# Solución del problema 8

Los datos del fichero lirios.RData (https://matematicas.uam.es/~joser.berrendero/datos/lirios.RData) corresponden a la longitud y anchura del pétalo y del sépalo de 100 lirios, 50 de ellos correspondientes a la especie *versicolor* y otros 50 de la especie *virginica*.

- Considera primero únicamente las dos variables correspondientes al sépalo. Calcula los coeficientes de la función discriminante lineal de Fisher y estima la probabilidad de error de esta regla mediante el riesgo empírico $\hat{L}_n$ y la tasa de error por validación cruzada $\hat{L}_n^{\mathrm{VC}}$. Compara los valores de estos estimadores con el estimador *paramétrico* basado en el resultado del problema anterior: $1-\Phi(\hat{\Delta}^2/2)$, donde $\hat{\Delta}^2=(\hat{\mu}_0-\hat{\mu}_1)'\hat{\Sigma}^{-1}(\hat{\mu}_0-\hat{\mu}_1)]$.
- Repite el apartado anterior pero considerando las cuatro variables.

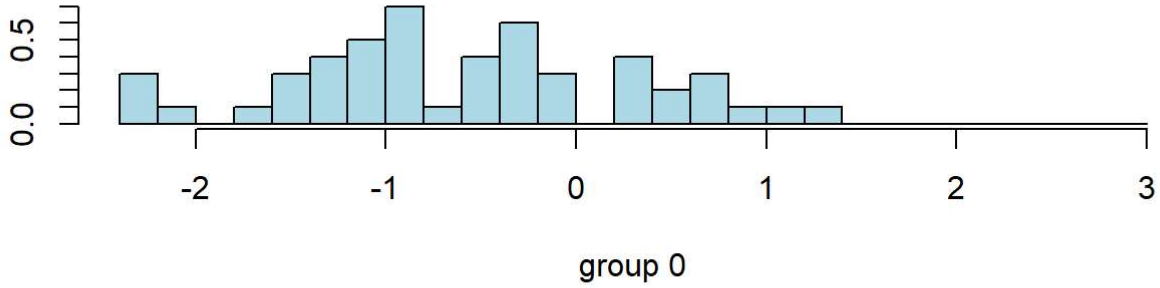
# Carga librerías y prepara los datos

```
library(tidyverse)
library(MASS)

datos <- 'https://matematicas.uam.es/~joser.berrendero/datos/lirios.RData'
load(url(datos))
lirios <- data.frame(cbind(lirios, especie = clases))
glimpse(lirios)
```

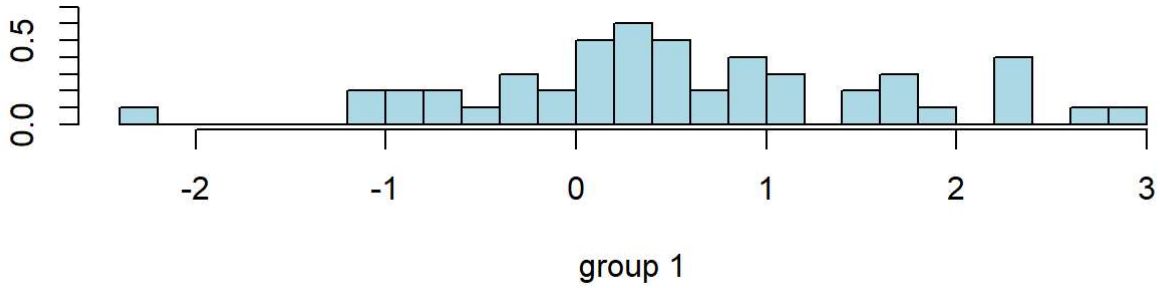
```
## Rows: 100
## Columns: 5
## $ Sepal.Length <dbl> 7.0, 6.4, 6.9, 5.5, 6.5, 5.7, 6.3, 4.9, 6.6, 5.2, 5.0, 5.~
## $ Sepal.Width  <dbl> 3.2, 3.2, 3.1, 2.3, 2.8, 2.8, 3.3, 2.4, 2.9, 2.7, 2.0, 3.~
## $ Petal.Length <dbl> 4.7, 4.5, 4.9, 4.0, 4.6, 4.5, 4.7, 3.3, 4.6, 3.9, 3.5, 4.~
## $ Petal.Width  <dbl> 1.4, 1.5, 1.5, 1.3, 1.5, 1.3, 1.6, 1.0, 1.3, 1.4, 1.0, 1.~
## $ especie      <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
```

# Tasas de error

```
lda_lirios <- lda(especie ~ Sepal.Length + Sepal.Width, data = lirios, prior = c(0.5, 0.5))

# Puntuaciones discriminantes
plot(lda_lirios, col ='lightblue')
```




group 0



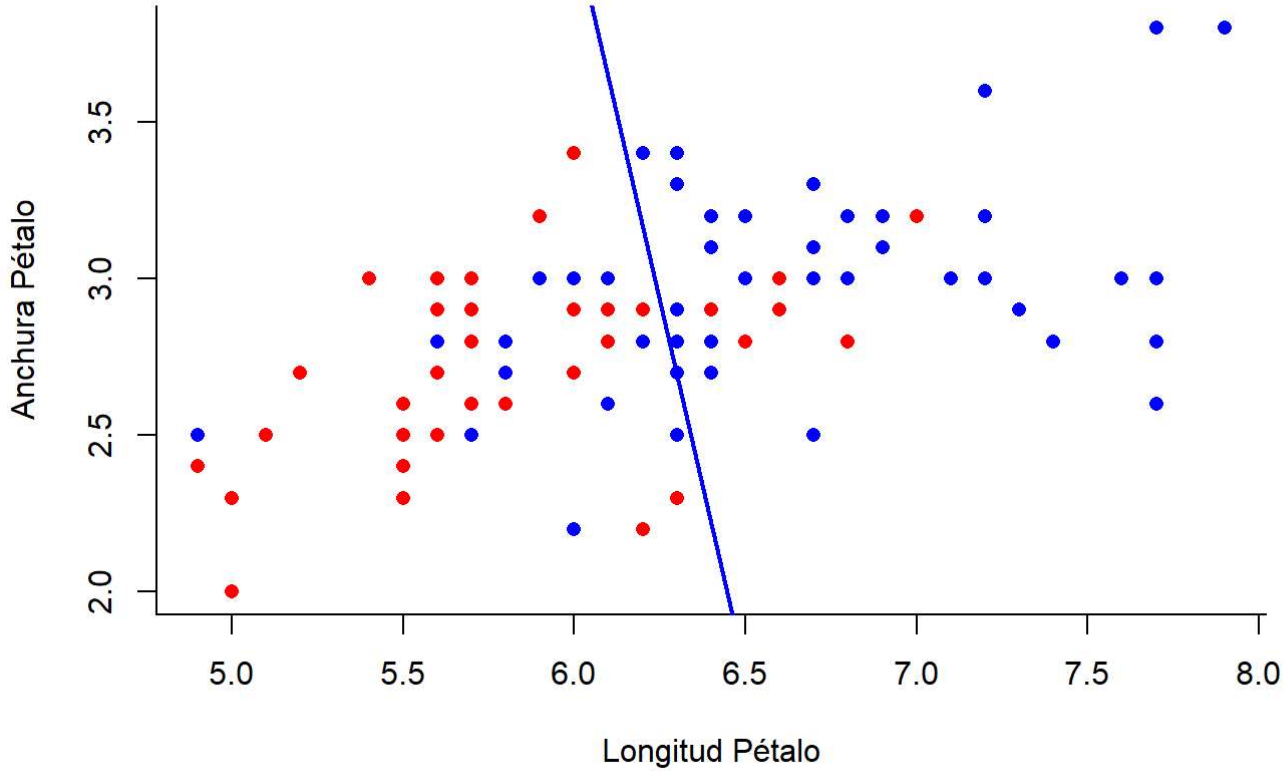

group 1

```
# Error aparente
especie_pred <- predict(lda_lirios)$class
mean(lirios$especie != especie_pred)
```

```
## [1] 0.25
```

```
# Prob. de error estimada por validacion cruzada
especie_pred <- lda(especie ~ Sepal.Length + Sepal.Width, data = lirios, prior = c(0.5, 0.5),
CV =TRUE)$class
mean(lirios$especie != especie_pred)
```

```
## [1] 0.27
```




# Usando expresión teórica

```
# Medias y covarianzas estimadas por grupos
mat1 <- lirios[1:50,1:2]
mat2 <- lirios[51:100,1:2]

mu1 <- colMeans(mat1)
mu2 <- colMeans(mat2)

S1 <- cov(mat1)
S2 <- cov(mat2)


# estimador combinado
n1 = dim(mat1)[1]
n2 = dim(mat2)[1]
Sp = ((n1-1)*S1 + (n2-1)*S2 )/(n1+n2-2)

# Dist Mahalanobis (al cuadrado) entre las medias
distancia = (mu1-mu2) %*% solve(Sp) %*% t(t(mu1-mu2))

# Error Bayes según fórmula
1-pnorm(sqrt(distancia)/2)
```

```
##           [,1]
## [1,] 0.2858654
```