

Inferencia estadística multivariante

Soluciones

Francesc Carmona y Josep Gregori*

12 de diciembre de 2023

Ejercicio 1

En un estudio sobre las plagas de diferentes artrópodos en las encinas de las dehesas, la siguiente tabla presenta la distribución conjunta de las variables aleatorias discretas: x_1 : diferentes tipos de artrópodos, que toma seis valores posibles $P_1 = \text{Orugas}$, $P_2 = \text{Hemipteros}$, $P_3 = \text{Hormigas}$, $P_4 = \text{Arañas}$, $P_5 = \text{Coleopteros}$ y $P_6 = \text{Otros}$ y x_2 : tipo de dehesa, que toma los valores $A = \text{dehesa pastada}$, $B = \text{dehesa con matorral}$ y $C = \text{dehesa cultivada}$.

	A	B	C
P_1	0.130	0.110	0.090
P_2	0.105	0.065	0.080
P_3	0.070	0.035	0.045
P_4	0.080	0.040	0.080
P_5	0.010	0.020	0.020
P_6	0.005	0.010	0.005

a) Calcular las distribuciones marginales¹.

```
> dae <- array(c(0.13,0.105,0.07,0.08,0.01,0.005,
+               0.11,0.065,0.035,0.04,0.02,0.01,
+               0.09,0.08,0.045,0.08,0.02,0.005), dim=c(6,3),
+               dimnames=list(paste("P",1:6, sep=""),c(LETTERS[1:3])))
> tbl <- addmargins(dae); tbl
```

	A	B	C	Sum
P1	0.130	0.110	0.090	0.33
P2	0.105	0.065	0.080	0.25
P3	0.070	0.035	0.045	0.15
P4	0.080	0.040	0.080	0.20
P5	0.010	0.020	0.020	0.05
P6	0.005	0.010	0.005	0.02
Sum	0.400	0.280	0.320	1.00

De modo que las probabilidades marginales para el tipo de artrópodo son

```
> (P.artropodo <- tbl[1:6,4])
```

P1	P2	P3	P4	P5	P6
0.33	0.25	0.15	0.20	0.05	0.02

* Alumno del curso 2009-10

¹En el caso discreto, las integrales son sumas.

Las probabilidades marginales para el tipo de dehesa son

```
> (P.dehesa <- tbl[7,1:3])
```

```
      A      B      C
0.40 0.28 0.32
```

- b) Calcular la distribución condicionada de las diferentes clases de artrópodos en la dehesa cultivada².

Las probabilidades condicionadas de los tipos de artrópodos en la dehesa cultivada (C) son

```
> tbl[1:6,"C"]/P.dehesa["C"]
```

```
      P1      P2      P3      P4      P5      P6
0.281250 0.250000 0.140625 0.250000 0.062500 0.015625
```

- c) Calcular la distribución condicionada de las dehesas para las arañas³.

Las probabilidades condicionadas de las dehesas para las arañas (P4) son

```
> tbl["P4",1:3]/P.artropodo["P4"]
```

```
      A      B      C
0.4 0.2 0.4
```

Ejercicio 2 (**)

Dada la función de densidad conjunta $f(x, y) = 6x$ definida en la región $R = \{(x, y) / 0 < x < 1, 0 < y < 1 - x\}$,

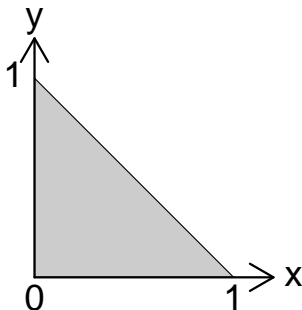
- a) Comprobar que las densidades marginales de ambas variables son $f(x) = 6x(1 - x)$, para $0 < x < 1$ y $f(y) = 3(1 - y)^2$, para $0 < y < 1$.

La región R es

```
> plot(c(0,0),c(1,1), type = "n", axes=FALSE,
+       xlab="", ylab="", asp=1, xlim=c(-0.2,1.4), ylim=c(-0.2,1.4))
> polygon(c(0,1,0), c(0,0,1), col=gray(0.8))
> arrows(x0=c(0,0), y0=c(0,0), x1=c(0,1.2), y1=c(1.2,0), lwd=2)
> text(x=1.2, y=0, labels = "x", cex=2, pos=4)
> text(x=0, y=1.2, labels = "y", cex=2, pos=3)
> text(x=1, y=0, labels = "1", cex=2, pos=1)
> text(x=0, y=0, labels = "0", cex=2, pos=1)
> text(x=0, y=1, labels = "1", cex=2, pos=2)
```

²La distribución condicionada de \mathbf{x}_1 , para un valor concreto de $\mathbf{x}_2 = \mathbf{x}_2^0$ es $f(\mathbf{x}_1|\mathbf{x}_2^0) = f(\mathbf{x}_1, \mathbf{x}_2^0)/f(\mathbf{x}_2^0)$.

³La distribución condicionada de \mathbf{x}_2 , para un valor concreto de $\mathbf{x}_1 = \mathbf{x}_1^0$ es $f(\mathbf{x}_2|\mathbf{x}_1^0) = f(\mathbf{x}_1^0, \mathbf{x}_2)/f(\mathbf{x}_1^0)$.



$$f(x) = \int_{R_x} f(x, y) dy = \int_0^{1-x} 6x dy = 6x[y]_0^{1-x} = 6x(1-x) \quad \text{para } 0 < x < 1$$

$$f(y) = \int_{R_y} f(x, y) dx = \int_0^{1-y} 6x dx = [3x^2]_0^{1-y} = 3(1-y)^2 \quad \text{para } 0 < y < 1$$

- b) Comprobar que las densidades condicionadas son $f(y|x) = \frac{1}{1-x}$, para $0 < y < 1-x$ y $f(x|y) = \frac{2x}{(1-y)^2}$, para $0 < x < 1-y$.

$$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{6x}{6x(1-x)} = \frac{1}{1-x} \quad \text{para } 0 < x < 1, 0 < y < 1-x$$

$$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{6x}{3(1-y)^2} = \frac{2x}{(1-y)^2} \quad \text{para } 0 < y < 1, 0 < x < 1-y$$

Ejercicio 3

Consideremos un vector aleatorio $\mathbf{x} = (x_1, x_2)$ con distribución uniforme en el rectángulo $[0, 2] \times [3, 4]$.

- a) Especificar la función de densidad conjunta de \mathbf{x} y calcular $E(\mathbf{x})$ y $\text{var}(\mathbf{x})$.

La función de densidad conjunta es

$$f(x, y) = \frac{1}{\text{área}} = \frac{1}{2 \times 1} = \frac{1}{2}$$

Las funciones de densidad marginales son uniformes univariantes independientes (la demostración es opcional):

$$f(x) = \int_3^4 \frac{1}{2} dy = \frac{1}{2}[y]_3^4 = \frac{1}{2} \quad x \in [0, 2]$$

$$f(y) = \int_0^2 \frac{1}{2} dx = \frac{1}{2}[x]_0^2 = 1 \quad y \in [3, 4]$$

La independencia se deduce del hecho de que $f(x, y) = f(x)f(y)$.

Entonces la esperanza es $E(\mathbf{x}) = (E(x_1), E(x_2)) = (1, 3.5)$ y la matriz de varianzas-covarianzas es

$$\text{var}(\mathbf{x}) = \begin{pmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) \\ \text{cov}(x_2, x_1) & \text{var}(x_2) \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/12 \end{pmatrix}$$

b) Dada una muestra aleatoria simple $\mathbf{x}_1, \dots, \mathbf{x}_{30}$ de \mathbf{x} , calcular $E(\bar{\mathbf{x}})$ y $\text{var}(\bar{\mathbf{x}})$.

Sabemos que $E(\bar{\mathbf{x}}) = E(\mathbf{x}) = (E(x_1), E(x_2))$, de modo que

$$E(\bar{\mathbf{x}}) = (1, 3.5)$$

y la matriz de varianzas-covarianzas de la media $\bar{\mathbf{x}}$ es

$$\text{var}(\bar{\mathbf{x}}) = \begin{pmatrix} \text{var}(\bar{x}_1) & \text{cov}(\bar{x}_1, \bar{x}_2) \\ \text{cov}(\bar{x}_2, \bar{x}_1) & \text{var}(\bar{x}_2) \end{pmatrix} = \begin{pmatrix} \text{var}(x_1)/n & 0 \\ 0 & \text{var}(x_2)/n \end{pmatrix} = \begin{pmatrix} 1/90 & 0 \\ 0 & 1/360 \end{pmatrix}$$

c) Generar una muestra aleatoria de \mathbf{x} de tamaño 30, dibujarla en un gráfico de dispersión y marcar los puntos $\bar{\mathbf{x}}$ y $E(\bar{\mathbf{x}})$. Calcular también la matriz de covarianzas muestrales \mathbf{S} .

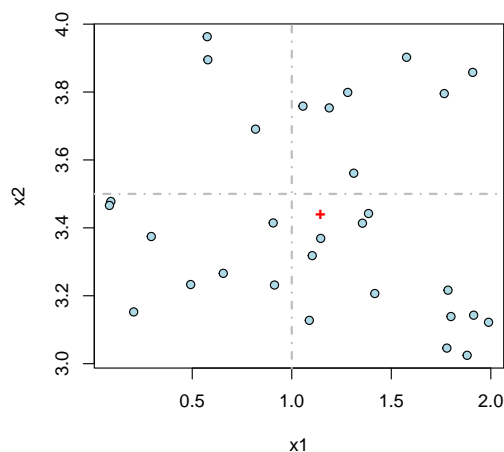
```
> # Generación de la muestra
> set.seed(123)
> x <- data.frame(x1=runif(30,0,2), x2=runif(30,3,4))
> # Media de la muestra
> media <- colMeans(x); media

      x1      x2
1.144800 3.438641

> # matriz de covarianzas de la muestra
> cov(x)

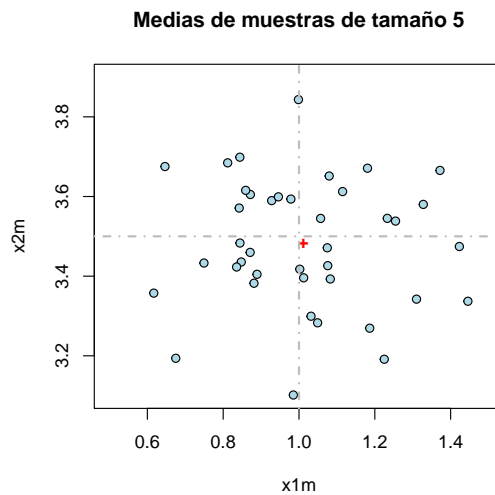
      x1      x2
x1 0.33956812 -0.02584134
x2 -0.02584134 0.08407819

> # Representación gráfica
> plot(x, pch=21, bg="lightblue")
> points(media[1], media[2], col="red", pch="+", font=2)
> abline(h=3.5, v=1, lty=4, lwd=2, col="gray")
```

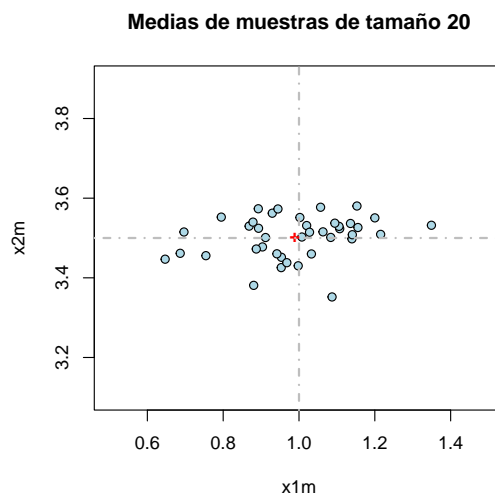


- d) Generar 40 muestras de tamaño 5, calcular sus medias muestrales y dibujarlas en un gráfico de dispersión, donde también se representa $E(\bar{x})$. Repetir este proceso en gráficos distintos para 40 muestras de tamaño 20 y otras 40 de tamaño 50. ¿Qué se observa?

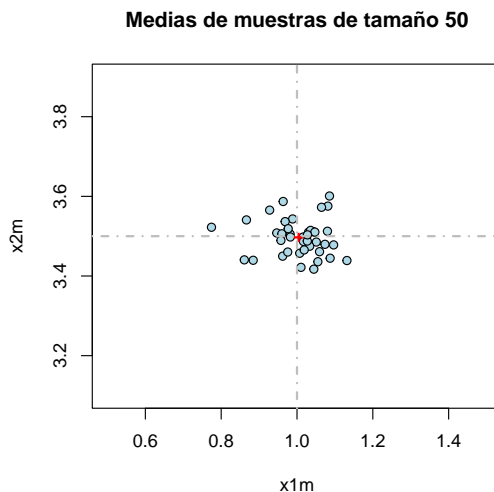
```
> grafico.de.medias <- function(n=5){
+   x1 <- array(runif(40*n,0,2), dim = c(40,n))
+   x2 <- array(runif(40*n,3,4), dim = c(40,n))
+   x1m <- apply(x1,1,mean)
+   x2m <- apply(x2,1,mean)
+   plot(x1m, x2m, pch=21, bg="lightblue", xlim=c(0.5,1.5), ylim=c(3.1,3.9))
+   points(mean(x1m),mean(x2m), col="red", pch="+", font=2)
+   abline(h=3.5, v=1, lty=4, lwd=2, col="gray")
+   title(paste("Medias de muestras de tamaño",n))
+ }
> grafico.de.medias()
```



```
> grafico.de.medias(20)
```



```
> grafico.de.medias(50)
```



Al aumentar el tamaño de la muestra disminuye la variabilidad de la media muestral. Como consecuencia la nube de puntos que representa las medias de un conjunto determinado de muestras aparece mas concentrada alrededor de la media poblacional al aumentar dicho tamaño muestral.

Ejercicio 4 (*) Esperanza y varianza condicionadas

Para calcular la esperanza y la matriz de covarianzas del vector \mathbf{x}_1 dados los valores del vector \mathbf{x}_2 , particionamos el vector aleatorio en dos partes $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)'$ con vector de medias $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)'$ y la matriz de covarianzas del vector \mathbf{x} en bloques asociados a estos dos vectores, como:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

Entonces, la distribución condicionada $\mathbf{x}_1|\mathbf{x}_2$ tiene esperanza

$$E[\mathbf{x}_1|\mathbf{x}_2] = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

y matriz de varianzas y covarianzas

$$\text{var}[\mathbf{x}_1|\mathbf{x}_2] = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

- a) Obtener las distribuciones condicionadas en la normal bivalente⁴ de media cero y matriz de covarianzas

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$E(x_1|x_2) = \mu_1 + \rho(x_2 - \mu_2)$$

$$E(x_2|x_1) = \mu_2 + \rho(x_1 - \mu_1)$$

$$\text{var}(x_1|x_2) = 1 - \rho^2$$

$$\text{var}(x_2|x_1) = 1 - \rho^2$$

⁴La distribución condicionada de una normal es normal.

b) Con la matriz de covarianzas

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ & 1 & 0 \\ & & 1 \end{pmatrix}$$

probar que la distribución condicionada de (x_1, x_2) dada x_3 , tiene como vector de medias $(\mu_1 + \rho^2(x_3 - \mu_3), \mu_2)'$ y matriz de covarianzas

$$\begin{pmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\begin{aligned} E((x_1, x_2)|x_3) &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho^2 \\ 0 \end{pmatrix} (x_3 - \mu_3) = \begin{pmatrix} \mu_1 + \rho^2(x_3 - \mu_3) \\ \mu_2 \end{pmatrix} \\ \text{var}((x_1, x_2)|x_3) &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho^2 \\ 0 \end{pmatrix} \begin{pmatrix} \rho^2 & 0 \end{pmatrix} = \begin{pmatrix} 1 - \rho^4 & \rho \\ \rho & 1 \end{pmatrix} \end{aligned}$$

Ejercicio 5

a) Generar una muestra de tamaño n de una normal bivalente $(X_1, X_2) \sim N_2(\mathbf{0}, \mathbf{I})$.⁵

```
> # Generación de una muestra de tamaño 200 de N2(0;I)
> set.seed(123)
> n <- 200
> N2 <- cbind(x1=rnorm(n,0,1), x2=rnorm(n,0,1))
```

b) Comprobar que las variables transformadas

$$\begin{aligned} Y_1 &= \mu_1 + \sigma_1 X_1 \\ Y_2 &= \mu_2 + \sigma_2(\rho X_1 + \sqrt{1 - \rho^2} X_2) \end{aligned}$$

siguen una distribución $N_2(\mu, \Sigma)$ con

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ & \sigma_2^2 \end{pmatrix}$$

Dado que $E(X_1) = E(X_2) = 0$, la media de las nuevas variables es:

$$\begin{aligned} E(Y_1) &= E(\mu_1 + \sigma_1 X_1) = \mu_1 + \sigma_1 E(X_1) = \mu_1 \\ E(Y_2) &= E(\mu_2 + \sigma_2(\rho X_1 + \sqrt{1 - \rho^2} X_2)) = \mu_2 + \sigma_2 E(\rho X_1 + \sqrt{1 - \rho^2} X_2) \\ &= \mu_2 + \sigma_2(\rho E(X_1) + \sqrt{1 - \rho^2} E(X_2)) = \mu_2 \end{aligned}$$

Dado que $\text{var}(X_1) = \text{var}(X_2) = 1$ y $\text{cov}(X_1, X_2) = 0$, las varianzas y covarianzas son:

$$\begin{aligned} \text{var}(Y_1) &= \text{var}(\mu_1 + \sigma_1 X_1) = \sigma_1^2 \text{var}(X_1) = \sigma_1^2 \\ \text{var}(Y_2) &= \text{var}(\mu_2 + \sigma_2(\rho X_1 + \sqrt{1 - \rho^2} X_2)) = \sigma_2^2 \text{var}(\rho X_1 + \sqrt{1 - \rho^2} X_2) \\ &= \sigma_2^2 \rho^2 \text{var}(X_1) + \sigma_2^2 (1 - \rho^2) \text{var}(X_2) + 2\sigma_2^2 \rho \sqrt{1 - \rho^2} \text{cov}(X_1, X_2) \\ &= \sigma_2^2 \rho^2 + \sigma_2^2 (1 - \rho^2) = \sigma_2^2 \\ \text{cov}(Y_1, Y_2) &= \text{cov}(\sigma_1 X_1, \sigma_2(\rho X_1 + \sqrt{1 - \rho^2} X_2)) \\ &= \sigma_1 \sigma_2 \rho \text{var}(X_1) + \sigma_1 \sigma_2 \sqrt{1 - \rho^2} \text{cov}(X_1, X_2) = \sigma_1 \sigma_2 \rho \end{aligned}$$

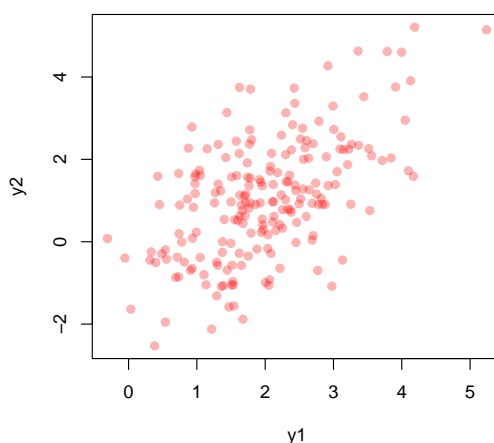
⁵Con la orden `rnorm()` de **R** se genera una muestra de una distribución normal estandar univariante.

- c) Con unos valores concretos $\boldsymbol{\mu} = (2, 1)'$, $\sigma_1 = 1$, $\sigma_2 = 1.5$ y $\rho = 0.6$, hallar con la transformación explicada en el apartado (b) una muestra aleatoria de (Y_1, Y_2) a partir de la muestra del primer apartado.

Tenemos que

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2 \rho & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

```
> transform <- function(x,mu1,mu2,s1,s2,rho){
+   y1 <- mu1 + s1*x[1]
+   y2 <- mu2 + s2*rho*x[1] + s2*sqrt(1-rho^2)*x[2]
+   c(y1,y2)
+ }
> N2Y <- t(apply(N2,1,transform,mu1=2,mu2=1,s1=1,s2=1.5,rho=0.6))
> colnames(N2Y) <- c("y1","y2")
> plot(N2Y, pch=19, col=rgb(1,0,0,alpha = 0.3))
```



Ejercicio 6

Generar muestras de una distribución normal bivalente $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ por el método siguiente: (1) generar un valor al azar de la distribución marginal de la primera variable con media μ_1 y varianza σ_1^2 ; (2) generar un valor al azar de la distribución univariante de la segunda variable dada la primera.

Para generar el segundo valor dado un valor x_1 , hay que tener en cuenta que

$$E(X_2|x_1) = \mu_2 + \sigma_{21}\sigma_{11}^{-1}(x_1 - \mu_1) = \mu_2 + \sigma_1^{-1}\sigma_2\rho(x_1 - \mu_1)$$

$$\text{var}(X_2|x_1) = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12} = \sigma_2^2(1 - \rho^2)$$

- a) Aplicar este método para generar valores al azar de un vector aleatorio con vector de medias $\boldsymbol{\mu} = (0, 5)'$, desviaciones típicas $(\sigma_1, \sigma_2) = (2, 3)$ y correlación $\rho = 0.5$.

```
> rnorm2 <- function(n,mu1,mu2,s1,s2,rho){
+   x1 <- rnorm(n, mu1, s1)
+   x2 <- rnorm(n, mu2 + rho*s2*(x1-mu1)/s1, s2*sqrt(1-rho^2))
+   cbind(x1=x1,x2=x2)
+ }
> set.seed(123)
> smp1 <- rnorm2(500,0,5,2,3,0.5)
```


Con esta muestra los estadísticos muestrales son:

```
> colMeans(smpl)

      x1      x2
0.0691809 5.0458199

> cov(smpl)

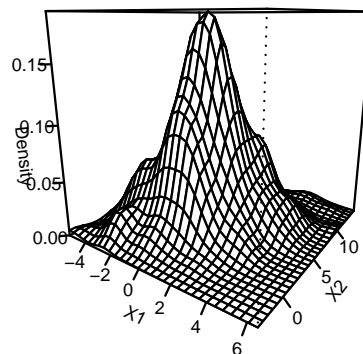
      x1      x2
x1 3.785121 2.573455
x2 2.573455 8.629048

> cor(smpl)

      x1      x2
x1 1.0000000 0.4502931
x2 0.4502931 1.0000000
```

- b) Utilizar los datos para estimar una densidad bivalente con la función `bivden()` de Everitt(2005)⁶ y representarla con un gráfico en perspectiva con la función `persp()`.

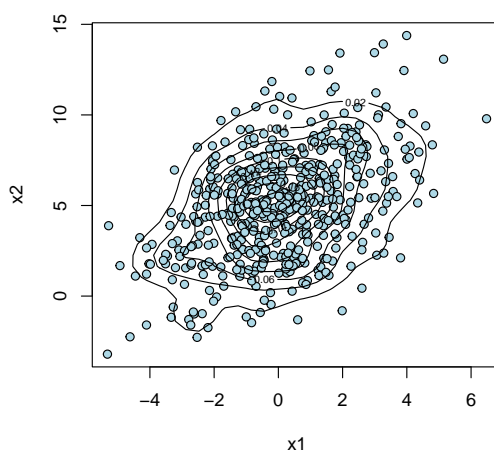
```
> den1 <- bivden(smpl[,1], smpl[,2])
> persp(den1$seqx, den1$seqy, den1$den, xlab="X1", ylab="X2", zlab="Density",
+       lwd=2, ticktype="detailed", theta=35)
> # Try different values of theta (between 0 and 360) to get the best perspective.
```



- c) Dibujar con los mismos datos un diagrama de dispersión y añadir un gráfico de niveles con la función `contour()`.

```
> plot(smpl, pch=21, bg="lightblue")
> contour(den1$seqx, den1$seqy, den1$den, nlevels=10, add=T)
```

⁶Ver la página web <http://biostatistics.iop.kcl.ac.uk/publications/everitt/>.



Ejercicio 7

Para una normal bivalente de media $(1, 2)'$ y matriz de covarianzas

$$\begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$$

a) Representarla con un gráfico en tres dimensiones⁷.

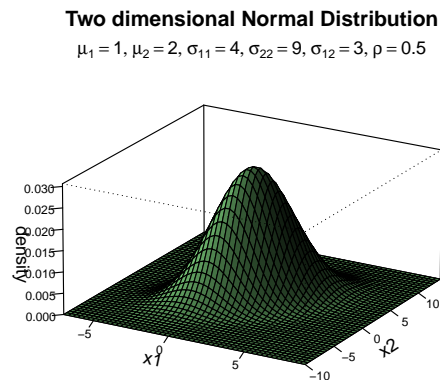
```
> mu1<-1 # setting the expected value of x1
> mu2<-2 # setting the expected value of x2
> s11<-4 # setting the variance of x1
> s12<-3 # setting the covariance between x1 and x2
> s22<-9 # setting the variance of x2
> rho<-s12/sqrt(s11*s22) # setting the correlation coefficient between x1 and x2
>
> x1 <- seq(mu1-4*sqrt(s11), mu1+4*sqrt(s11), length=51) # the vector series x1
> x2 <- seq(mu2-4*sqrt(s22), mu2+4*sqrt(s22), length=51) # the vector series x2
> #
> f<-function(x1,x2){
+   Q <- (x1-mu1)^2/s11 - 2*rho*(x1-mu1)*(x2-mu2)/sqrt(s11*s22) + (x2-mu2)^2/s22
+   1/(2*pi*sqrt(s11*s22*(1-rho^2)))*exp(-Q/(2*(1-rho^2)))
+ } # setting up the function of the multivariate normal density
> #
> z <- outer(x1,x2,f) # calculating the density values
> #
> persp(x1, x2, z, # produces the 3-D plot
+       main="Two dimensional Normal Distribution",
+       col="lightgreen",
+       theta=30, phi=20,
+       r=50,
+       d=0.1,
+       expand=0.5,
+       ltheta=90, lphi=180,
```

⁷Por ejemplo, en la página <http://www.stat.cmu.edu/~kass/KEB/RHTML/R/bivariateNormalPerspectives.r.html> tenemos un código que se puede adaptar. No confundir la correlación entre dos variables con su covarianza.

```

+     shade=0.75,
+     ticktype="detailed",
+     cex.axis=0.7,
+     zlab="density")
> # adding a text line to the graph
> mtext(expression(list(mu[1]==1,mu[2]==2,sigma[11]==4,sigma[22]==9,
+                       sigma[12]==3,rho==0.5))), side=3)

```

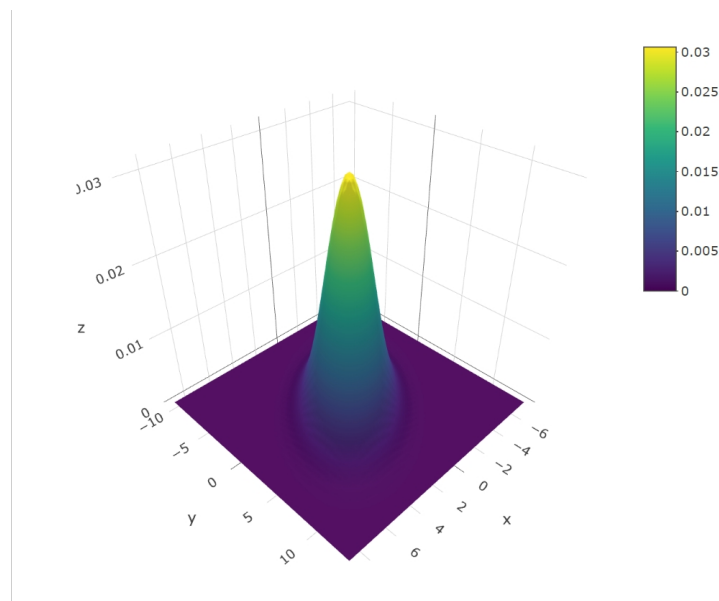


Con las funciones del paquete `plotly` el gráfico es más sencillo e interactivo.

```

> library(plotly)
> plot_ly() %>% add_surface(x = x1, y = x2, z = t(z))

```



b) Calcular la probabilidad del rectángulo $P(1 < X_1 \leq 1.5, 2 < X_2 \leq 2.75)$.

Se puede utilizar la función de distribución `pmvnorm()` del paquete `mvtnorm` de **R**.

```
> library(mvtnorm)
> Sm <- matrix(c(4,3,3,9), ncol=2)
> pmvnorm(lower=c(1,2),upper=c(1.5,2.75),mean=c(1,2),sigma=Sm)[[1]]

[1] 0.01128891
```

Ejercicio 8 (**)

Dibujar el gráfico de la sección 4.6 del curso online STAT 505

<https://newonlinecourses.science.psu.edu/stat505/node/36/>

Fijamos los parámetros de la normal que podemos ajustar a nuestro gusto.

```
> mu1<-2 # setting the expected value of x1
> mu2<-2 # setting the expected value of x2
> s11<-4 # setting the variance of x1
> s22<-4 # setting the variance of x2
> rho<-0.7 # setting the correlation coefficient between x1 and x2
> s12<-sqrt(s11)*sqrt(s22)*rho # setting the covariance between x1 and x2
```

Calculamos los vectores y valores propios de la matriz de covarianzas:

```
> cov.mat <- matrix(c(s11,s12,s12,s22),ncol=2)
> ee <- eigen(cov.mat)
> ee$vectors

      [,1]      [,2]
[1,] 0.7071068 -0.7071068
[2,] 0.7071068  0.7071068
```

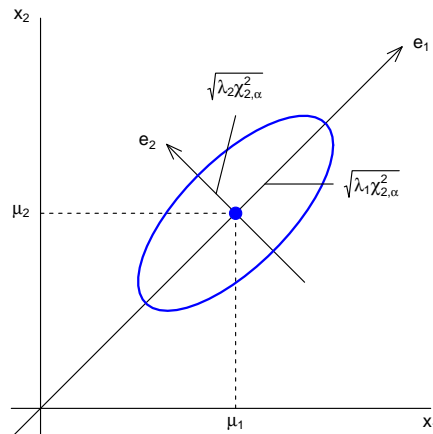
El gráfico muestra una elipse de equiprobabilidad para una distribución normal bivalente. El interior de la elipse es la región de predicción centrada en las medias poblacionales con probabilidad $(1 - \alpha) \times 100\%$.

```
> library(car)
> par(mar=c(2.1,2.1,2.1,2.1))
> plot.new()
> plot.window(xlim=c(-0.3,4), ylim=c(-0.3,4))
> segments(-0.3,0,4,0)
> segments(0,-0.3,0,4)
> text(4,0,expression(x[1]),pos=1)
> text(0,4,expression(x[2]),pos=2)
> segments(2,0,2,2,lty=2)
> segments(0,2,2,2,lty=2)
> arrows(x0 = -0.3,y0=-0.3,x1 = 3+ee$vectors[1,1],y1=3+ee$vectors[2,1],
+       length = 0.15, angle = 20)
> text(3+ee$vectors[1,1],3+ee$vectors[2,1],expression(e[1]),pos = 4)
> arrows(x0 = 2-ee$vectors[1,2],y0=2-ee$vectors[2,2],
+       x1 = 2+ee$vectors[1,2],y1=2+ee$vectors[2,2],
+       length = 0.15, angle = 20)
> text(2+ee$vectors[1,2],2+ee$vectors[2,2],expression(e[2]),pos = 2)
> text(2,0,expression(mu[1]),pos = 1)
```

```

> text(0,2,expression(mu[2]),pos = 2)
> ellipse(center=c(mu1,mu2), shape = cov.mat, radius=0.5)
> text(3.4,2.3,expression(sqrt(paste(lambda[1],chi[paste(2,"",alpha)]^2))))
> segments(3,2.3,2.3,2.3)
> text(2,3.3,expression(sqrt(paste(lambda[2],chi[paste(2,"",alpha)]^2))))
> segments(2,3,1.8,2.2)

```



Para producir el gráfico anterior se han utilizado la mayoría de funciones gráficas de **R**. En particular la función `ellipse` del paquete `car`.

Ejercicio 9 (**) Geometría de la Distribución Normal Multivariante

Calcular los semiejes l_j del elipsoide al 95 % para las variables numéricas de los datos *crabs* del paquete *MASS* para los individuos machos de la especie *Blue* tal como se explica en la página:

<https://newonlinecourses.science.psu.edu/stat505/node/36/>

Calcular también el volumen del mismo elipsoide.

Los datos para este ejercicio son:

```

> library(MASS)
> data(crabs)
> crabs5BM <- crabs[crabs$sp=="B" & crabs$sex=="M",4:8]
> p <- dim(crabs5BM)[2]

```

Los semiejes del elipsoide (elipse con dos variables) son

$$l_j = \sqrt{\lambda_j \chi^2_{p,\alpha}}$$

donde λ_j son los valores propios de la descomposición de la matriz de varianzas-covarianzas. Así pues, necesitamos los valores propios de la matriz de covarianzas.

```

> val_propios <- eigen(cov(crabs5BM))$values

```

Entonces los semiejes son:

```
> sqrt(val_propios*qchisq(0.95,df=p))
[1] 40.4363294 1.5570380 1.0730279 0.9219061 0.7703722
```

El volumen (área con dos variables) en mm^5 es

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)}(\chi_{p,\alpha}^2)^{p/2}|\Sigma|^{1/2}$$

donde $|\Sigma|^{1/2}$ es la raíz cuadrada de la varianza generalizada.
Los cálculos en este caso son así:

```
> ((2*pi^(p/2))/(p*gamma(p/2))*qchisq(0.95,df=p)^(p/2)*det(cov(crabs5BM))^(1/2)
[1] 252.5618
```

o lo que es equivalente

```
> ((2*pi^(p/2))/(p*gamma(p/2))*qchisq(0.95,df=p)^(p/2)*prod(val_propios)^(1/2)
[1] 252.5618
```

ya que el determinante de la matriz de covarianzas se puede calcular con el producto de sus valores propios.

Ejercicio 10

En la Tabla 1 se muestran los datos de 28 alcornoques que miden los depósitos de corcho (en centigramos) en cada uno de los puntos cardinales: N , E , S , W . El vector de medias es

$$\bar{\mathbf{x}} = (50.536, 46.179, 49.679, 45.179)'$$

y la matriz de covarianzas

$$\mathbf{S} = \begin{pmatrix} 280 & 216 & 278 & 218 \\ & 212 & 221 & 165 \\ & & 337 & 250 \\ & & & 218 \end{pmatrix}$$

Las siguientes variables compuestas explican diferentes aspectos de la variabilidad de los datos:

$$\begin{array}{ll} \text{Contraste eje } N - S \text{ con eje } E - W: & Y_1 = N + S - E - W \\ \text{Contraste } N - S: & Y_2 = N - S \\ \text{Contraste } E - W: & Y_3 = E - W \end{array}$$

- a) Realizar una matriz de dispersiones de las variables originales dos a dos. Añadir las rectas de regresión y los histogramas en la diagonal.

```
> corcho <- read.table("corcho.dat", header=TRUE)
> colMeans(corcho)

      N      E      S      W
50.53571 46.17857 49.67857 45.17857

> cov(corcho)
```

```

      N      E      S      W
N 290.4061 223.7526 288.4378 226.2712
E 223.7526 219.9299 229.0595 171.3743
S 288.4378 229.0595 350.0040 259.5410
W 226.2712 171.3743 259.5410 226.0040

```

```

> n <- nrow(corcho)
> Sm <- cov(corcho)*(n-1)/n; round(Sm,0)

```

```

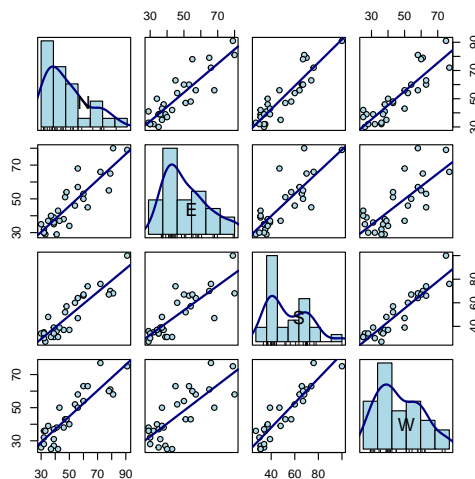
      N      E      S      W
N 280 216 278 218
E 216 212 221 165
S 278 221 338 250
W 218 165 250 218

```

```

> # Función de panel diagonal: histograma y densidad
> dgp.fn <- function(x,...){
+   par(new=TRUE)
+   hist(x,col="lightblue",probability=TRUE,axes=FALSE,main="")
+   lines(density(x),col="navy",lwd=2)
+   rug(x)
+ }
> # Función de panel no diagonal: diagrama dispersión y recta de regresión
> pn.fn <- function(x,y,...){
+   points(x,y,pch=21,bg="lightblue")
+   abline(lm(y~x),col="navy",lwd=2);
+ }
> # Matriz de diagramas de dispersión
> pairs(corcho, panel=pn.fn, diag.panel=dgp.fn, label.pos=0.3, cex.labels=1.5)

```



b) Calcular el vector de medias y la matriz de covarianzas de las variables compuestas.

```

> Y <- data.frame(Y1=corcho$N+corcho$S-corcho$E-corcho$W,
+                 Y2=corcho$N-corcho$S,

```

```
+          Y3=corcho$E-corcho$W)
> colMeans(Y)

      Y1      Y2      Y3
8.8571429 0.8571429 1.0000000

> cov(Y)

      Y1      Y2      Y3
Y1 128.71958 -21.02116 -26.92593
Y2 -21.02116  63.53439  27.96296
Y3 -26.92593  27.96296 103.18519
```

- c) Lo mismo para las variables transformadas y normalizadas (la suma de cuadrados de los coeficientes es 1).

```
> a <- c(1,-1,1,-1) # N,E,S,W
> norma.a <- sqrt(sum(a^2)) # sqrt(a[1]^2+a[2]^2+a[3]^2+a[4]^2)
> b <- c(1,0,-1,0) # N,E,S,W
> norma.b <- sqrt(sum(b^2))
> c <- c(0,1,0,-1) # N,E,S,W
> norma.c <- sqrt(sum(c^2))
> Y.norm <- data.frame(Z1=Y[,1]/norma.a,Z2=Y[,2]/norma.b,Z3=Y[,3]/norma.c)
> # Y.norm <- data.frame( as.matrix(Y) %*% diag(c(1/norma.a,1/norma.b,1/norma.c)))
> colMeans(Y.norm)

      Z1      Z2      Z3
4.4285714 0.6060915 0.7071068

> cov(Y.norm)

      Z1      Z2      Z3
Z1 32.179894 -7.432104 -9.519752
Z2 -7.432104  31.767196 13.981481
Z3 -9.519752 13.981481 51.592593
```

Ejercicio 11 (**)

Probar que si la matriz de varianzas y covarianzas muestral de unos datos normales es

$$\mathbf{S} = \frac{1}{n} \mathbf{X}' \mathbf{H} \mathbf{X}$$

donde \mathbf{H} es la matriz de centrado, entonces

$$n\mathbf{S} \sim W_p(\boldsymbol{\Sigma}, n-1)$$

Supongamos que \mathbf{X} es un conjunto de datos de una distribución normal multivariante con matriz de varianzas-covarianzas $\boldsymbol{\Sigma}$.

Recordemos que la matriz de centrado es simétrica $\mathbf{H}' = \mathbf{H}$ e idempotente $\mathbf{H}^2 = \mathbf{H}$, de modo que

$$n\mathbf{S} = \mathbf{X}' \mathbf{H} \mathbf{X} = (\mathbf{H} \mathbf{X})' \mathbf{H} \mathbf{X}$$

donde $\mathbf{Z} = \mathbf{H}\mathbf{X}$ es la matriz de datos de una normal multivariante (por ser transformación lineal de una normal) con distribución $N_p(0, \Sigma)$. Entonces por definición, $n\mathbf{S} = \mathbf{Z}'\mathbf{Z}$ es una Whishart con matriz de escala Σ .

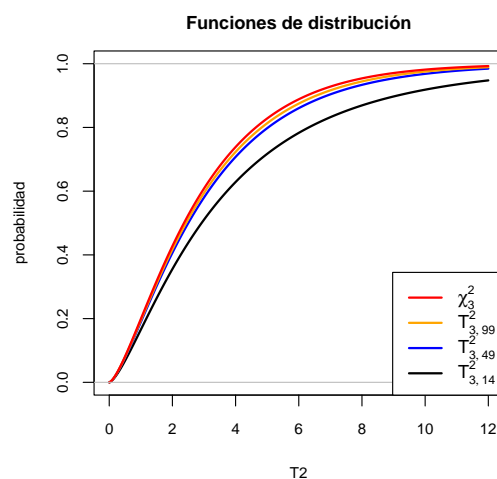
En cuanto a los grados de libertad, son $n - 1$ ya que $\text{rango}(\mathbf{H}) = n - 1$ y esto añade una dependencia a las observaciones (filas de la matriz de datos).

Ejercicio 12

Para observar gráficamente que $T^2(p, n - 1) \rightarrow \chi_p^2$, representar en un mismo gráfico las densidades $T^2(3, 14)$, $T^2(3, 49)$ y χ_3^2

La aproximación se considera válida si el cociente $n/p > 15$.

```
> # Función de distribución de la T2
> pT2 <- function(T2,p,m){
+   F <- (m-p+1)/(m*p)*T2
+   pf(F,p,m-p+1)
+ }
> # Representación gráfica
> T2 <- seq(0,12,by=0.005)
> par(cex.main=1,cex.lab=0.8,cex.axis=0.8)
> plot(T2, pT2(T2,3,14), type="l", lwd=2, ylab="probabilidad", ylim=c(0,1))
> points(T2,pT2(T2,3,49), type="l", col="blue", lwd=2)
> points(T2,pT2(T2,3,99), type="l", col="orange", lwd=2)
> points(T2,pchisq(T2,3), type="l", col="red", lwd=2)
> abline(h=c(0,1), col="gray")
> legend("bottomright", legend=c(expression(chi[3]^{2}),
+                               expression(T[list(3,99)]^{2}),
+                               expression(T[list(3,49)]^{2}),
+                               expression(T[list(3,14)]^{2})),
+       lwd=2, col=c("red","orange","blue","black"), bg="white")
> title(main="Funciones de distribución",line=1)
```



Ejercicio 13

Con la matriz de datos siguiente

X_1	X_2	X_3
6.0	4.7	3.3
5.0	4.2	2.9
4.5	5.1	3.2
6.2	4.9	4.0
5.4	4.5	3.7
5.7	4.6	2.2
7.2	5.1	4.4
6.5	6.4	5.3
2.8	3.9	2.5
5.6	4.6	3.2
6.1	4.9	4.0
4.5	4.3	3.5

- a) Calcular el vector de medias, la matriz de covarianzas y la matriz de correlaciones.

```
> datos <- matrix(c(6.0,4.7,3.3,
+                   5.0,4.2,2.9,
+                   4.5,5.1,3.2,
+                   6.2,4.9,4.0,
+                   5.4,4.5,3.7,
+                   5.7,4.6,2.2,
+                   7.2,5.1,4.4,
+                   6.5,6.4,5.3,
+                   2.8,3.9,2.5,
+                   5.6,4.6,3.2,
+                   6.1,4.9,4.0,
+                   4.5,4.3,3.5), byrow = T, ncol=3)
> colnames(datos) <- c("X1", "X2", "X3")
> (medias <- colMeans(datos))
```

	X1	X2	X3
	5.458333	4.766667	3.516667

```
> (cov.datos <- cov(datos))
```

	X1	X2	X3
X1	1.3244697	0.4612121	0.6253030
X2	0.4612121	0.3951515	0.4278788
X3	0.6253030	0.4278788	0.7142424

```
> (cor.datos <- cor(datos))
```

	X1	X2	X3
X1	1.0000000	0.6375260	0.6429048
X2	0.6375260	1.0000000	0.8054083
X3	0.6429048	0.8054083	1.0000000

- b) Calcular la varianza generalizada, la varianza total y el coeficiente de dependencia global $\eta^2 = 1 - |\mathbf{R}|$.

```

> sum(diag(cov.datos)) # varianza total

[1] 2.433864

> det(cov.datos) # varianza generalizada

[1] 0.07168729

> 1-det(cor.datos) # coeficiente de dependencia global

[1] 0.8082255

```

- c) Contrastar las hipótesis univariantes $H_0^{(1)} : \mu_1 = 5$, $H_0^{(2)} : \mu_2 = 4$ y $H_0^{(3)} : \mu_3 = 3$. Utilizar, por ejemplo, la función `t.test` de **R**.

```

> t.test(datos[,1], mu = 5)

One Sample t-test

data:  datos[, 1]
t = 1.3796, df = 11, p-value = 0.1951
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
 4.727114 6.189552
sample estimates:
mean of x
 5.458333

> t.test(datos[,2], mu = 4)

One Sample t-test

data:  datos[, 2]
t = 4.2249, df = 11, p-value = 0.001425
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 4.367266 5.166067
sample estimates:
mean of x
 4.766667

> t.test(datos[,3], mu = 3)

One Sample t-test

data:  datos[, 3]
t = 2.1178, df = 11, p-value = 0.05779
alternative hypothesis: true mean is not equal to 3

```

```

95 percent confidence interval:
 2.979698 4.053636
sample estimates:
mean of x
 3.516667

```

Con un nivel de significación $\alpha = 0.05$, únicamente la segunda hipótesis es significativa.

- d) Contrastar la hipótesis $H_0 : (\mu_1, \mu_2, \mu_3)' = (5, 4, 3)'$ con el estadístico T^2 de Hotelling y hallar el estadístico F correspondiente.

```

> # Test multivariante
> n <- nrow(datos)
> p <- ncol(datos)
> mus <- c(5,4,3)
> F.star <- (n-p)/p * t(medias-mus) %*% solve(cov.datos) %*%(medias-mus)
> p.valor <- pf(F.star, p, n-p, lower.tail = F)
> c(F.star=F.star,p.valor=p.valor)

      F.star      p.valor
5.96093644 0.01601301

```

La hipótesis multivariante es significativa.

Ejercicio 14 (**) Contrastes sobre la matriz de covarianzas de una población normal

En una población normal multivariante $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ el contraste

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \quad H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$$

se realiza con la razón de verosimilitudes.

Si el logaritmo de la verosimilitud es

$$l = \log L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{x}) = -\frac{n}{2} \log |2\pi\boldsymbol{\Sigma}| - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

Bajo la hipótesis nula, los estimadores de los parámetros son $\bar{\mathbf{x}}$ y $\boldsymbol{\Sigma}_0$, mientras que en general son $\bar{\mathbf{x}}$ y \mathbf{S} . De modo que

$$\begin{aligned}
 l_0 &= -\frac{n}{2} \log |2\pi\boldsymbol{\Sigma}_0| - \frac{n}{2} \text{tr}(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}) \\
 l_1 &= -\frac{n}{2} \log |2\pi\mathbf{S}| - \frac{np}{2}
 \end{aligned}$$

Así que

$$-2 \log \lambda = 2(l_1 - l_0) = n \log \frac{|\boldsymbol{\Sigma}_0|}{|\mathbf{S}|} + n \text{tr}(\boldsymbol{\Sigma}_0^{-1}\mathbf{S}) - np$$

que asintóticamente tiene distribución $\chi^2_{p(p+1)/2}$.

- a) Comprobar los cálculos anteriores.

$$\lambda = \frac{L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{x})}{L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{x})}$$

$$\log \lambda = \log(L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{x})) - \log(L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{x})) = l_0 - l_1$$

donde

$$\begin{aligned} l_0 &= -\frac{n}{2} \log |2\pi \Sigma_0| - \frac{n}{2} \text{tr}(\Sigma_0^{-1} \mathbf{S}) \\ l_1 &= -\frac{n}{2} \log |2\pi \mathbf{S}| - \frac{n}{2} \text{tr}(\mathbf{S}^{-1} \mathbf{S}) \\ &= -\frac{n}{2} \log |2\pi \mathbf{S}| - \frac{n}{2} \text{tr}(\mathbf{I}) \\ &= -\frac{n}{2} \log |2\pi \mathbf{S}| - \frac{n}{2} p \end{aligned}$$

Finalmente

$$\begin{aligned} -2 \log \lambda &= 2(l_1 - l_0) = 2\left(-\frac{n}{2} \log |2\pi \mathbf{S}| - \frac{n}{2} p + \frac{n}{2} \log |2\pi \Sigma_0| + \frac{n}{2} \text{tr}(\Sigma_0^{-1} \mathbf{S})\right) \\ &= n \log \frac{|\Sigma_0|}{|\mathbf{S}|} + n \text{tr}(\Sigma_0^{-1} \mathbf{S}) - np \end{aligned}$$

b) Concretar el test $H_0 : \Sigma = \mathbf{I}$.

En este caso tenemos:

$$\begin{aligned} -2 \log \lambda &= n \log \frac{|\mathbf{I}|}{|\mathbf{S}|} + n \text{tr}(\mathbf{I}^{-1} \mathbf{S}) - np \\ &= -n \log |\mathbf{S}| + n \text{tr}(\mathbf{S}) - np \end{aligned}$$

c) Dado que para poblaciones normales incorrelación implica independencia, el contraste de independencia es $H_0 : \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$.

La estimación de los parámetros bajo esa hipótesis es $\bar{\mathbf{x}}$ y $\text{diag}(\mathbf{S})$.

Comprobar que el estadístico para contrastar la independencia es

$$-2 \log \lambda = -n \log |\mathbf{R}|$$

cuya distribución asintótica es $\chi_{p(p-1)/2}^2$.

Sea $\mathbf{D} = \text{diag}(\mathbf{S})$ la estimación de la matriz $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$ bajo la hipótesis nula.

$$\begin{aligned} -2 \log \lambda &= n \log \frac{|\mathbf{D}|}{|\mathbf{S}|} + n \text{tr}(\mathbf{D}^{-1} \mathbf{S}) - np \\ &= -n \log \frac{|\mathbf{S}|}{|\mathbf{D}|} + np - np = -n \log |\mathbf{R}| \end{aligned}$$

ya que $\mathbf{D}^{-1} \mathbf{S}$ tiene todos los elementos de la diagonal iguales a 1, con lo que $\text{tr}(\mathbf{D}^{-1} \mathbf{S}) = np$.

Además

$$\frac{|\mathbf{S}|}{|\mathbf{D}|} = \frac{|\mathbf{S}|}{|\mathbf{D}^{1/2}| |\mathbf{D}^{1/2}|} = |\mathbf{D}^{-1/2}| |\mathbf{S}| |\mathbf{D}^{-1/2}| = |\mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2}| = |\mathbf{R}|$$

Ejercicio 15

Los vectores de medias y las matrices de covarianzas de $n_1 = 24$ tortugas macho y $n_2 = 24$ tortugas hembra, respecto a las variables $X_1 = \text{longitud}$, $X_2 = \text{anchura}$ y $X_3 = \text{altura del caparazón en mm}$, son:

$$\begin{aligned} \bar{\mathbf{x}}_m &= \begin{pmatrix} 113.38 \\ 88.29 \\ 40.71 \end{pmatrix} & \mathbf{S}_m &= \begin{pmatrix} 132.99 & 75.85 & 35.82 \\ & 47.96 & 20.75 \\ & & 10.79 \end{pmatrix} \\ \bar{\mathbf{x}}_h &= \begin{pmatrix} 136.00 \\ 102.58 \\ 51.96 \end{pmatrix} & \mathbf{S}_h &= \begin{pmatrix} 432.58 & 259.87 & 161.67 \\ & 164.57 & 98.99 \\ & & 63.87 \end{pmatrix} \end{aligned}$$

a) Calcular la matriz de covarianzas común \mathbf{S} .

```
> # Datos
> p <- 3
> n <- n1 <- n2 <- 24
> nms <- c("X1", "X2", "X3")
> means.m <- c(113.38, 88.29, 40.71)
> S.m <- matrix(c(132.99, 75.85, 35.82, 75.85, 47.96, 20.75, 35.82, 20.75, 10.79), nrow=3)
> means.h <- c(136.00, 102.58, 51.96)
> S.h <- matrix(c(432.58, 259.87, 161.67, 259.87, 164.57, 98.99, 161.67, 98.99, 63.87), nrow=3)
> # Matriz de varianzas común
> S <- ((n1-1)*S.m+(n2-1)*S.h)/(n1+n2-2)
> rownames(S) <- colnames(S) <- nms; S
```

	X1	X2	X3
X1	282.785	167.860	98.745
X2	167.860	106.265	59.870
X3	98.745	59.870	37.330

b) Descartar la independencia de las variables en los dos grupos con el test $-2 \log \lambda = -n \log |\mathbf{R}|$ que sigue asintóticamente⁸ una distribución $\chi^2_{p(p-1)/2}$.

```
> # Test de independencia de variables
> # Para los machos
> D <- diag(1/sqrt(diag(S.m)))
> R <- D %*% S.m %*% D
> rownames(R) <- colnames(R) <- nms; R
```

	X1	X2	X3
X1	1.0000000	0.9497439	0.9455954
X2	0.9497439	1.0000000	0.9121532
X3	0.9455954	0.9121532	1.0000000

```
> stat.1 <- -n1*log(det(R))
> p.val1 <- pchisq(stat.1, p*(p-1)/2, lower.tail=FALSE)
> c(stat=stat.1, p.valor=p.val1)
```

	stat	p.valor
	1.101111e+02	1.038548e-23

```
> # Para las hembras
> D <- diag(1/sqrt(diag(S.h)))
> R <- D %*% S.h %*% D
> rownames(R) <- colnames(R) <- nms; R
```

	X1	X2	X3
X1	1.0000000	0.9739745	0.9726302
X2	0.9739745	1.0000000	0.9655339
X3	0.9726302	0.9655339	1.0000000

⁸Ver Peña (2002) pág. 295 o Cuadras(2018) pág. 52.

```
> stat.2 <- -n2*log(det(R))
> p.val2 <- pchisq(stat.2, p*(p-1)/2, lower.tail=FALSE)
> c(stat=stat.2,p.valor=p.val2)

      stat      p.valor
1.443602e+02 4.337409e-31
```

En ambos casos rechazamos la hipótesis de independencia.

- c) Comparar de forma univariante las medias de X_1 , X_2 y X_3 en las dos poblaciones.

En los siguientes test t univariantes, las varianzas se consideran iguales

```
> t <- (means.h-means.m) / ((sqrt(diag(S)))*sqrt(1/n1+1/n2)); t

      X1      X2      X3
4.659669 4.802062 6.378435

> pt(abs(t),n1+n2-2,lower.tail=FALSE)*2

      X1      X2      X3
2.727426e-05 1.702562e-05 7.825857e-08
```

Todas las diferencias resultan significativas al nivel 0.05.

- d) Comparar de forma multivariante las medias en las dos poblaciones.

Aquí consideramos las matrices de varianzas-covarianzas iguales.

```
> D2 <- t(means.h-means.m) %*% solve(S) %*% (means.h-means.m)
> D2 <- as.numeric(D2)
> T2 <- n1*n2/(n1+n2) * D2
> F <- (n1+n2-1-p)/((n1+n2-2)*p) * T2
> p.valor <- pf(F,p,n1+n2-1-p,lower.tail=FALSE)
> c(D2,T2,F,p.valor)

[1] 5.813085e+00 6.975702e+01 2.224137e+01 6.452240e-09
```

Los vectores de medias resultan significativamente distintos en la métrica común.

- e) Comparar las matrices de covarianzas de las dos poblaciones con el test de la razón de verosimilitudes:

$$-2 \log \lambda = n \log |\mathbf{S}| - n_1 \log |\mathbf{S}_m| - n_2 \log |\mathbf{S}_h|$$

que asintóticamente sigue una distribución $\chi^2_{p(p+1)/2}$.

Para ello debemos considerar las estimaciones de máxima verosimilitud de las covarianzas:

```
> S1 <- (n1-1)*S.m/n1
> S2 <- (n2-1)*S.h/n2
> S <- (n1*S1+n2*S2)/(n1+n2)
> llr <- (n1+n2)*log(det(S))-n1*log(det(S1))-n2*log(det(S2))
> p.valor <- pchisq(llr, p*(p+1)/2, lower.tail=FALSE)
> c(llr,p.valor)

[1] 2.698611e+01 1.456775e-04
```

Las matrices de covarianzas deben considerarse distintas lo que invalida el test de Hotelling del apartado anterior.

Como en el caso univariante, hay algunos tests para comparar medias con covarianzas distintas.

url: Two-sample Test for Multivariate Means by Krishnamoorthy and Yu (2004)

Ejercicio 16

La tabla 2 contiene las medidas de 5 variables biométricas sobre gorriones hembra, recogidos casi moribundos después de una tormenta. Los primeros 21 sobrevivieron mientras que los 28 restantes no lo consiguieron. Las variables observadas son X_1 = longitud total, X_2 = extensión del ala, X_3 = longitud del pico y de la cabeza, X_4 = longitud del húmero y X_5 = longitud del esternón. Suponiendo normalidad multivariante,

- a) Comparar las covarianzas entre el grupo de supervivientes y el de no supervivientes.

```
> load("gorriones.RData")
> str(gorriones)

'data.frame': 49 obs. of 6 variables:
 $ x1      : num  156 154 153 153 155 163 157 155 164 158 ...
 $ x2      : num  245 240 240 236 243 247 238 239 248 238 ...
 $ x3      : num  31.6 30.4 31 30.9 31.5 32 30.9 32.8 32.7 31 ...
 $ x4      : num  18.5 17.9 18.4 17.7 18.6 19 18.4 18.6 19.1 18.8 ...
 $ x5      : num  20.5 19.6 20.6 20.2 20.3 20.9 20.2 21.2 21.1 22 ...
 $ superviv: Factor w/ 2 levels "N","S": 2 2 2 2 2 2 2 2 2 2 ...

> names(gorriones) <- c("length", "wing", "head", "humerus", "sternum", "alive")
> attach(gorriones)
> # Dimensiones
> p <- 5
> n1 <- sum(alive=="S")
> n2 <- sum(alive=="N")
> n <- n1 + n2
> nms <- colnames(gorriones[,1:5])
>
> # Estadísticos por grupo
> medias1 <- colMeans(gorriones[alive=="S",1:5]); round(medias1,2)

length    wing    head humerus sternum
157.38    241.00    31.43    18.50    20.81

> S1 <- cov(gorriones[alive=="S",1:5]); round(S1,2)

      length    wing    head humerus sternum
length  11.05    9.10    1.56    0.87    1.29
wing     9.10   17.50    1.91    1.31    0.88
head     1.56    1.91    0.53    0.19    0.24
humerus  0.87    1.31    0.19    0.18    0.13
sternum  1.29    0.88    0.24    0.13    0.57

> medias2 <- colMeans(gorriones[alive=="N",1:5]); round(medias2,2)
```



```

length    wing    head humerus sternum
158.43   241.57   31.48  18.45  20.84

> S2 <- cov(gorriones[alive=="N",1:5]); round(S2,2)

      length  wing head humerus sternum
length  15.07 17.19 2.24   1.75   2.93
wing    17.19 32.55 3.40   2.95   4.07
head     2.24  3.40 0.73   0.47   0.56
humerus  1.75  2.95 0.47   0.43   0.51
sternum  2.93  4.07 0.56   0.51   1.32

> # Matriz de varianzas común
> S <- ((n1-1)*S1+(n2-1)*S2)/(n1+n2-2); round(S,2)

      length  wing head humerus sternum
length  13.36 13.75 1.95   1.37   2.23
wing    13.75 26.15 2.76   2.25   2.71
head     1.95  2.76 0.64   0.35   0.42
humerus  1.37  2.25 0.35   0.32   0.35
sternum  2.23  2.71 0.42   0.35   1.00

> # Matrices de covarianzas sesgadas (estimaciones maximoverosímiles)
> S1mv <- (n1-1)*S1/n1; S2mv <- (n2-1)*S2/n2
> Smv <- (n1*S1mv+n2*S2mv)/(n1+n2)
> # Comparación de matrices de covarianzas
> llr <- (n1+n2)*log(det(Smv))-n1*log(det(S1mv))-n2*log(det(S2mv))
> p.valor <- pchisq(llr, p*(p+1)/2, lower.tail=FALSE)
> c(llr=llr,p.valor=p.valor)

      llr    p.valor
12.5322335 0.6383781

```

Podemos aceptar la igualdad de varianzas.

Ahora vamos a aplicar el test M de Box que mejora notablemente el anterior.

```

> library(heplots)

Loading required package: broom

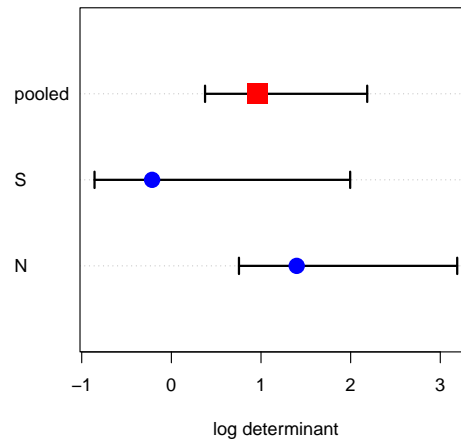
> M.test <- boxM(gorriones[,1:5],alive)
> M.test

Box's M-test for Homogeneity of Covariance Matrices

data:  gorriones[, 1:5]
Chi-Sq (approx.) = 10.408, df = 15, p-value = 0.7933

> # summary(M.test)
> plot(M.test)

```



b) Comparar las medias de las dos poblaciones con el estadístico T^2 de Hotelling.

```
> # Comparación multivariante de medias por Hotelling
> D2 <- t(medias1-medias2) %*% solve(S) %*% (medias1-medias2);
> D2 <- as.numeric(D2)
> T2 <- n1*n2/(n1+n2) * D2
> F <- (n1+n2-1-p)/((n1+n2-2)*p) * T2
> p.valor <- pf(F,p,n1+n2-1-p,lower.tail=FALSE)
> c(D2=D2,T2=T2,F=F,p.valor=p.valor)
```

	D2	T2	F	p.valor
	0.2353082	2.8236983	0.5166767	0.7621709

No se puede rechazar la igualdad de medias entre los dos grupos.

c) (*) Comparar las medias de las dos poblaciones con el estadístico Λ de Wilks.

```
> # Comparación multivariante de medias por Wilks
> # 1 - Dispersión entre grupos
> n.i <- c(n1,n2)
> medias <- colMeans(gorriones[,1:5])
> diff <- rbind(medias1,medias2)-matrix(medias, nrow=2, ncol=p, byrow=T)
> B <- t(diff) %*% diag(n.i) %*% diff; round(B,3)
```

	length	wing	head	humerus	sternum
length	13.170	7.184	0.569	-0.673	0.374
wing	7.184	3.918	0.310	-0.367	0.204
head	0.569	0.310	0.025	-0.029	0.016
humerus	-0.673	-0.367	-0.029	0.034	-0.019
sternum	0.374	0.204	0.016	-0.019	0.011

```
> # 2 - Dispersión total
> diff <- as.matrix(gorriones[,1:5])-matrix(medias,nrow=n,ncol=p,byrow=T)
> T <- t(diff) %*% diff; round(T,3)
```

```

      length      wing      head humerus sternum
length 640.980  653.327  92.259  63.869 105.227
wing    653.327 1232.776 130.253 105.490 127.576
head     92.259  130.253  30.318  16.429  19.903
humerus  63.869  105.490  16.429  15.284  16.290
sternum 105.227  127.576  19.903  16.290  47.176

> # 3 - Dispersión inter grupos
> W <- T-B; round(W,3)

      length      wing      head humerus sternum
length 627.810  646.143  91.690  64.543 104.852
wing    646.143 1228.857 129.943 105.857 127.371
head     91.690  129.943  30.294  16.458  19.887
humerus  64.543  105.857  16.458  15.250  16.309
sternum 104.852  127.371  19.887  16.309  47.165

> # 4 - Lambda de Wilks y gr de libertad
> lambda <- det(W) / det(W+B)
> p <- nrow(B) # Dimensión de las matrices
> r <- 2       # Niveles del factor
> nf <- r-1    # gl de la dispersión entre grupos B
> m <- n-r     # gl residuales
> c(p,m,nf)

[1] 5 47 1

> # 5 - Conversión Lambda(p,m,n=1) a F(p,m-p+1)
> F <- (1-lambda)/lambda * (m-p+1)/p
> # 6 - p-valor
> p.val <- pf(F,p,m-p+1,lower.tail=FALSE)
> c(lambda,F,p.valor=p.val)

                p.valor
0.9433262 0.5166767 0.7621709

```

De modo que aceptamos la hipótesis de igualdad de medias.

El mismo test se puede hacer con la función `manova()`

```

> maov1 <- manova(cbind(length,wing,head,humerus,sternum) ~ alive, data=gorriones)
> summary(maov1, test="Wilks")

      Df    Wilks approx F num Df den Df Pr(>F)
alive    1 0.94333  0.51668     5    43 0.7622
Residuals 47

```

Ejercicio 17

Con los datos de los gorriones hembra del ejercicio anterior y teniendo en cuenta que el test *M* de Box es muy sensible a la no normalidad de los datos:

- a) Comparar la variabilidad de los dos grupos de datos (supervivientes y no supervivientes) con el test de Levene múltiple (no multivariante) que se puede aplicar con la función `leveneTests()` del paquete `heplots`.

La gracia de esta función es que aplica el test de Levene a cada variable con una única instrucción.

```
> leveneTests(gorrones[,1:5],alive)

Levene's Tests for Homogeneity of Variance (center = median)

      df1 df2 F value  Pr(>F)
length   1  47  1.4470 0.23503
wing     1  47  1.4030 0.24218
head     1  47  0.6638 0.41934
humerus  1  47  3.6559 0.06198 .
sternum  1  47  1.9840 0.16555
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- b) (**) Comparar la variabilidad de los dos grupos con una generalización multivariante del test de Levene llamada test de Van Valen.

Seguiremos el algoritmo paso a paso:

```
> # Estandarizamos las variables con la función scale()
> matestandar <- scale(gorrones[,1:5])
> # Calculamos el vector de medianas para cada grupo
> matsup <- matestandar[alive == "S",]
> matnosup <- matestandar[alive == "N",]
> medianas.sup <- apply(matsup, 2, median)
> medianas.nosup <- apply(matnosup, 2, median)
> # Preparamos los datos restando las medianas
> matabsdevsup <- abs(sweep(matsup,2,medianas.sup,"-"))
> matabsdevnosup <- abs(sweep(matnosup,2,medianas.nosup,"-"))
> matabsdev.all <- rbind(matabsdevsup,matabsdevnosup)
> matabsdev.all <- data.frame(alive, matabsdev.all)
> # Calculamos las distancias euclideas al origen de coordenadas
> norma.euclidea <- function(x) sqrt(sum(x^2))
> distancias <- data.frame(alive, dist=apply(matabsdev.all[,1],1,norma.euclidea))
> # Hacemos el test de comparación de los dos grupos
> dist_1 <- distancias$dist[alive=="S"]
> dist_2 <- distancias$dist[alive=="N"]
> t.test(dist_1, dist_2, var.equal=T, alt="less", data=distancias)

Error in !var.equal && nx <2: 'length = 25' in coercion to 'logical(1)'
```

En este test hemos tomado como alternativa la hipótesis unilateral de que la media de las distancias para los supervivientes es menor que la media de las distancias para los no supervivientes.

También podemos hacer un test no paramétrico:

```
> wilcox.test(dist_1,dist_2, alt="less", data=distancias)

Wilcoxon rank sum exact test

data:  dist_1 and dist_2
W = 211, p-value = 0.0479
alternative hypothesis: true location shift is less than 0

> detach(gorriones)
```

En ambos contrastes rechazamos⁹ la igualdad de la variabilidad multivariante de los dos grupos de gorriones. Por lo tanto, estos dos test indican mayor variabilidad para los no sobrevivientes que para los sobrevivientes.

⁹Gracias al alumno Javier Blecua del curso 2020-2021-1s que alertó sobre un error que había pasado desapercibido en documentos anteriores a éste.

Según Manley (pág. 66):

“Una explicación para el resultado significativo con este test, pero sin resultado significativo con la prueba de Levene, no es difícil de encontrar. La prueba de Levene no es direccional y no tiene en cuenta la expectativa de que los supervivientes, si acaso, serán menos variables que los no supervivientes. Por otro lado, la prueba de Van Valen es específicamente para una menor variación en la Muestra 1 que en la Muestra 2, para todas las variables. En este caso, todas las variables muestran menos variación en la Muestra 1 que en la Muestra 2. La prueba de Van Valen ha enfatizado este hecho, pero la prueba de Levene no.”

Ejercicio 18 Los datos de Iris de Edgar Anderson

El `data.frame` `iris` de **R** contiene los famosos datos de Iris de Fisher(1936) o de Anderson(1935) con las variables longitud del sépalo, anchura del sépalo, longitud del pétalo y anchura del pétalo, medidas en centímetros sobre tres especies de flores del género Iris: Iris setosa, Iris versicolor e Iris virginica.

```
data(iris)
help(iris)
attach(iris)
```

a) Realizar un tratamiento descriptivo de los datos.

```
> data(iris)
> attach(iris)
```

Con la función `str` tendremos una primera impresión de las variables y los datos en el `data.frame`:

```
> str(iris)

'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 ...
 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Para resumir los datos, podemos utilizar algunas funciones como `summary`:

```
> 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
```

también y sobre las variables numéricas podemos utilizar el `numSummary` del paquete `Rcmdr` o algunas otras funciones como `fivenum` o `sd`:

```
> t(apply(iris[, -5], 2, fivenum))
```

	[,1]	[,2]	[,3]	[,4]	[,5]
Sepal.Length	4.3	5.1	5.80	6.4	7.9
Sepal.Width	2.0	2.8	3.00	3.3	4.4
Petal.Length	1.0	1.6	4.35	5.1	6.9
Petal.Width	0.1	0.3	1.30	1.8	2.5

```
> apply(iris[, -5], 2, sd)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
	0.8280661	0.4358663	1.7652982	0.7622377

Con todo ello estudiaremos los datos y los valores de las medias, desviaciones, rangos, etc. para comentar dichos valores y destacar los que nos parezcan extraños o especiales.

Por cierto: una cosa imprescindible es comprobar el número de datos faltantes (missing) que tenemos:

```
> sum(is.na(iris))
```

```
[1] 0
```

A continuación podemos estudiar los valores medios disgregados en función del factor. Por ejemplo, para la variable `Sepal.Length` tenemos:

```
> tapply(iris$Sepal.Length, iris$Species, mean)
```

	setosa	versicolor	virginica
	5.006	5.936	6.588

Lo mismo podemos hacer para otras variables y con otros estadísticos como la mediana o la desviación estándar.

A continuación calculamos la matriz de varianzas-covarianzas de las variables numéricas:

```
> round(cov(iris[, -5]), 3)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	0.686	-0.042	1.274	0.516
Sepal.Width	-0.042	0.190	-0.330	-0.122
Petal.Length	1.274	-0.330	3.116	1.296
Petal.Width	0.516	-0.122	1.296	0.581

y la matriz de correlaciones:

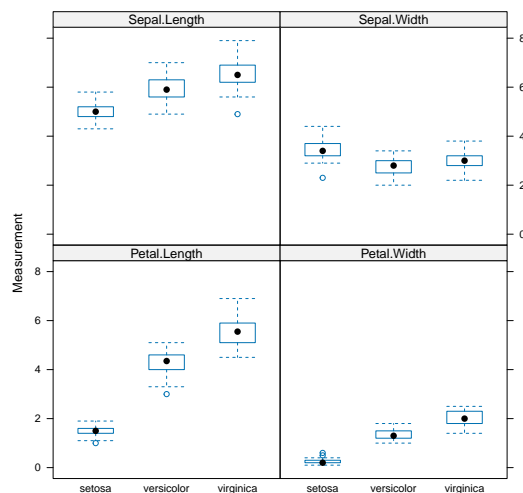
```
> round(cor(iris[, -5]), 4)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Sepal.Length	1.0000	-0.1176	0.8718	0.8179
Sepal.Width	-0.1176	1.0000	-0.4284	-0.3661
Petal.Length	0.8718	-0.4284	1.0000	0.9629
Petal.Width	0.8179	-0.3661	0.9629	1.0000

Observamos que hay correlaciones muy altas. Los signos también son importantes.

b) Dibujar un boxplot múltiple con el paquete *lattice*.

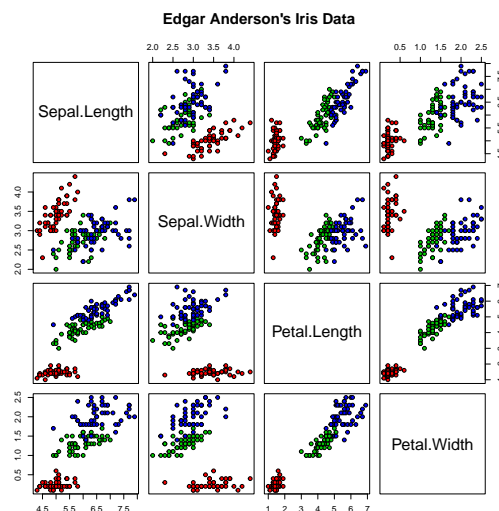
```
> library(lattice)
> reshaped.iris <- reshape(iris, varying=1:4, v.names="Measurement",
+                           timevar="Dimension", times=names(iris)[1:4],
+                           idvar="Measure ID", direction="long")
> bwplot(Measurement ~ Species | Dimension, data=reshaped.iris)
```



c) Representar las tres poblaciones conjuntamente en gráficos de dispersión para las variables numéricas dos a dos.

Para estudiar gráficamente las relaciones dos a dos de las variables numéricas podemos utilizar la función `pairs()` o alguna de sus variantes más sofisticadas:

```
> pairs(iris[1:4], main = "Edgar Anderson's Iris Data", pch = 21,
+       bg = c("red", "green3", "blue")[unclass(iris$Species)])
```

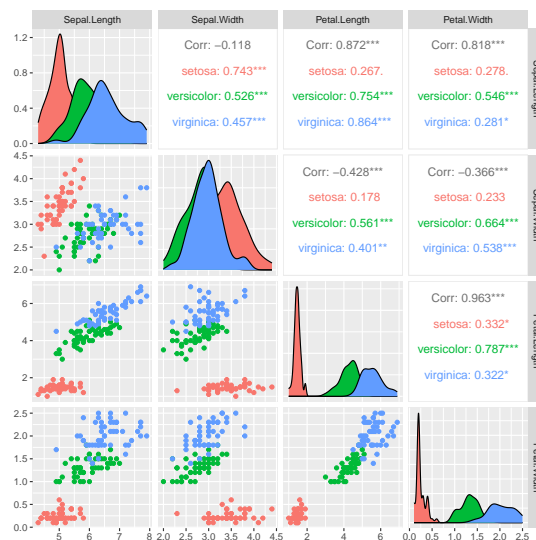


Con el paquete GGally basado en el ggplot2.

```
> library(GGally)

Loading required package: ggplot2
Registered S3 method overwritten by 'GGally':
method from
+.gg ggplot2

> pm <- ggpairs(data = iris,
+               mapping = aes(color = Species),
+               columns = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width"))
> pm
```



d) Comparar las matrices de covarianzas de las tres especies con el test de la razón de verosimilitudes.

En primer lugar calculamos el número de casos en cada nivel:

```
> table(iris$Species)

setosa versicolor virginica
50      50          50
```

Se trata de un caso balanceado con 50 réplicas por nivel.

Para utilizar el test de la razón de verosimilitudes es preciso utilizar las estimaciones máximo-verosímiles de las matrices de covarianzas:

```
> Sseto <- cov(iris[iris$Species=="setosa",-5])
> Svers <- cov(iris[iris$Species=="versicolor",-5])
> Svirg <- cov(iris[iris$Species=="virginica",-5])
> Sseto <- 49*Sseto/50
> Svers <- 49*Svers/50
> Svirg <- 49*Svirg/50
```

Así, la matriz de covarianzas común es

```
> S <- (50*Sseto + 50*Svers + 50*Svirg)/150
```

y el estadístico¹⁰

```
> llr <- 150*log(det(S)) - (50*log(det(Sseto)) + 50*log(det(Svers))
+                               + 50*log(det(Svirg)))
```

cuyo p -valor es

```
> pchisq(llr, df=(3-1)*4*5/2, lower=FALSE)
```

```
[1] 7.313689e-22
```

de modo que no podemos aceptar la igualdad de matrices de covarianzas.

Lo mismo ocurre con el test M de Box:

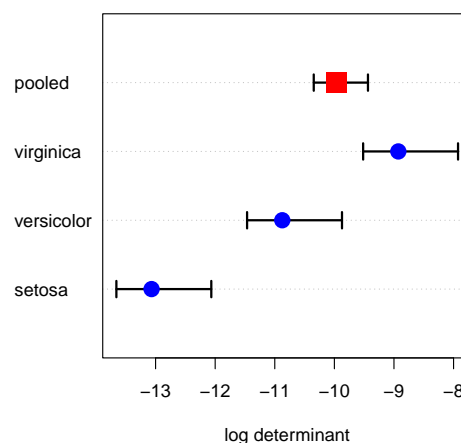
```
> M.test <- boxM(iris[,1:4],iris$Species)
> M.test
```

Box's M-test for Homogeneity of Covariance Matrices

data: iris[, 1:4]

Chi-Sq (approx.) = 140.94, df = 20, p-value < 2.2e-16

```
> plot(M.test)
```



- e) Realizar un MANOVA de un factor, la especie, y contrastar si las medias de las especies se pueden considerar estadísticamente distintas.

A pesar del resultado del apartado anterior y dada la robustez del MANOVA cuando los tamaños muestrales son parecidos, el modelo que consideramos es:

¹⁰Este estadístico puede ser sesgado y convendría aplicar la corrección de Box.

```
> Y <- as.matrix(iris[, -5])
> g <- manova( Y ~ Species, data=iris)
> summary(g)
```

	Df	Pillai	approx F	num Df	den Df	Pr(>F)
Species	2	1.1919	53.466	8	290	< 2.2e-16 ***
Residuals	147					

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El resultado del MANOVA nos indica que debemos rechazar la hipótesis nula y concluir que hay diferencias significativas entre los vectores de medias de los 3 grupos.

También se pueden utilizar otros test como el de Wilks (aunque la traza de Pillai es más robusta):

```
> summary(g, test="Wilks")
```

	Df	Wilks	approx F	num Df	den Df	Pr(>F)
Species	2	0.023439	199.15	8	288	< 2.2e-16 ***
Residuals	147					

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- f) (**) Realizar una representación canónica de las tres poblaciones con regiones confidenciales para los individuos medios de cada grupo.

Utilizar la función `candisc()` del paquete `candisc` o consultar

<http://erre-que-erre-paco.blogspot.com/2010/03/analisis-canonical-de-poblaciones.html>

Tabla 1: Depósitos de corcho en centigramos de 28 alcornoques en las cuatro direcciones cardinales.

N	E	S	W	N	E	S	W
72	66	76	77	91	79	100	75
60	53	66	63	56	68	47	50
56	57	64	58	79	65	70	61
41	29	36	38	81	80	68	58
32	32	35	36	78	55	67	60
30	35	34	26	46	38	37	38
39	39	31	27	39	35	34	37
42	43	31	25	32	30	30	32
37	40	31	25	60	50	67	54
33	29	27	36	35	37	48	39
32	30	34	28	39	36	39	31
63	45	74	63	50	34	37	40
54	46	60	52	43	37	39	50
47	51	52	43	48	54	57	43

Tabla 2: Medidas biométricas sobre unos gorriones.

Supervivientes					No supervivientes				
X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5
156	245	31.6	18.5	20.5	155	240	31.4	18.0	20.7
154	240	30.4	17.9	19.6	156	240	31.5	18.2	20.6
153	240	31.0	18.4	20.6	160	242	32.6	18.8	21.7
153	236	30.9	17.7	20.2	152	232	30.3	17.2	19.8
155	243	31.5	18.6	20.3	160	250	31.7	18.8	22.5
163	247	32.0	19.0	20.9	155	237	31.0	18.5	20.0
157	238	30.9	18.4	20.2	157	245	32.2	19.5	21.4
155	239	32.8	18.6	21.2	165	245	33.1	19.8	22.7
164	248	32.7	19.1	21.1	153	231	30.1	17.3	19.8
158	238	31.0	18.8	22.0	162	239	30.3	18.0	23.1
158	240	31.3	18.6	22.0	162	243	31.6	18.8	21.3
160	244	31.1	18.6	20.5	159	245	31.8	18.5	21.7
161	246	32.3	19.3	21.8	159	247	30.9	18.1	19.0
157	245	32.0	19.1	20.0	155	243	30.9	18.5	21.3
157	235	31.5	18.1	19.8	162	252	31.9	19.1	22.2
156	237	30.9	18.0	20.3	152	230	30.4	17.3	18.6
158	244	31.4	18.5	21.6	159	242	30.8	18.2	20.5
153	238	30.5	18.2	20.9	155	238	31.2	17.9	19.3
155	236	30.3	18.5	20.1	163	249	33.4	19.5	22.8
163	246	32.5	18.6	21.9	163	242	31.0	18.1	20.7
159	236	31.5	18.0	21.5	156	237	31.7	18.2	20.3
					159	238	31.5	18.4	20.3
					161	245	32.1	19.1	20.8
					155	235	30.7	17.7	19.6
					162	247	31.9	19.1	20.4
					153	237	30.6	18.6	20.4
					162	245	32.5	18.5	21.1
					164	248	32.3	18.8	20.9