## 117	6.5	3.0	5.5	1.8 Virginica
## 118	7.7	3.8	6.7	2.2 Virginica
## 119	7.7	2.6	6.9	2.3 Virginica
## 120	6.0	2.2	5.0	1.5 Virginica
## 121	6.9	3.2	5.7	2.3 Virginica
## 122	5.6	2.8	4.9	2.0 Virginica
## 123	7.7	2.8	6.7	2.0 Virginica
## 124	6.3	2.7	4.9	1.8 Virginica
## 125	6.7	3.3	5.7	2.1 Virginica
## 126	7.2	3.2	6.0	1.8 Virginica
## 127	6.2	2.8	4.8	1.8 Virginica
## 128	6.1	3.0	4.9	1.8 Virginica
## 129	6.4	2.8	5.6	2.1 Virginica
## 130	7.2	3.0	5.8	1.6 Virginica
## 131	7.4	2.8	6.1	1.9 Virginica
## 132	7.9	3.8	6.4	2.0 Virginica
## 133	6.4	2.8	5.6	2.2 Virginica

```
## 134      6.3      2.8      5.1      1.5 Virginica
## 135      6.1      2.6      5.6      1.4 Virginica
## 136      7.7      3.0      6.1      2.3 Virginica
## 137      6.3      3.4      5.6      2.4 Virginica
## 138      6.4      3.1      5.5      1.8 Virginica
## 139      6.0      3.0      4.8      1.8 Virginica
## 140      6.9      3.1      5.4      2.1 Virginica
## 141      6.7      3.1      5.6      2.4 Virginica
## 142      6.9      3.1      5.1      2.3 Virginica
## 143      5.8      2.7      5.1      1.9 Virginica
## 144      6.8      3.2      5.9      2.3 Virginica
## 145      6.7      3.3      5.7      2.5 Virginica
## 146      6.7      3.0      5.2      2.3 Virginica
## 147      6.3      2.5      5.0      1.9 Virginica
## 148      6.5      3.0      5.2      2.0 Virginica
## 149      6.2      3.4      5.4      2.3 Virginica
## 150      5.9      3.0      5.1      1.8 Virginica
```

*#Instrucción de tarea*

*#Primer contacto con R*

*#Explorar la base de datos iris usando funciones como head(), Summary()*  
`head(data_sub)`

```
##      sepal.length sepal.width petal.length petal.width  variety
## 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)`

```
##      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
##      variety
## Length:100
## Class :character
## Mode  :character
##
##
##
```

*#Identificar las variables Petal.Length y determina las estadísticas descriptivas  
#para las dos especie*

`tapply(data_sub$petal.length, data_sub$variety, summary)` *#Resumen estadístico*



```
## $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
```

```
tapply(data_sub$petal.length, data_sub$variety, mean) #Solo práctica
```

```
## Versicolor  Virginica
##      4.260      5.552
```

```
tapply(data_sub$petal.length, data_sub$variety, sd) #Solo práctica
```

```
## Versicolor  Virginica
##  0.4699110  0.5518947
```

```
tapply(data_sub$petal.length, data_sub$variety, var) #Solo práctica
```

```
## Versicolor  Virginica
##  0.2208163  0.3045878
```

```
by(data_sub[3], data_sub$variety,summary)
```

```
## data_sub$variety: Versicolor
##   petal.length
##   Min.   :3.00
##   1st Qu.:4.00
##   Median :4.35
##   Mean   :4.26
##   3rd Qu.:4.60
##   Max.   :5.10
## -----
## data_sub$variety: Virginica
##   petal.length
##   Min.   :4.500
##   1st Qu.:5.100
##   Median :5.550
##   Mean   :5.552
##   3rd Qu.:5.875
##   Max.   :6.900
```

*#Prueba estadística*

*#+Defina una pregunta de investigación sobre la variable Petal.Length*

*#¿Hay diferencia significativa en la longitud de los pétalos (Petal.Length)  
# entre las variedades Versicolor y Virginica de la base de datos iris?*

*#+Plantee formalmente las hipótesis estadísticas para una prueba t de dos*

```

#muestras independientes (two.sided).

# + H0 (nula): No existen diferencias significativas entre la longitud de
#los petalos de las variedades Versicolor y Virginica de la base de datos iris.

# + H1 (alternativa): Existen diferencias significativas entre la longitud
#de los petalos de las variedades Versicolor y Virginica de la base de datos iris.

#Ejecute la prueba en R justificando el tipo de prueba (Welch cuando las
#varianzas son diferentes o clásica, cuando las varianzas son iguales).

# Revisar homogeneidad
var.test(data_sub$petal.length ~ data_sub$variety)

```

```

##
## F test to compare two variances
##
## data: data_sub$petal.length by data_sub$variety
## 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

```

```

# Observar datos
# Al utilizar F test para comparar dos varianzas, la información que nos arroja
#son valores de P = 0.2637 siendo >0.05, estos datos nos dicen que no existen
#diferencias significativas entre las varianzas de las dos especies; por lo
#tanto si hay homogeneidad y se utilizará la prueba de T clásica.

# Prueba de T
t.test(data_sub$petal.length ~ data_sub$variety, alternative = "two.sided",
var.equal = T)

```

```

##
## Two Sample t-test
##
## data: data_sub$petal.length by data_sub$variety
## 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

```

```

t.test(data_sub$petal.length ~ data_sub$variety, alternative = "two.sided",
var.equal = F)

```

```

##

```

```
## Welch Two Sample t-test
##
## data: data_sub$petal.length by data_sub$variety
## 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
```

```
#Con esta prueba de T podemos decir que el valor de p (p=<2.2e-16) es < que
#0.05, rechazamos la hipótesis nula y decimos que si hay diferencia
#significativa entre las variedades Virginica y Versicolor en la variable
#petal.length de la base de datos iris.
```

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

```
# Medir el efecto del efecto
```

```
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(df_Versicolor$petal.length, df_Virginica$petal.length)
d1_cal
```

```
## [1] -2.520756
```

```
abs(d1_cal)
```

```
## [1] 2.520756
```

```
#Este valor de cohens no dice que hay un diferencia enorme en la
#variable petal.length y esto nos lleva al inicio en objetivos que
#podemos decir que tanto en términos estadísticos como en biológicos
#si hay diferencia entre las variedades y el largo del petalo.
```

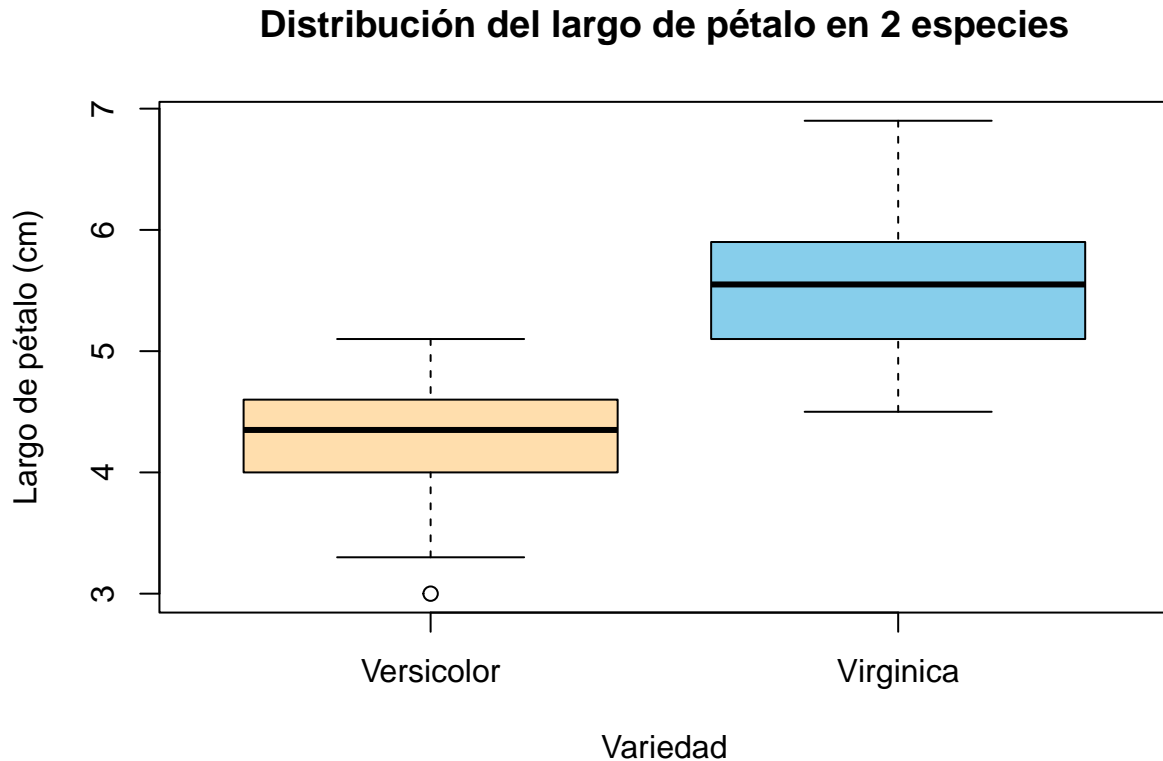
```
#Visualización
```

```
# Genere una gráfica comparativa (boxplot, violinplot, etc.) que muestre
#las diferencias entre especies.
```

```
colores <-c ("navajowhite", "skyblue")
```

```
# Crear un boxplot data_sub
```

```
boxplot (data_sub$petal.length~ data_sub$variety, col = colores,
        main = "Distribución del largo de pétalo en 2 especies",
        xlab = "Variedad",
        ylab = "Largo de pétalo (cm)")
```



*#Se realizó una prueba t para muestras independientes ( Versicolor vs Virginica),  
 #comprobando varianzas iguales. Se encontró una diferencia,  $t(98) = -12.604$ ,  
 # $p \leq 2.2e-16$ . El grupo Virginica mostró una media mayor (5.552) que el grupo  
 #Versicolor (4.26). La diferencia de medias fue de 1.292 y el IC 95% =  
 #[-0.23, -0.04]. El tamaño del efecto fue grande ( $d=-2.520756$ ) lo que indica  
 #que la variedad tuvo un efecto sustancial sobre el largo del petalo.*

# Informe Escrito -----

#Informe escrito:

#Redacte una síntesis (máx. 1 cuartilla) que incluya:

#• Planteamiento del problema y de las hipótesis.

#• Resultados numéricos y gráficos.

#• Interpretación estadística y biológica.

#• Planteamiento del problema y de las hipótesis.

#De la base de datos iris saber si hay diferencia del el largo de petalo entre

#las variedades Versicolor y Virginica. De aquí la pregunta que me realice fue:

#¿Hay diferencia significativa en la longitud de los pétalos (Petal.Length)

#entre las variedades Versicolor y Virginica de la base de datos iris?

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

#muestras independientes (two.sided).

# + H0 (nula): No existen diferencias significativas entre la longitud de

#los petalos de las variedades Versicolor y Virginica de la base de datos iris.

# + H1 (alternativa): Existen diferencias significativas entre la longitud

#de los petalos de las variedades Versicolor y Virginica de la base de datos iris.

#• Resultados numéricos y gráficos.

#Los resultados numéricos fueron:

#Se realizó una prueba t para muestras independientes ( Versicolor vs Virginica),

#comprobando varianzas iguales. Se encontró una diferencia,  $t(98) = -12.604$ ,

# $p = 2.2e-16$ . El grupo Virginica mostró una media mayor (5.552) que el grupo

#Versicolor (4.26). La diferencia de medias fue de 1.292 y el IC 95% =

# $[-0.23, -0.04]$ . El tamaño del efecto fue grande ( $d = -2.520756$ ) lo que indica

#que la variedad tuvo un efecto sustancial sobre el largo del petalo.

#Los resultados gráficos mostraron

# Las gráficas tanto la de inicio (practica) como la de resultados muestran

# claramente que si hay una muy significativa diferencia entre las variedades cuando

# las evaluamos por el largo de petalo, aunque se puede observar que pudiera haber

# algo de valores iguales en tanto que pudieramos decir que la variedades versicolor

# y virginica puede haber erros en cuanto a la clasificación de variedades ya que

# los valores máximos de versicolor pueden confundir con los valores mínimos de

# virginica y es por eso que se uso la prueba de T clásica para estas dos variedades

#• Interpretación estadística y biológica.

#Estadísticamente determinamos que si hay diferencia enorme con el valor del largo

#de petalo entre as variedades versicolor y virginica ya que el valor del efecto

#del efecto evaluado por el método Cohen's no dio un valor por arriba del criterio

#de efecto grande.

#Hablando biologicamente el largo del petalo de la especie iris es fundamental para

#determinar a la variedad que corresponde y estadísticamente esta respaldado.