

Asignación 3: Contraste de medias

Dr. Marco Aurelio González Tagle

Alumno: César Enrique Martínez Sánchez

Objetivos

- 1.- Describir y comprender la estructura de un conjunto de datos reales.
- 2.- Aplicar pruebas estadísticas básicas (prueba t de dos muestras) para contrastar hipótesis sobre medias poblacionales.
- 3.- 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).
- 4.- Desarrollar habilidades prácticas en la escritura de código reproducible en R y en la presentación de resultados mediante reportes en formato PDF.

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.

Instrucción de tarea

Primer contacto con R:

- Explorar la base de datos de iris usando funciones como head(), summary().

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

```
head(data_sub)
```

```
> 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
> |
```

8576901 (Identidad confirmada).

```
summary(data_sub)
```

```
> summary(data_sub)
  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width      Species
Min.   :4.900   Min.   :2.000   Min.   :3.000   Min.   :1.000   setosa    : 0
1st Qu.:5.800   1st Qu.:2.700   1st Qu.:4.375   1st Qu.:1.300   versicolor:50
Median :6.300   Median :2.900   Median :4.900   Median :1.600   virginica :50
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
> |
```

- Identificar las variables Petal.Length y determina las estadísticas descriptivas de las dos especies.

```
petal_length <- subset(
  iris,
  Species %in% c("versicolor","virginica"),
  select = c(Species, Petal.Length)
)
```

Se instala el paquete “**dplyr**” para el cálculo de las estadísticas descriptivas (media, mediana, máximo, mínimo, desviación estándar, cuartiles, rango intercuartil:

```
install.packages("dplyr")
data_sub <- subset(iris, Species %in% c("versicolor","virginica"))
petal_length <- subset(data_sub, select = c(Species, Petal.Length))
```

```
library(dplyr)
petal_length %>%
  group_by(Species) %>%
  summarise(
    n = n(),
    mean = mean(Petal.Length),
    median = median(Petal.Length),
    sd = sd(Petal.Length),
    min = min(Petal.Length),
    q1 = quantile(Petal.Length, .25),
    q3 = quantile(Petal.Length, .75),
    max = max(Petal.Length),
    IQR = IQR(Petal.Length),
    .groups = "drop"
  )
```

```
# A tibble: 2 × 10
  Species      n mean median   sd  min   q1   q3  max  IQR
  <fct>    <int> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 versicolor   50  4.26   4.35 0.470   3     4   4.6   5.1 0.600
2 virginica    50  5.55   5.55 0.552   4.5   5.1  5.88   6.9 0.775
> )
```

576901 (Identidad confirmada).

Planteamiento de problema

¿La media de la longitud de pétalo (Petal.Length) difiere entre ambas especies?

H₀ (nula): No existe diferencia en la media de la longitud del pétalo entre **versicolor** y **virginica**.

- **H₁ (alternativa):** Existe una diferencia en la media de la longitud del pétalo entre **versicolor** y **virginica**.

Se realizó la prueba **F-test** para comparar diferencias significativas entre las varianzas de la longitud de pétalo entre ambas especies y seleccionar **Welch** o **t de Student**.

La prueba **F** para homogeneidad de varianzas no fue significativa ($F=0.725$, $p=0.264$; IC95% del cociente de varianzas [0.411, 1.278]), por lo que se asumió igualdad de varianzas y se aplicó una prueba **t de Student** de dos muestras ($\alpha=0.05$).

Se observó una diferencia significativa de medias de Petal.Length entre **virginica** y **versicolor** ($t(98)=-12.60$, $p<0.001$), con una diferencia media de 1.29 (IC95% 1.09–1.50) a favor de **virginica**. El efecto de tamaño de Cohen (Cohen's d) es una medida estandarizada de la diferencia entre dos medias e indica cuántas desviaciones estándar separan las medias de dos grupos. En este caso se observó un tamaño de efecto muy grande (Cohen's d ≈ 2.5).

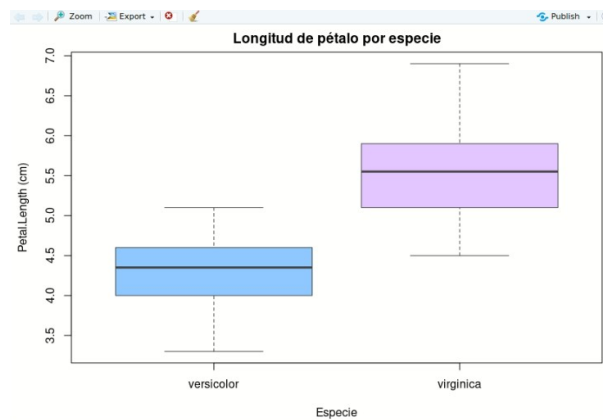


Figura 1. Longitud del pétalo (cm) por especie en el conjunto iris (boxplot).

La diferencia en la longitud del pétalo es **estadística y biológicamente significativa** y de **gran magnitud**. Morfológicamente, las muestras de **virginica** presenta pétalos notablemente más largos que las de **versicolor**, con implicaciones claras para **taxonomía y clasificación**.