

Normal multivariante

Francesc Carmona

Análisis multivariante

Máster universitario de Bioinformática y Bioestadística

19 de diciembre de 2020

1. Normal bivariate

1.1. Definición

Consideremos una variable aleatoria $X \sim N(0, 1)$ normal univariante estándar, es decir, la que tiene media 0 y varianza 1. Entonces la variable $Y = \mu + \sigma X$ tiene distribución

$$Y \sim N(\mu, \sigma^2)$$

donde μ es la media y σ^2 la varianza.

Del mismo modo, podemos considerar la siguiente generalización con dos variables independientes X_1, X_2 ambas $N(0, 1)$. Si fijamos unas constantes $\mu_1, \mu_2, \sigma_1 > 0, \sigma_2 > 0$ y $|\rho| < 1$, diremos que las variables transformadas

$$Y_1 = \mu_1 + \sigma_1 X_1$$

$$Y_2 = \mu_2 + \sigma_2(\rho X_1 + \sqrt{1 - \rho^2} X_2)$$

siguen una distribución conjunta normal bivalente $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ y matriz de varianzas-covarianzas

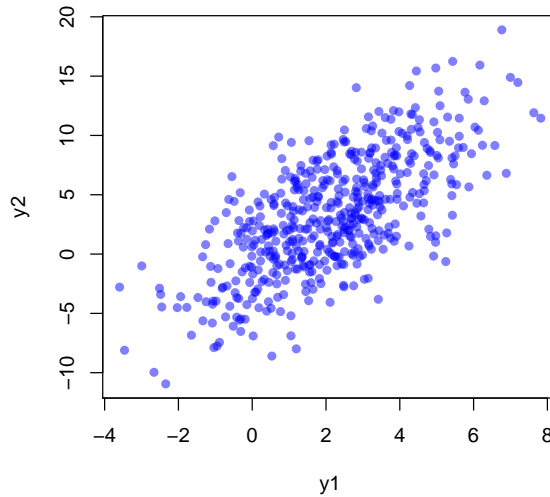
$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}$$

ya que gracias a las propiedades de la esperanza y de la varianza se puede demostrar que $\mathbf{y} = (Y_1, Y_2)'$ tiene como parámetros

$$E(Y_1) = \mu_1, E(Y_2) = \mu_2, \text{var}(Y_1) = \sigma_1^2, \text{var}(Y_2) = \sigma_2^2, \text{cov}(Y_1, Y_2) = \sigma_1 \sigma_2 \rho, \text{cor}(Y_1, Y_2) = \rho$$

Ejemplo 1.1 *Vamos a realizar una simulación con R.*

```
> set.seed(321)
> x1 <- rnorm(500)
> x2 <- rnorm(500)
> mu1 <- 2
> mu2 <- 3
> sigma1 <- 2
> sigma2 <- 5
> rho <- 0.7
> # con estos datos calculamos y1 e y2
> y1 <- mu1 + sigma1 * x1
> y2 <- mu2 + sigma2 * (rho*x1 + sqrt(1-rho^2) * x2)
> plot(y1, y2, col = rgb(red = 0, green = 0, blue = 1, alpha = 0.5),
+      pch = 16)
```



Observamos una cierta correlación entre los datos y una mayor dispersión para el eje Y_2 que para el Y_1 . ¿Es esta la forma que adoptan las muestras en una normal bivalente? Para contestar la pregunta habrá que definir la densidad teórica de la normal bivalente y compararla con estos datos.

La función de densidad conjunta de $\mathbf{y} = (Y_1, Y_2)'$ es

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}Q(y_1, y_2)\right)$$

donde

$$Q(y_1, y_2) = (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ = \frac{1}{1-\rho^2} \left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(y_1 - \mu_1)(y_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} \right)$$

La representación gráfica de la densidad es la llamada superficie de Gauss que tiene su máximo en el punto $\boldsymbol{\mu} = (\mu_1, \mu_2)'$.

Ejemplo 1.2 Vamos a dibujar la densidad de una normal bivalente con los mismos parámetros que en el ejemplo anterior.

```
> x1 <- seq(mu1-4*sigma1, mu1+4*sigma1, length=51)
> x2 <- seq(mu2-4*sigma2, mu2+4*sigma2, length=51)
> f <- function(x1,x2){
+   Q <- ((x1-mu1)^2/sigma1^2
+         - 2*rho*(x1-mu1)*(x2-mu2)/(sigma1*sigma2)
+         + (x2-mu2)^2/sigma2^2)/(1-rho^2)
+   1/(2*pi*sigma1*sigma2*sqrt(1-rho^2)) * exp(-Q/2)
+ }
> z <- outer(x1,x2,f) # calculamos los valores de la densidad
> persp(x1, x2, z,
+   main="Distribución normal bivalente",
+   col="lightgreen",
+   theta=30, phi=20,
```

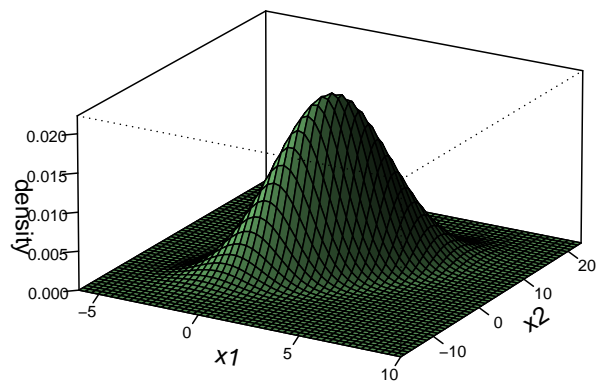
```

+ r=50,
+ d=0.1,
+ expand=0.5,
+ ltheta=90, lphi=180,
+ shade=0.75,
+ ticktype="detailed",
+ cex.axis=0.7,
+ zlab="density")
> # podemos añadir una línea de texto al gráfico
> mtext(expression(list(mu[1]==2,mu[2]==3,sigma[1]==2,sigma[2]==5,
+ rho==0.7)), side=3)

```

Distribución normal bivalente

$$\mu_1 = 2, \mu_2 = 3, \sigma_1 = 2, \sigma_2 = 5, \rho = 0.7$$



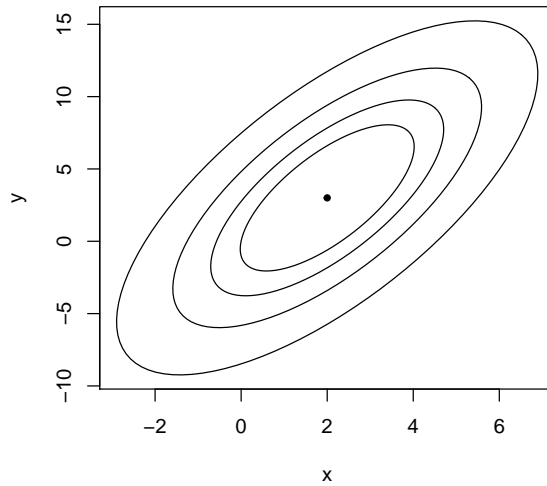
De este modo no podemos ver si los datos generados en el ejemplo 1.1 se ajustan a la densidad teórica que vemos en este gráfico.

Ahora bien, resulta que la relación $Q(y_1, y_2) = c^2$, donde c^2 es una constante positiva, define elipses concéntricas y homotéticas en el plano \mathbb{R}^2 . La función de densidad es constante sobre tales elipses, constituyendo las llamadas curvas de equiprobabilidad.

```

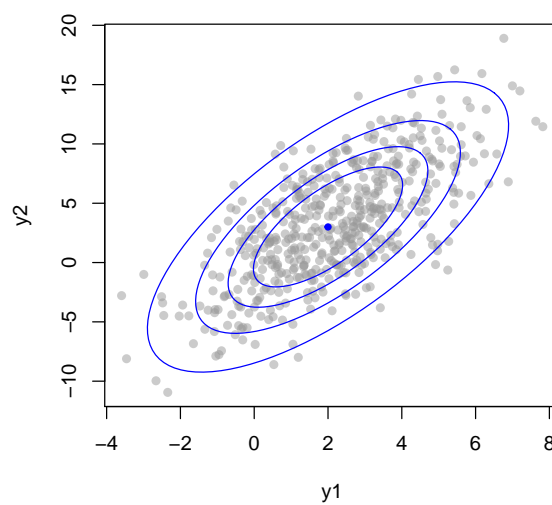
> library(ellipse)
> Sigma <- matrix(c(sigma1^2, sigma1*sigma2*rho,
+                   sigma1*sigma2*rho, sigma2^2), ncol=2)
> ee95 <- ellipse(Sigma, centre = c(2, 3), level = 0.95)
> ee80 <- ellipse(Sigma, centre = c(2, 3), level = 0.80)
> ee60 <- ellipse(Sigma, centre = c(2, 3), level = 0.60)
> ee40 <- ellipse(Sigma, centre = c(2, 3), level = 0.40)
> plot(ee95, type = "l")
> lines(ee80)
> lines(ee60)
> lines(ee40)
> points(2,3, pch=20)

```



Ejemplo 1.3 Podemos superponer las elipses de equiprobabilidad a los datos representados en el diagrama de dispersión.

```
> plot(y1, y2, col = rgb(red = 0.6, green = 0.6, blue = 0.6, alpha = 0.5),
+       pch = 16)
> lines(ee95, type = "l", col="blue")
> lines(ee80, col="blue")
> lines(ee60, col="blue")
> lines(ee40, col="blue")
> points(2,3, pch=20, col="blue")
```

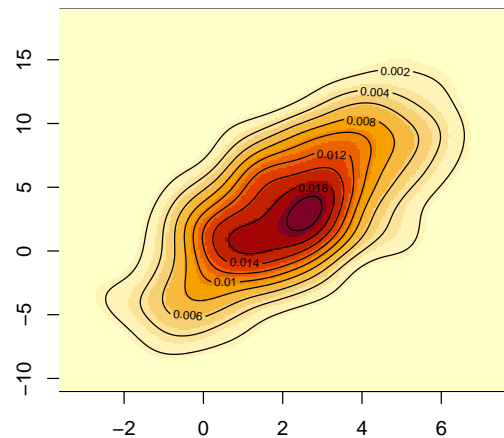


A simple vista, el ajuste es muy notable. Para una mayor precisión habrá que estudiar más propiedades de la normal bivariante o multivariante y otras distribuciones asociadas.

También podemos comprobar que las curvas de nivel de estos datos tienen la forma de las elipses de

equiprobabilidad. Para ello disponemos de la función `kde2d()` del paquete *MASS* que permite estimar la llamada **densidad kernel**.

```
> library(MASS)
> bivn.kde <- kde2d(y1, y2, n=200)
> image(bivn.kde)
> contour(bivn.kde, add=T)
```



1.2. Propiedades de la normal bivalente

1. La primera propiedad que podemos destacar es que las distribuciones marginales Y_1 e Y_2 son también normales univariantes con $Y_1 \sim N(\mu_1, \sigma_1^2)$ y $Y_2 \sim N(\mu_2, \sigma_2^2)$.

Para demostrarlo hay que calcular las integrales:

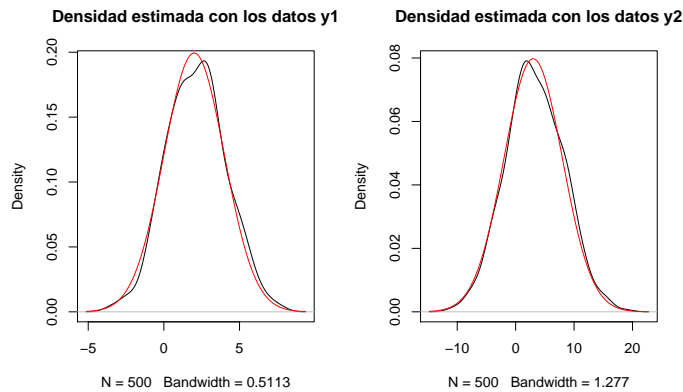
$$f(y_1) = \int_{-\infty}^{+\infty} f(y_1, y_2) dy_2$$

$$f(y_2) = \int_{-\infty}^{+\infty} f(y_1, y_2) dy_1$$

Sin embargo, en el siguiente ejemplo podemos ver una aproximación con los datos simulados del ejemplo 1.1.

Ejemplo 1.4 Vamos a dibujar las funciones de densidad aproximadas de los datos y_1 e y_2 del ejemplo 1.1. Además añadiremos la densidad exacta (en rojo) que deben tener esas variables con los parámetros aportados.

```
> oldpar <- par(mfrow=c(1,2))
> plot(density(y1), main="Densidad estimada con los datos y1")
> curve(dnorm(x, mean = mu1, sd = sigma1), add=T, col="red")
> plot(density(y2), main="Densidad estimada con los datos y2")
> curve(dnorm(x, mean = mu2, sd = sigma2), add=T, col="red")
> par(oldpar)
```



El ajuste no deja lugar a dudas.

2. El parámetro ρ es el coeficiente de correlación entre Y_1 e Y_2 .

$$\text{cor}(Y_1, Y_2) = \frac{\text{cov}(Y_1, Y_2)}{\sqrt{\text{var}(Y_1)\text{var}(Y_2)}} = \frac{\sigma_1\sigma_2\rho}{\sqrt{\sigma_1^2\sigma_2^2}} = \rho$$

Ejemplo 1.5 Veamos que la correlación de los datos se aproxima a la correlación teórica poblacional que hemos fijado para ellos $\rho = 0.7$.

```
> cor(y1,y2)
[1] 0.6896275
```

3. Para esta distribución incorrelación implica independencia estocástica. El recíproco siempre es cierto.

En efecto: si $\rho = 0$, es fácil ver que

$$f(y_1, y_2) = f(y_1) \cdot f(y_2)$$

donde $f(y_i)$ es la densidad marginal de Y_i , $i = 1, 2$.

Ejemplo 1.6 En este ejemplo vamos a generar unos datos con $\rho = 0$.

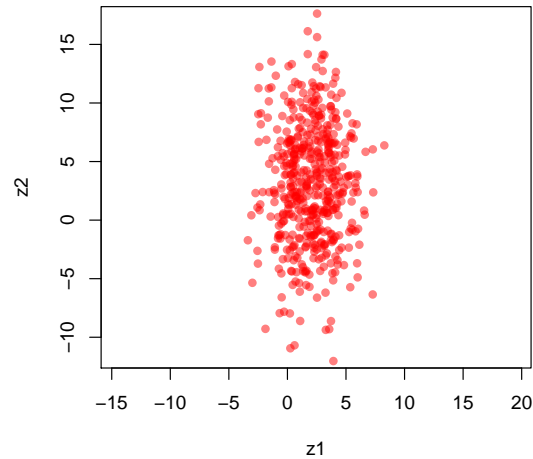
```
> z1 <- rnorm(n=500, mean=2, sd=2)
> z2 <- rnorm(n=500, mean=3, sd=5)
```

Si calculamos por separado dos muestras independientes, la distribución conjunta tiene correlación $\rho = 0$.

```
> cor(z1,z2)
[1] 0.005561068
```

Para dibujar los puntos utilizaremos la función `eqsplot()` del paquete *MASS* de modo que los ejes tengan la misma escala en los dos ejes.

```
> eqscplot(z1, z2, col = rgb(red = 1, green = 0, blue = 0, alpha = 0.5),
+          pch = 16)
```



Se observa que la nube de puntos es vertical ya que $\rho = 0$ y la varianza de la variable Y_2 es mayor que la de Y_1 . En caso contrario, la nube sería horizontal. Cuando ambas varianzas son iguales y $\rho = 0$, la nube es circular, es decir, las elipses de equiprobabilidad son circunferencias.

4. La distribución de Y_2 condicionada a $Y_1 = y_1$ es también normal univariante con esperanza

$$E(Y_2|Y_1 = y_1) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)$$

y varianza

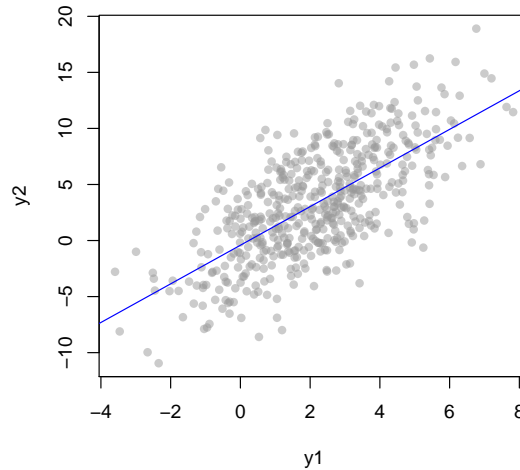
$$\text{var}(Y_2|Y_1 = y_1) = \sigma_2^2(1 - \rho^2)$$

Esto implica que la curva de regresión de la media $y_2 = E(Y_2|Y_1 = y_1)$ es la recta de regresión

$$y_2 - \mu_2 = \rho \frac{\sigma_2}{\sigma_1}(y_1 - \mu_1)$$

Ejemplo 1.7 Para los datos del ejemplo 1.1, podemos calcular la recta de regresión así

```
> fit <- lm(y2 ~ y1)
> plot(y1, y2, col = rgb(red = 0.6, green = 0.6, blue = 0.6, alpha = 0.5),
+      pch = 16)
> abline(fit, col="blue")
```



La pendiente de esta recta es

```
> coef(fit)[2]

      y1
1.722517
```

parecida al valor teórico

```
> rho * sigma2 / sigma1

[1] 1.75
```

Esta recta es la “mejor” curva de regresión sobre estos datos, ya que su distribución es normal bivalente.

2. Normal multivariante

2.1. Definición

Esta es la más importante de las distribuciones multivariantes, es decir, las distribuciones con varias variables conjuntamente. Es una generalización k -dimensional de la normal univariante y la normal bivalente que hemos visto en el apartado anterior. Se asimilan a esta distribución conjuntos de variables tales como medidas biométricas, variables socioeconómicas, resultados cuantitativos de la eficacia de fármacos, etc. Para que los datos procedan de esta distribución es necesario que cada variable sea normal y que el conjunto de las variables tengan una buena unidad experimental, de modo que sea apropiado estudiarlas conjuntamente.

Diremos que las variables aleatorias $\mathbf{y} = (Y_1, Y_2, \dots, Y_k)'$ siguen la distribución normal multivariante, si existen k variables aleatorias independientes $\mathbf{x} = (X_1, X_2, \dots, X_k)'$ con distribución $N(0, 1)$ y se verifica

$$Y_i = \mu_i + a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ik}X_k \quad i = 1, 2, \dots, k$$

donde $\mu_1, \mu_2, \dots, \mu_k$ son constantes y $\mathbf{A} = (a_{ij})$ es una matriz cuadrada $k \times k$ no singular. En notación matricial podemos escribir

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}$$

donde $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_k)'$ y $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$ es simétrica y definida positiva.

Por razones que seguidamente justificaremos, esta distribución se designa en la forma $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y depende únicamente de los k parámetros de $\boldsymbol{\mu}$ y de los $k(k+1)/2$ parámetros distintos de $\boldsymbol{\Sigma}$.

Ejemplo 2.1 Veamos que en la definición de normal bivalente que hemos hecho en la sección anterior verifica la propiedad $\mathbf{A}\mathbf{A}' = \boldsymbol{\Sigma}$.

En este caso, la matriz \mathbf{A} es

$$\mathbf{A} = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2 \rho & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix}$$

de modo que

$$\mathbf{A}\mathbf{A}' = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2 \rho & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix} \begin{pmatrix} \sigma_1 & \sigma_2 \rho \\ 0 & \sigma_2 \sqrt{1 - \rho^2} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix} = \boldsymbol{\Sigma}$$

Ejemplo 2.2 A partir de los parámetros de la normal bivalente definidos en el ejemplo 1.1 $\sigma_1 = 2$, $\sigma_2 = 5$ y $\rho = 0.7$ se puede calcular la matriz \mathbf{A} gracias a la diagonalización de la matriz $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}'$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2^2 & 2 \cdot 5 \cdot 0.7 \\ 2 \cdot 5 \cdot 0.7 & 5^2 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 7 & 25 \end{pmatrix}$$

de forma que

$$\mathbf{A} = \mathbf{V}\mathbf{D}^{1/2}$$

```
> Sigma <- matrix(c(sigma1^2, sigma1*sigma2*rho,
+                   sigma1*sigma2*rho, sigma2^2), ncol=2)
> eigenS <- eigen(Sigma)
> A <- eigenS$vectors %*% diag(sqrt(eigenS$values))
> A

      [,1]      [,2]
[1,] 1.509089 -1.3124976
[2,] 4.984183  0.3973923

> A %*% t(A)

      [,1] [,2]
[1,]    4    7
[2,]    7   25
```

2.2. Propiedades

2.2.1. Vector de medias y matriz de covarianzas

En una normal multivariante $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ el vector de medias es $\boldsymbol{\mu}$ y la matriz de varianzas-covarianzas es $\boldsymbol{\Sigma}$.

En efecto, como $E(X_i) = \mu_i$, $i = 1, \dots, k$, la esperanza de Y_i es

$$\begin{aligned} E(Y_i) &= E(\mu_i + a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ik}X_k) \\ &= \mu_i + a_{i1}E(X_1) + a_{i2}E(X_2) + \dots + a_{ik}E(X_k) = \mu_i \end{aligned}$$

En cuanto a las covarianzas, como $E(X_i \cdot X_j) = 0$, $i \neq j$, y $E(X_i^2) = 1$, tenemos

$$\begin{aligned}\sigma_{ij} &= E[(Y_i - \mu_i)(Y_j - \mu_j)] = E \left[\left(\sum_h a_{ih} X_h \right) \left(\sum_g a_{jg} X_g \right) \right] \\ &= E \left(\sum_h \sum_g a_{ih} a_{jg} X_h X_g \right) = \sum_{h=1}^k a_{ih} a_{jh}\end{aligned}$$

relación que equivale al producto matricial

$$\Sigma = \mathbf{A} \mathbf{A}'$$

2.2.2. Función de densidad

La función de densidad conjunta de Y_1, Y_2, \dots, Y_k es

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right)$$

donde $\mathbf{y} = (y_1, y_2, \dots, y_k)'$.

La demostración se basa en que la distribución conjunta de las variables independientes X_1, X_2, \dots, X_k $N(0, 1)$ tiene densidad multivariante

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} \exp \left(-\frac{1}{2} \mathbf{x}' \mathbf{x} \right)$$

y en que las Y_1, Y_2, \dots, Y_k son una combinación lineal de éstas (ver Cuadras[2] pág. 30).

2.2.3. Marginales

Las distribuciones marginales de todas las Y_i , $i = 1, \dots, k$ son $N(\mu_i, \sigma_{ii})$, donde σ_{ii} son los elementos de la diagonal de la matriz Σ .

La normalidad univariante de cada una de las Y_i no es condición suficiente de normalidad multivariante. Se pueden construir distribuciones multivariantes no normales con marginales normales.

2.2.4. Combinación lineal

Toda combinación lineal del tipo

$$Z = c_1 Y_1 + c_2 Y_2 + \dots + c_k Y_k$$

es normal univariante.

2.2.5. Transformación lineal

Si \mathbf{y} es $N_k(\boldsymbol{\mu}, \Sigma)$ y \mathbf{B} es una matriz $k \times k$ y \mathbf{v} es un vector de k constantes, entonces

$$\mathbf{z} = \mathbf{B} \mathbf{y} + \mathbf{v}$$

es normal multivariante $N_k(\mathbf{B} \boldsymbol{\mu} + \mathbf{v}, \mathbf{B}' \Sigma \mathbf{B})$.

Cualquier subconjunto formado por $m < k$ variables entre las Y_1, \dots, Y_k es también normal multivariante.

2.2.6. Incorrelación

Incorrelación implica independencia estocástica, es decir, si $\Sigma = \mathbf{I}$, las variables Y_1, \dots, Y_k son estocásticamente independientes.

Basta con substituir Σ con \mathbf{I} en la función de densidad conjunta para ver que es igual al producto de las marginales.

2.2.7. Transformación lineal ortogonal

Si el vector de variables $\mathbf{y} = (Y_1, \dots, Y_k)'$ es $N_k(\boldsymbol{\mu}, \mathbf{I})$, y por lo tanto independientes, y $\mathbf{z} = \mathbf{T}\mathbf{y}$ es una transformación lineal definida por una matriz ortogonal \mathbf{T} (es decir, $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$), entonces las variables $\mathbf{z} = (Z_1, \dots, Z_k)$ son también independientes.

En efecto, por la propiedad 2.2.5 la distribución de \mathbf{z} es normal multivariante con matriz de covarianzas $\mathbf{T}\mathbf{T}' = \mathbf{I}$.

2.2.8. Distribución ji-cuadrado

Si \mathbf{x} es $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad \text{tiene distribución } \chi_k^2$$

Si aplicamos la transformación $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$, por la propiedad 2.2.5, \mathbf{z} es $N(\mathbf{0}, \mathbf{I})$, luego

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{z}'\mathbf{z} = Z_1^2 + \dots + Z_k^2$$

es la suma de cuadrados de k variables independientes $N(0, 1)$, es decir, una ji-cuadrado son k grados de libertad.

Nota: Cuando \mathbf{x} es una observación concreta, la expresión $D^2 = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ se conoce como la distancia de Mahalanobis (al cuadrado) entre dicha observación y la media teórica. En la práctica, la distancia de Mahalanobis se calcula sustituyendo los parámetros $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ por sus estimaciones.

3. Distribución de Wishart

3.1. Introducción

La distribución de Wishart es la distribución que siguen ciertas matrices simétricas que se utilizan en muchos análisis multivariantes. Su uso es similar al que hacemos con la distribución ji-cuadrado en el caso univariante. Recordemos su utilidad.

Sea $\mathbf{z} = (z_1, \dots, z_m)'$ un vector con m observaciones independientes con distribución $N(0, 1)$. La distribución de

$$\mathbf{z}'\mathbf{z} = \sum_{i=1}^m z_i^2 \quad \text{es una } \chi_m^2$$

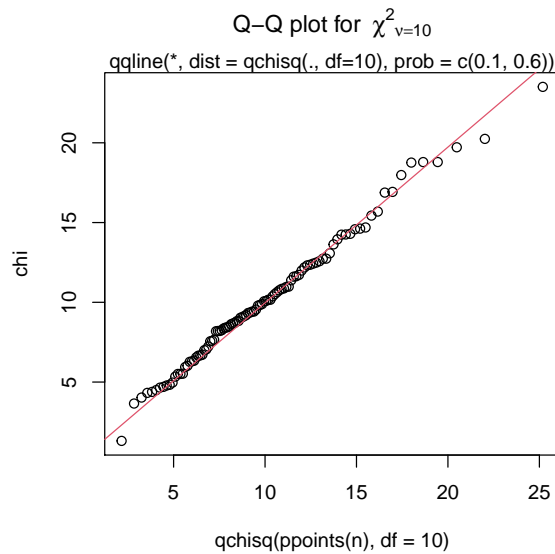
Ejemplo 3.1 Aunque **R** dispone de una función `rchisq()` para generar una muestra aleatoria de valores según la distribución ji-cuadrado, vamos a simular la creación de una muestra a partir de su definición.

```
> m <- 10
> n <- 100
> z <- matrix(rnorm(n*m), ncol=10)
```

La matriz \mathbf{z} contiene $n = 100$ observaciones de $m = 10$ muestras $N(0, 1)$ independientes.

Si elevamos al cuadrado cada elemento y sumamos las filas de la matriz \mathbf{z} el resultado debe ser una muestra de 100 valores según una χ_{10}^2 .

```
> chi <- apply(z^2, 1, sum)
> ## Q-Q plot for Chi^2 data against true theoretical distribution:
> qqplot(qchisq(ppoints(n), df = 10), chi,
+        main = expression("Q-Q plot for" ~~ {chi^2}[nu == 10]))
> qqline(chi, distribution = function(p) qchisq(p, df = 10),
+        prob = c(0.1, 0.6), col = 2)
> mtext("qqline(*, dist = qchisq(., df=10), prob = c(0.1, 0.6))")
```



Ciertamente, el ajuste en el Q-Q plot es notable.

Un resultado clásico de muestreo con variables normales es el siguiente:

Sea x_1, x_2, \dots, x_n una muestra aleatoria simple de una variable $N(\mu, \sigma^2)$ y consideremos la media y la varianza muestrales

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

entonces \bar{x} y s^2 son estocásticamente independientes y

$$\frac{ns^2}{\sigma^2} \text{ sigue la distribución } \chi^2_{n-1}$$

Ejemplo 3.2 Veamos que esta propiedad permite hallar un intervalo de confianza para la varianza σ^2 de una población normal dada una muestra aleatoria simple.

Consideremos una muestra de $n = 30$ observaciones de una población normal $N(5, 4)$ de modo que sabemos que $\sigma^2 = 4$, pero vamos a pensar que ese parámetro fuera desconocido. El intervalo de confianza con un nivel $(1 - \alpha) \cdot 100\%$ para σ^2 (que suponemos desconocido) es

$$\frac{ns^2}{b} < \sigma^2 < \frac{ns^2}{a}$$

donde $P(\chi^2 < a) = \alpha/2$ y $P(\chi^2 > b) = \alpha/2$.

```
> n <- 30
> x <- rnorm(n, mean=5, sd=2)
> alpha <- 0.05
> a <- qchisq(alpha/2, df=n-1)
> b <- qchisq(1-alpha/2, df=n-1)
> c((n-1)*var(x)/b, (n-1)*var(x)/a)

[1] 2.572129 7.328666
```

3.2. Definición

La generalización multivariante de la distribución ji-cuadrado consiste en definir una matriz

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{pmatrix}$$

cuyas filas siguen la distribución normal multivariante $N_m(\mathbf{0}, \mathbf{\Sigma})$ y son estocásticamente independientes. Se llama entonces distribución de Wishart $W_m(\mathbf{\Sigma}, n)$, para m variables y de parámetros $\mathbf{\Sigma}$ y grados de libertad n , la que sigue la matriz simétrica $m \times m$

$$\mathbf{W} = \mathbf{Z}'\mathbf{Z}$$

Esta distribución fue introducida por Wishart en 1928 y juega un papel fundamental en análisis multivariante, especialmente, en MANOVA.

La función de densidad de esta distribución tiene una expresión bastante compleja que puede verse en Cuadras[2] pág. 33.

Cuando $m = 1$, la distribución de Wishart $W_1(\sigma^2, n)$ equivale a una ji-cuadrado en la forma $\sigma^2 \chi_n^2$.

3.3. Propiedades

Algunas de las propiedades de la distribución de Wishart son:

3.3.1. Esperanza y varianza

Se puede demostrar (ver Mardia[3] pág. 66) que la esperanza de \mathbf{W} es

$$E(\mathbf{W}) = n\mathbf{\Sigma}$$

y la varianza de sus componentes

$$\text{var}(w_{ij}) = n(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj})$$

Ejemplo 3.3 En **R** tenemos la función `rWishart()` que permite obtener muestras aleatorias según una distribución Wishart de parámetros conocidos. Supongamos que tenemos una matriz de covarianzas $\mathbf{\Sigma}$ de una población normal multivariante como

```
> (Sigma <- toeplitz((5:1)/5))
```

```
      [,1] [,2] [,3] [,4] [,5]
[1,]  1.0  0.8  0.6  0.4  0.2
[2,]  0.8  1.0  0.8  0.6  0.4
[3,]  0.6  0.8  1.0  0.8  0.6
[4,]  0.4  0.6  0.8  1.0  0.8
[5,]  0.2  0.4  0.6  0.8  1.0
```

Calculamos ahora una muestra de $N = 1000$ de una distribución $W_5(\mathbf{\Sigma}, 20)$

```
> R <- rWishart(n = 1000, df = 20, Sigma = Sigma)
> dim(R)

[1]      5      5 1000
```

Observemos que R es un array que contiene 1000 matrices según una $W_5(\mathbf{\Sigma}, 20)$. Ahora calculamos la media:

```
> (mR <- apply(R, 1:2, mean))
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	19.890886	15.905002	11.96245	7.775967	3.838337
[2,]	15.905002	20.026420	15.86321	11.642863	7.658348
[3,]	11.962446	15.863212	19.85494	15.703408	11.710116
[4,]	7.775967	11.642863	15.70341	19.614630	15.710259
[5,]	3.838337	7.658348	11.71012	15.710259	19.820339

El resultado debe ser una aproximación al valor teórico 20Σ

```
> 20*Sigma
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	20	16	12	8	4
[2,]	16	20	16	12	8
[3,]	12	16	20	16	12
[4,]	8	12	16	20	16
[5,]	4	8	12	16	20

3.3.2. Suma

Si \mathbf{W}_1 es $W_m(\Sigma, n_1)$ y \mathbf{W}_2 es $W_m(\Sigma, n_2)$ y son estocásticamente independientes, entonces

$$\mathbf{W}_1 + \mathbf{W}_2 \text{ es } W_m(\Sigma, n_1 + n_2)$$

3.3.3. Matriz de covarianzas muestral de una normal

La distribución Wishart nació asociada a la distribución muestral del estimador de máxima verosimilitud de la matriz de covarianzas de una distribución normal multivariante.

Sea \mathbf{S} la matriz de varianzas-covarianzas muestrales para una muestra de tamaño n de una distribución $N_m(\boldsymbol{\mu}, \Sigma)$, entonces el vector de medias muestrales $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_m)'$ y la matriz \mathbf{S} son estocásticamente independientes y

$$n\mathbf{S} \text{ sigue la distribución } W_m(\Sigma, n-1)$$

3.3.4. Transformación

Si \mathbf{W} es $W_m(\Sigma, n)$ y \mathbf{T} es una matriz $m \times q$, entonces

$$\mathbf{T}'\mathbf{W}\mathbf{T} \text{ es } W_q(\mathbf{T}'\Sigma\mathbf{T}, n)$$

Además, si \mathbf{W} es $W_m(\Sigma, n)$ y $\mathbf{a} = (a_1, \dots, a_m)$ es un vector de constantes tal que $\mathbf{a}'\Sigma\mathbf{a} \neq 0$, entonces

$$\frac{\mathbf{a}'\mathbf{W}\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} \text{ es } \chi_n^2$$

Esta última propiedad es consecuencia de la anterior, ya que $\mathbf{a}'\mathbf{W}\mathbf{a}$ es $W_1(\mathbf{a}'\Sigma\mathbf{a}, n)$.

4. Distribución T^2 de Hotelling

4.1. Introducción

Vamos a estudiar la distribución asociada a funciones del tipo $\mathbf{d}'\mathbf{M}^{-1}\mathbf{d}$, donde \mathbf{d} es normal, \mathbf{M} es Wishart y \mathbf{d} y \mathbf{M} son independientes. Por ejemplo, \mathbf{d} puede ser un vector de medias y \mathbf{M} proporcional a la matriz de covarianzas muestral. Este es el caso más importante que da sentido a esta distribución como generalización de la t de Student univariante.

4.2. Definición

El estudio de las distribuciones asociadas a formas cuadráticas como la anterior se inició con Hotelling (1931).

Si \mathbf{d} y \mathbf{M} son independientes, $N_p(\mathbf{0}, \mathbf{I})$ y $W_p(\mathbf{I}, m)$ respectivamente, entonces

$$m\mathbf{d}'\mathbf{M}^{-1}\mathbf{d} \text{ tiene distribución } T^2 \text{ de Hotelling}$$

con parámetros p y m que escribiremos $T^2(p, m)$.

Del mismo modo, si \mathbf{x} y \mathbf{M} son independientes, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y $W_p(\boldsymbol{\Sigma}, m)$ respectivamente, entonces

$$m(\mathbf{x} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \sim T^2(p, m)$$

Finalmente, si $\bar{\mathbf{x}}$ y \mathbf{S} son el vector de medias y la matriz de covarianzas de una muestra de $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ de tamaño n y $\hat{\mathbf{S}} = (n/(n-1))\mathbf{S}$, entonces

$$(n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) = n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\hat{\mathbf{S}}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim T^2(p, n-1)$$

Este resultado es consecuencia del anterior si tomamos $\mathbf{M} = n\mathbf{S}$, $m = n-1$ y $\mathbf{x} - \boldsymbol{\mu}$ por $n^{1/2}\bar{\mathbf{x}} - \boldsymbol{\mu}$.

4.3. Propiedades

4.3.1. Invariancia

El estadístico T^2 es invariante para cualquier transformación lineal del tipo $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x} + \mathbf{b}$.

4.3.2. Relación con la distribución F

Se puede demostrar que cuando $p = 1$, la t de Student al cuadrado es $t_m^2 = T^2(1, m)$.

Del mismo modo

$$T^2(p, m) = \frac{mp}{m-p+1} F_{p, m-p+1}$$

4.3.3. Muestra de una normal

Si $\bar{\mathbf{x}}$ y \mathbf{S} son el vector de medias y la matriz de covarianzas de una muestra de $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ de tamaño n , entonces

$$\frac{n-p}{p}(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim F_{p, n-p}$$

Ejemplo 4.1 Hacemos una simulación con muestras de tamaño $n = 30$ para normales $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ donde $\boldsymbol{\mu} = (2, 3)'$ y $\boldsymbol{\Sigma}$ es la matriz de covarianzas del ejemplo 1.1. La simulación consiste en calcular $N = 200$ valores del estadístico y comprobar que sigue una distribución $F_{2, 30-2}$.

La función `mvrnorm()` del paquete *MASS* nos ayudará a calcular las $n = 30$ muestras en cada iteración.

```
> library(MASS)
> N <- 200
> n <- 30
> p <- 2
> mu <- c(2,3)
> Sigma <- matrix(c(4,7,7,25), ncol=2)
> F.est <- rep(0,N)
> for(i in 1:N){
+ X <- mvrnorm(n, mu, Sigma)
+ x.bar <- colMeans(X)
```

```

+ S <- ((n-1)/n)*cov(X)
+ S.inv <- solve(S)
+ F.est[i] <- ((n-p)/p)*t(x.bar-mu) %*% S.inv %*% (x.bar-mu)
+ }

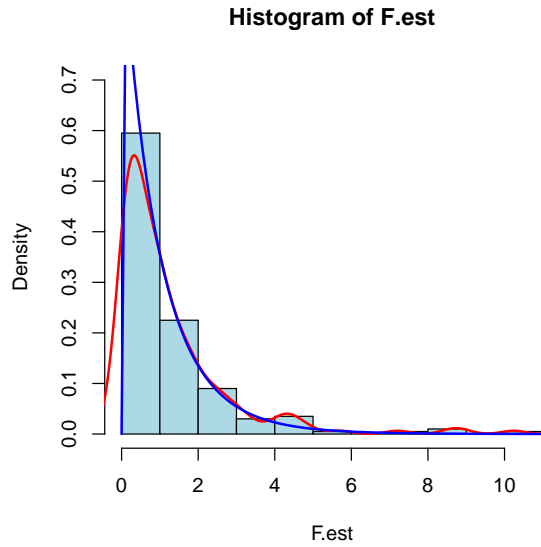
```

Ahora veamos si se ajusta a la densidad de una F de Fisher (en azul):

```

> hist(F.est, freq = FALSE, col = "lightblue", ylim=c(0,0.7))
> lines(density(F.est), col="red", lwd=2)
> curve(df(x, df1 = p, df2 = n-p), col="blue", lwd=2, add=T)

```



4.3.4. Distancia de Mahalanobis

La distancia de Mahalanobis (al cuadrado) entre dos poblaciones con medias μ_1 y μ_2 con matriz de covarianzas común Σ es

$$\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$$

Por supuesto que difícilmente conoceremos los parámetros de las poblaciones, de modo que los estimaremos con sus correspondientes valores muestrales. Supongamos que los tamaños muestrales son n_1 y n_2 respectivamente y $n = n_1 + n_2$. La distancia de Mahalanobis muestral (al cuadrado) es

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_c^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

donde $\mathbf{S}_c = (n_1 \mathbf{S}_1 + n_2 \mathbf{S}_2) / (n - 2)$ es un estimador insesgado de Σ .

Bajo unas determinadas condiciones (hipótesis nula), la distribución de D^2 es una T^2 de Hotelling. Si \mathbf{X}_1 y \mathbf{X}_2 son dos matrices de datos independientes y si las n_i filas de \mathbf{X}_i son iid $N_p(\mu_i, \Sigma_i)$, $i = 1, 2$, entonces cuando $\mu_1 = \mu_2$ y $\Sigma_1 = \Sigma_2$,

$$\frac{n_1 n_2}{n} D^2 \sim T^2(p, n - 2)$$

El estadístico

$$\frac{n_1 n_2}{n} D^2 = \frac{n_1 n_2}{n} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_c^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

es el llamado *estadístico T^2 de Hotelling para dos muestras*. Como la distribución está relacionada con la F de Fisher, resulta

$$\frac{n_1 n_2 (n - p - 1)}{n(n - 2)p} D^2 \sim F_{p, n-p-1}$$

5. Distribución lambda de Wilks

Esta distribución se puede considerar una versión multivariante de la distribución F de Fisher-Snedecor y juega un papel similar en el contraste de hipótesis.

5.1. Definición

Supongamos que la matriz \mathbf{A} es $W_p(\mathbf{I}, m)$ y la matriz \mathbf{B} es $W_p(\mathbf{I}, n)$ y son independientes y $m \geq p$, entonces se dice que

$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A} + \mathbf{B}|} = |\mathbf{I} + \mathbf{A}^{-1}\mathbf{B}|^{-1} \sim \Lambda(p, m, n)$$

tiene una distribución lambda de Wilks con parámetros p , m y n .

La familia de distribuciones lambda de Wilks aparece frecuentemente en el contexto de los contrastes de la razón de verosimilitudes. El parámetro m representa los grados de libertad del “error” y n los grados de libertad de la “hipótesis”. Luego la suma $m + n$ representa los grados de libertad del “total”.

Ejemplo 5.1 Consideremos una muestra de $N = 1000$ matrices $\mathbf{A} \sim W_3(\mathbf{I}, 30)$ y $k = 1000$ matrices $\mathbf{B} \sim W_3(\mathbf{I}, 4)$

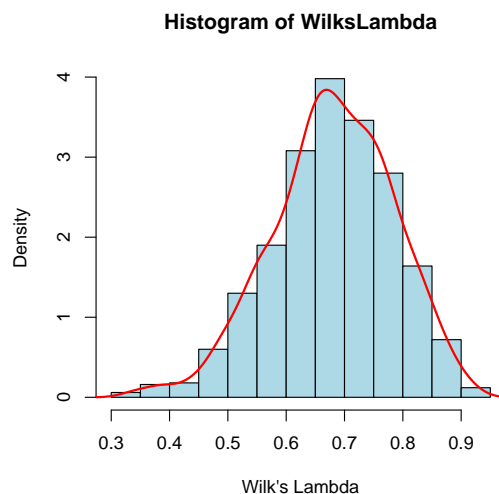
```
> N <- 1000
> m <- 30
> n <- 4
> I <- diag(rep(1,3))
> A <- rWishart(N, df = m, I)
> B <- rWishart(N, df = n, I)
```

Con estas dos muestras independientes podemos calcular la lambda de Wilks para cada pareja de matrices

```
> WilksLambda <- apply(A, 3, det)/apply(A + B, 3, det)
```

y con estos valores podemos dibujar un histograma y la densidad aproximada para ellos:

```
> hist(WilksLambda, col = "lightblue", freq = F,
+       xlab = "Wilk's Lambda", ylab = "Density")
> lines(density(WilksLambda), col = "red", lwd = 2)
```



La función de densidad de la distribución Λ se puede calcular (ver teorema 3.7.3 de Mardia[3]) como el producto de n variables independientes con distribución beta.

$$\prod_{i=1}^n U_i \sim \Lambda(p, m, n)$$

donde $U_i \sim B(\frac{1}{2}(m+i-p), \frac{1}{2}p)$, $i = 1, \dots, n$ y son independientes.

5.2. Propiedades

5.2.1. Cambio de escala

La distribución lambda de Wilks es invariante por cambios de escala en los parámetros de \mathbf{A} y \mathbf{B} . Es decir, si la matriz \mathbf{A} es $W_p(\Sigma, m)$ y la matriz \mathbf{B} es $W_p(\Sigma, n)$ y son independientes y $m \geq p$, entonces

$$\Lambda = \frac{|\mathbf{A}|}{|\mathbf{A} + \mathbf{B}|} = |\mathbf{I} + \mathbf{A}^{-1}\mathbf{B}|^{-1} \sim \Lambda(p, m, n)$$

5.2.2. Valores propios de $\mathbf{A}^{-1}\mathbf{B}$

Si $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ son los valores propios de $\mathbf{A}^{-1}\mathbf{B}$ y $k = \min(n, p)$ es el número de valores propios no nulos, entonces

$$\Lambda = |\mathbf{I} + \mathbf{A}^{-1}\mathbf{B}|^{-1} = \prod_{i=1}^p (1 + \lambda_i)^{-1} = \prod_{i=1}^k (1 + \lambda_i)^{-1}$$

Luego Λ es una función de los valores propios no nulos de $\mathbf{A}^{-1}\mathbf{B}$.

Como consecuencia, se puede demostrar (ver teorema 3.7.4 de Mardia[3]) que

$$\Lambda(p, m, n) = \Lambda(n, m + n - p, p)$$

5.2.3. Relación con la distribución F

Cuando $n = 1$ o $n = 2$ el estadístico Λ se puede expresar en términos de la distribución F . Del mismo modo, cuando $p = 1$ o $p = 2$.

$$\begin{aligned} n = 1 & \quad \frac{1 - \Lambda}{\Lambda} \cdot \frac{m - p + 1}{p} \sim F_{p, m-p+1} \\ p = 1 & \quad \frac{1 - \Lambda}{\Lambda} \cdot \frac{m}{n} \sim F_{n, m} \\ n = 2 & \quad \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \cdot \frac{m - p + 1}{p} \sim F_{2p, 2(m-p+1)} \\ p = 2 & \quad \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \cdot \frac{m - 1}{n} \sim F_{2n, 2(m-1)} \end{aligned}$$

5.2.4. Aproximaciones asintóticas

Para otros valores de n y p que no sean 1 o 2, siempre que m sea grande, se puede utilizar la aproximación asintótica de Bartlett (1947) según la cual

$$U = - \left[m - \frac{1}{2}(p - n + 1) \right] \log \Lambda(p, m, n) \sim \chi_{np}^2$$

Otra aproximación asintótica más precisa se debe a Rao (1951). Si $\Lambda \sim \Lambda(p, t - q, q)$, entonces

$$V = \frac{ms - 2c}{pq} \cdot \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \sim F_{pq, ms-2c}$$

donde $m = t - (p + q + 1)/2$, $s^2 = (p^2q^2 - 4)/(p^2 + q^2 - 5)$ y $c = (pq - 2)/4$.

El estadístico V tiene una distribución F con pq y $ms - 2c$ grados de libertad, incluso aunque $ms - 2c$ no sea un número entero.

Ejemplo 5.2 Supongamos que para decidir en un determinado test se utiliza un estadístico Λ que sigue una distribución $\Lambda(3, 30, 4)$. Para la muestra observada, el valor de dicho estadístico es $\Lambda = 0.35$. Vamos a utilizar la aproximación de Rao para calcular el p -valor para ese estadístico.

```
> L <- 0.35
> p <- 3
> q <- 4
> t <- 34 # 34 - 4 = 30
> m <- t - (p+q+1)/2
> s <- sqrt((p^2*q^2-4)/(p^2+q^2-5))
> c <- (p*q-2)/4
> V = ((m*s-2*c)/(p*q)) * (1-L^(1/s))/L^(1/s)
> pf(V, df1 = p*q, df2 = m*s-2*c, lower.tail = FALSE)

[1] 0.001711322
```

Se tiende a rechazar la hipótesis nula si el estadístico Λ es pequeño. Así que la cola de rechazo para el estadístico Λ es la de la izquierda, mientras que para V es la de la derecha.

6. Inferencia

6.1. Estimación de los parámetros

Supongamos que disponemos de una muestra aleatoria de tamaño n con datos procedentes de una normal $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Esos datos se pueden poner en forma de matriz \mathbf{X} de dimensión $n \times p$, donde cada fila es una observación de la muestra.

Las estimaciones de máxima verosimilitud de los parámetros $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ son

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} \quad \hat{\boldsymbol{\Sigma}} = \mathbf{S}$$

donde $\bar{\mathbf{x}}$ es el vector con las medias de cada una de las columnas de \mathbf{X} , es decir,

$$\bar{\mathbf{x}}' = \frac{1}{n} \mathbf{1}' \mathbf{X}$$

y \mathbf{S} es la matriz de varianzas-covarianzas muestral

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

donde $\mathbf{H} = \mathbf{I} - \frac{1}{n} \mathbf{1} \mathbf{1}'$ es la matriz de centrado.

Sin embargo, \mathbf{S} no es un estimador insesgado de $\boldsymbol{\Sigma}$ de modo que en la práctica se utiliza la estimación centrada

$$\hat{\mathbf{S}} = \mathbf{X}' \mathbf{H} \mathbf{X} / (n - 1)$$

Ejemplo 6.1 Con los datos generados en el ejemplo 1.1 podemos calcular el vector de medias muestrales y la matriz de varianzas-covarianzas.

```
> X <- cbind(y1,y2)
> dim(X)

[1] 500 2
```

```

> # medias
> colMeans(X)

      y1      y2
2.143973 3.273615

> # matriz de var-cov
> cov(X) # var(X)

      y1      y2
y1 3.876299 6.676989
y2 6.676989 24.183268

```

En \mathbf{R} , la matriz de covarianzas es la corregida o centrada.

Entre sus propiedades podemos destacar las siguientes:

1. La distribución de $\bar{\mathbf{x}}$ es $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$.
2. $\bar{\mathbf{x}}$ y \mathbf{S} son estocásticamente independientes.
3. $n\mathbf{S}$ sigue la distribución de Wishart.

6.2. Contraste sobre la media en una población normal

Supongamos que las filas de \mathbf{X} son muestras independientes de una población $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Queremos un test para contrastar la hipótesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

donde $\boldsymbol{\mu}_0$ es un vector con valores conocidos. Hay dos situaciones:

1. Si $\boldsymbol{\Sigma}$ es conocida (cosa poco probable), como $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$, el estadístico de contraste es

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi_p^2$$

2. Si $\boldsymbol{\Sigma}$ es desconocida, como $(n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim T^2(p, n-1)$, el estadístico de contraste es

$$\frac{n-p}{p} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim F_{p, n-p}$$

En ambos casos se rechaza H_0 para valores grandes del estadístico, es decir, la zona de rechazo es la cola derecha.

6.3. Comparación de medias en dos poblaciones normales

Supongamos ahora que tenemos dos matrices de datos independientes \mathbf{X} $n_1 \times p$, \mathbf{Y} $n_2 \times p$ que provienen de distribuciones $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ y sea $n = n_1 + n_2$. Queremos contrastar la hipótesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$$

1. Si $\boldsymbol{\Sigma}$ es conocida (cosa poco probable), como $\bar{\mathbf{x}} - \bar{\mathbf{y}} \sim N_p(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, (1/n_1 + 1/n_2)\boldsymbol{\Sigma})$, el estadístico de contraste es

$$\frac{n_1 n_2}{n} (\bar{\mathbf{x}} - \bar{\mathbf{y}})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \sim \chi_p^2$$

2. Si $\boldsymbol{\Sigma}$ es desconocida, resulta que

$$\frac{n_1 n_2}{n} (\bar{\mathbf{x}} - \bar{\mathbf{y}})' \mathbf{S}_c^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \sim T^2(p, n-2)$$

donde $\mathbf{S}_c = (n_1\mathbf{S}_1 + n_2\mathbf{S}_2)/(n-2)$ es el estimador combinado e insesgado de $\mathbf{\Sigma}$. Este es el **test T^2 de Hotelling**.

En consecuencia, el estadístico de contraste es

$$\frac{n-p-1}{(n-2)p} \frac{n_1 n_2}{n} (\bar{\mathbf{x}} - \bar{\mathbf{y}})' \mathbf{S}_c^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}}) \sim F_{p, n-p-1}$$

3. Si las poblaciones son $N_p(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ y las matrices de covarianzas de las dos poblaciones son distintas, el estadístico

$$(\bar{\mathbf{x}} - \bar{\mathbf{y}})' \left(\frac{\mathbf{S}_1}{n_1 - 1} + \frac{\mathbf{S}_2}{n_2 - 1} \right)^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}})$$

sigue aproximadamente una distribución $T^2(p, \nu)$, donde ν tiene una fórmula ciertamente compleja. Para valores grandes de n_1 y n_2 se puede aproximar a una χ_p^2 . Para valores moderados de n_1 y n_2 existe una transformación a F más precisa.

4. Dadas las duras condiciones para utilizar el test T^2 de Hotelling, también se puede calcular un test de permutaciones con el estadístico T^2 para obtener un p -valor no paramétrico.

Ejemplo 6.2 Con las dos muestras de gorriones de la tabla 1.1 de Manly[4] vamos a contrastar si hay diferencias entre los dos grupos (supervivientes y no supervivientes). Suponemos normalidad multivariante e igualdad de covarianzas.

```
> load("gorriones.RData")
> 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
> # medias por grupo
> medias1 <- colMeans(gorriones[alive=="S", 1:5])
> medias2 <- colMeans(gorriones[alive=="N", 1:5])
> # covarianzas por grupo
> S1 <- cov(gorriones[alive=="S", 1:5])
> S2 <- cov(gorriones[alive=="N", 1:5])
> # matriz de covarianzas común
> S <- ((n1-1)*S1+(n2-1)*S2)/(n1+n2-2)
> # comparación de medias
> 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)
> # resultados
> c(D2=D2, T2=T2, F=F, p.valor=p.valor)
```

D2	T2	F	p.valor
0.2353082	2.8236983	0.5166767	0.7621709

Con este p -valor no se puede rechazar la igualdad de medias entre los dos grupos.

Con la función `hotelling.test()` del paquete *Hotelling* el contraste es inmediato:

```

> library(Hotelling)
> fit <- hotelling.test(. ~ alive, data=gorriones)
> fit$stat[[1]] # T2

[1] 2.823698

> fit

Test stat: 0.51668
Numerator df: 5
Denominator df: 43
P-value: 0.7622

```

La misma función nos ofrece el test de permutaciones.

```

> fitPerm <- hotelling.test(. ~ alive, data=gorriones, perm=TRUE, progBar=FALSE)
> fitPerm

Test stat: 0.51668
Numerator df: 5
Denominator df: 43
Permutation P-value: 0.7685
Number of permutations : 10000

```

En este caso el p-valor es muy similar y la decisión es la misma.

6.4. Comparación de medias en poblaciones normales

Se trata de contrastar la hipótesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_m$$

es decir, una generalización del análisis de la varianza de un factor que en el caso univariante se contrasta con el test F .

Para ello se calculan las matrices de dispersión entre grupos \mathbf{B} , dentro de los grupos \mathbf{W} y total $\mathbf{T} = \mathbf{B} + \mathbf{W}$ como se explica en Cuadras[2] (apartado 3.3.3) o en Manly[4] (apartado 4.7). Entonces, si la hipótesis nula es cierta, se verifica que

$$\mathbf{B} \sim W_p(\Sigma, m-1), \quad \mathbf{W} \sim W_p(\Sigma, n-m), \quad \mathbf{T} \sim W_p(\Sigma, n-1)$$

\mathbf{B} y \mathbf{W} son estocásticamente independientes.

Por lo tanto, si H_0 es cierta, el estadístico Λ de Wilks es

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} \sim \Lambda(p, n-m, m-1)$$

Rechazaremos H_0 si Λ es un valor pequeño en la cola izquierda o la transformación a F cae en la cola de la derecha.

Sean $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ los valores propios de la matriz $\mathbf{W}^{-1}\mathbf{B}$, entonces la Λ de Wilks se puede expresar en la forma

$$\Lambda = \prod_{i=1}^p \frac{1}{1 + \lambda_i}$$

Otro estadístico para el mismo contraste es el mayor valor propio λ_1 , que se conoce con el nombre de **test de la raíz mayor de Roy**. Se basa en la idea de que la combinación lineal de las variables que maximiza la razón entre la suma de cuadrados entre grupos y la suma de cuadrados dentro de los grupos

tiene como valor de esa razón λ_1 . Luego es un buen estadístico para valorar si la variabilidad entre grupos es significativamente grande. Algunos autores prefieren considerar $\lambda_1/(1 - \lambda_1)$. Para los detalles del test se puede consultar la tabla 4.4 de Manly[4].

El tercer estadístico con el mismo propósito es la **traza de Pillai**:

$$V = \sum_{i=1}^p \frac{\lambda_i}{1 + \lambda_i}$$

De nuevo, un valor grande de este estadístico indica que las muestras se pueden considerar de poblaciones con medias distintas. Consultar la tabla 4.4 de Manly[4].

Por último, el cuarto estadístico para contrastar la hipótesis de medias iguales es la **traza de Lawes-Hotelling**:

$$U = \sum_{i=1}^p \lambda_i$$

que es justamente la suma de los valores propios de la matriz $\mathbf{W}^{-1}\mathbf{B}$. Otra vez, un valor grande de este estadístico indica el rechazo de la hipótesis nula. Ver la aproximación al test F en la tabla 4.4 de Manly[4].

En el apartado 15.8 de Cuadras[2] se explican estos mismos estadísticos, pero hay que tener en cuenta que los valores propios que allí se utilizan son los de la matriz $\mathbf{R}_1^{-1}\mathbf{R}_0$ donde $\mathbf{R}_0 = \mathbf{W}$ y $\mathbf{R}_1 = \mathbf{T} = \mathbf{W} + \mathbf{B}$, es decir, $\tilde{\lambda}_i = 1/(1 + \lambda_i)$. Con este cambio en las fórmulas de Cuadras se obtienen los mismos estadísticos que en la tabla 4.4 de Manly. Con la notación de Cuadras la lambda de Wilks es

$$\Lambda = \frac{|\mathbf{R}_0|}{|\mathbf{R}_1|} = |\mathbf{T}^{-1}\mathbf{W}| = \prod_{i=1}^p \tilde{\lambda}_i$$

Generalmente, los cuatro test dan resultados similares en cuanto a la significación. Todos necesitan la hipótesis de normalidad multivariante de las poblaciones con igual matriz de varianzas-covarianzas. Son estadísticos bastante robustos si los tamaños muestrales son similares en las m muestras. Si hay dudas sobre la normalidad multivariante o sobre la igualdad de las matrices de covarianzas, parece que el estadístico traza de Pillai es el más robusto. Finalmente, siempre nos queda realizar un test de permutaciones.

Ejemplo 6.3 *Consideremos los datos sobre cráneos de varones egipcios de cinco épocas históricas que se pueden bajar desde la página del libro de Everitt(2005). Estos mismos datos se muestran en la tabla 1.2 del libro de Manly[4]. También se pueden obtener en el paquete **heplots** de **R**.*

```
> # Skulls <- source("chap5skulls.dat")$value
> # colnames(Skulls) <- c("epoch", "mb", "bh", "bl", "nh")
> data(Skulls, package = "heplots")
> attach(Skulls)
```

Vamos a comparar las medias de las cuatro variables en las cinco muestras. Primero podemos hacer los tests univariantes.

```
> summary(aov(mb ~ epoch))
> summary(aov(bh ~ epoch))
> summary(aov(bl ~ epoch))
> summary(aov(nh ~ epoch))
```

Los tests de las tres primeras medidas son significativos, luego hay diferencias entre las medias con respecto al tiempo.

Ahora vamos a considerar las cuatro variables juntas.

```

> n <- dim(Skulls)[1]
> # matriz Total
> T <- cov(Skulls[,-1])*(n-1)
> skullsdat <- split(Skulls[,-1], Skulls[,1])
> skullsdat <- lapply(skullsdat, as.matrix)
> skulls.var <- lapply(skullsdat, var)
> table(epoch)

epoch
c4000BC c3300BC c1850BC c200BC cAD150
      30      30      30      30      30

> # Matriz Within
> W <- (30-1) * Reduce("+", skulls.var)
> # Lambda de Wilks
> (Lambda <- det(W)/det(T))

[1] 0.6635858

```

Por otra parte, los valores propios de $\mathbf{T}^{-1}\mathbf{W}$ son

```

> val.propios <- eigen(solve(T) %*% W)$values

```

de modo que

```

> (Lambda <- prod(val.propios))

[1] 0.6635858

```

Estos cálculos “a mano” no son necesarios ya que **R** dispone de una función específica para realizar el MANOVA.

```

> skulls.manova <- manova(as.matrix(Skulls[,-1]) ~ epoch)
> summary(skulls.manova, test="Wilks")

          Df   Wilks approx F num Df den Df    Pr(>F)
epoch         4 0.66359   3.9009    16 434.45 7.01e-07 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

La función `manova()` de **R** proporciona los cuatro tests explicados si cambiamos el argumento `test` de la función `summary()` entre *Pillai* (por defecto), *Wilks*, *Hotelling-Lawley* o *Roy*.

Por si nos puede interesar para otros cálculos, el `summary` del objeto `manova` contiene algunos elementos interesantes como los valores propios de $\mathbf{W}^{-1}\mathbf{B}$ y las matrices **B** y **W**.

```

> skulls.manova.summ <- summary(skulls.manova)
> names(skulls.manova.summ)

[1] "row.names"  "SS"          "Eigenvalues" "stats"

> # matriz Within
> skulls.manova.summ$SS$Residuals

```



```

      mb      bh      bl      nh
mb 3061.066667  5.333333  11.46667 291.3000
bh   5.333333 3405.266667 754.00000 412.5333
bl  11.466667 754.000000 3505.96667 164.3333
nh 291.300000 412.533333 164.33333 1472.1333

> # matriz Between
> skulls.manova.summ$SS$epoch

      mb      bh      bl      nh
mb 502.8267 -228.14667 -626.6267 135.43333
bh -228.1467 229.90667 292.2800 -66.06667
bl -626.6267 292.28000 803.2933 -180.73333
nh 135.4333 -66.06667 -180.7333 61.20000

> # matriz Total
> skulls.manova.summ$SS$epoch + skulls.manova.summ$SS$Residuals

      mb      bh      bl      nh
mb 3563.8933 -222.8133 -615.16 426.7333
bh -222.8133 3635.1733 1046.28 346.4667
bl -615.1600 1046.2800 4309.26 -16.4000
nh 426.7333 346.4667 -16.40 1533.3333

```

6.5. Comparación de las varianzas de dos muestras

El test de la razón de verosimilitudes para la hipótesis $H_0 : \Sigma_1 = \Sigma_2$ es

$$\lambda_R = \frac{|\mathbf{S}_1|^{n_1/2} |\mathbf{S}_2|^{n_2/2}}{|\mathbf{S}_c|^{n/2}}$$

donde \mathbf{S}_i es la matriz de covarianzas de los datos de la población i , estimación máximo verosímil (sesgada) de Σ_i y

$$\mathbf{S}_c = (n_1 \mathbf{S}_1 + n_2 \mathbf{S}_2)/n$$

es la estimación máximo verosímil de la matriz de covarianzas común bajo H_0 . Entonces, aprovecharemos la distribución asintótica

$$-2 \log \lambda_R = n \log |\mathbf{S}| - n_1 \log |\mathbf{S}_1| - n_2 \log |\mathbf{S}_2| \sim \chi_{p(p+1)/2}^2$$

para decidir si el estadístico es significativo. Este es el **test de Bartlett**.

Ejemplo 6.4 Con los datos de los gorriones que hemos utilizado en el ejemplo 6.2 tenemos

```

> # Matrices de covarianzas sesgadas (estimaciones máximo verosímiles)
> 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

```

En el paquete *heplots* hay una función que calcula el test de Bartlett de forma univariante.

```
> heplots::bartlettTests(gorriones[,-6],gorriones[,6])

Bartlett's Tests for Homogeneity of Variance

      Chisq df Pr(>Chisq)
length 0.5314 1 0.46603
wing    2.0690 1 0.15032
head    0.5487 1 0.45887
humerus 4.2596 1 0.03903 *
sternum 3.6395 1 0.05642 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El test M de Box mejora notablemente el de Bartlett y lo veremos en el siguiente apartado, pero ambos tests son muy sensibles a la no normalidad de las muestras. Por ello, podemos construir un test más robusto usando los principios del test de Levene univariante.

El **test de Levene** multivariante consiste en calcular el valor absoluto de las desviaciones de los datos a sus medianas en cada grupo y comparar las medias de esas desviaciones con un test T^2 .

Ejemplo 6.5 Con los mismos datos del ejemplo anterior, vamos a calcular en primer lugar el test de Levene univariante aplicado a cada una de las variables.

```
> heplots::leveneTests(gorriones[,1:5],gorriones[,6])

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
```

Se trata de contrastes bilaterales, aunque en este caso los gorriones no supervivientes deberían ser más variables que los supervivientes si se da una selección estabilizadora, luego se requiere mejor un contraste unilateral (ver ejemplo 4.2 de Manly[4]). En todo caso, parece que no hay diferencias de variabilidad univariante.

Con los datos estandarizados de forma global, el test de Levene multivariante se puede aplicar así:

```
> gor.estandar <- scale(gorriones[,1:5])
> gor.estandar <- data.frame(gor.estandar, alive=gorriones[,6])
> gor.dat <- split(gor.estandar[,1:5],gor.estandar[,6])
> gor.dat <- lapply(gor.dat, as.matrix)
> medianas <- lapply(gor.dat, function(x) apply(x,2,median))
> abs.dev.S <- abs(sweep(gor.dat$S,2,medianas$S))
> abs.dev.N <- abs(sweep(gor.dat$N,2,medianas$N))
> fit <- hotelling.test(abs.dev.S,abs.dev.N)
> fit
```

```
Test stat: 0.86874
Numerator df: 5
Denominator df: 43
P-value: 0.5099
```

El estadístico es en realidad el estadístico F y no es significativo, luego no hay evidencia de una diferencia significativa entre las muestras.

Otra posibilidad es el **test de Van Valen** (1978) que propone calcular las distancias euclídeas de las observaciones al vector de medianas para cada muestra y luego comparar los dos grupos de distancias con un test t .

En este test estamos suponiendo que si las dos muestras difieren, entonces una muestra es más variable que la otra *para todas las variables*. El efecto de mayor variabilidad en unas variables y menor en otras, podría cancelar las diferencias en las distancias. Por ello, este test no es apropiado cuando el nivel de variación no sea consistente para todas las variables.

Ejemplo 6.6 Con los datos en valor absoluto de las desviaciones respecto a la mediana de cada muestra que hemos calculado en el ejemplo anterior, tenemos:

```
> abs.dev.all <- rbind(abs.dev.S,abs.dev.N)
> abs.dev.all <- data.frame(alive,abs.dev.all)
> norma.euclidea <- function(x) sqrt(sum(x^2))
> distancias <- data.frame(alive, dist=apply(abs.dev.all[,1],1,norma.euclidea))
> # test de comparación de los dos grupos
> dist_S <- distancias$dist[alive=="S"]
> dist_N <- distancias$dist[alive=="N"]
> t.test(dist_S, dist_N, var.equal=TRUE, alt="less", data=distancias)
```

Two Sample t-test

```
data: dist_S and dist_N
t = -1.9241, df = 47, p-value = 0.0302
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.06392549
sample estimates:
mean of x mean of y
 1.741959  2.241557
```

Aquí se ha aplicado un test unilateral ya que la variabilidad de los no supervivientes debería ser mayor que la de los supervivientes. El resultado es significativo y distinto al test de Levene. Como se ha dicho anteriormente, 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.

6.6. Comparación de varianzas de varias muestras

El test M de Box es un refinamiento del test de la razón de verosimilitudes o test de Bartlett para comparar varias matrices de covarianzas en poblaciones normales.

Para contrastar la hipótesis $H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_m$, el test de la razón de verosimilitudes es

$$\lambda_R = \frac{|\mathbf{S}_1|^{n_1/2} \dots |\mathbf{S}_m|^{n_m/2}}{|\mathbf{S}_c|^{n/2}}$$

donde \mathbf{S}_i es la matriz de covarianzas de los datos de la población i , estimación máximo verosímil de Σ_i , y

$$\mathbf{S}_c = (n_1 \mathbf{S}_1 + \cdots + n_m \mathbf{S}_m) / n$$

es la estimación máximo verosímil de la matriz de covarianzas común bajo la hipótesis H_0 . Entonces, el estadístico

$$-2 \log \lambda_R = n \log |\mathbf{S}| - (n_1 \log |\mathbf{S}_1| + \cdots + n_m \log |\mathbf{S}_m|) \sim \chi_q^2$$

donde $q = (m-1)p(p+1)/2$, puede decidir sobre dicha hipótesis.

El estadístico M de Box es

$$M = \frac{|\hat{\mathbf{S}}_1|^{(n_1-1)/2} \cdots |\hat{\mathbf{S}}_m|^{(n_m-1)/2}}{|\hat{\mathbf{S}}_c|^{(n-m)/2}}$$

donde $\hat{\mathbf{S}}_i = [n_i / (n_i - 1)] \mathbf{S}_i$ y $\hat{\mathbf{S}}_c$ es la estimación insesgada de la matriz de covarianzas común bajo H_0 .

$$\begin{aligned} \hat{\mathbf{S}}_c &= (n_1 \mathbf{S}_1 + \cdots + n_m \mathbf{S}_m) / (n - m) \\ &= ((n_1 - 1) \hat{\mathbf{S}}_1 + \cdots + (n_m - 1) \hat{\mathbf{S}}_m) / (n - m) \end{aligned}$$

Entonces, se puede mejorar la aproximación asintótica a la distribución χ^2 en la forma

$$-2 \cdot c \log M$$

para una constante c (ver Cuadras[2] apartado 7.5.2).

Ejemplo 6.7 Con los datos de los cráneos egipcios, vamos a comparar las matrices de covarianzas de las cinco épocas.

```
> library(biotools)
---
biotools version 3.1

> # test ji-cuadrado de homogeneidad de matrices de covarianzas
> chitest.boxM <- boxM(Skulls[, -1], Skulls[, 1])
> chitest.boxM
```

Box's M-test for Homogeneity of Covariance Matrices

```
data: Skulls[, -1]
Chi-Sq (approx.) = 45.667, df = 40, p-value = 0.2483
```

El test no es significativo y no hay evidencias de cambio en las matrices de covarianzas con el tiempo.

En el libro de Manly[4] (apartado 4.8) se explica una aproximación del mismo estadístico M a una distribución F .

El test M de Box es muy sensible a la falta de normalidad multivariante de los datos. En el ejemplo propuesto, parece razonable utilizarlo ya que las variables son medidas biométricas y no hay motivos para pensar que no cumplan este requisito. Sin embargo, se pueden generalizar las alternativas más robustas propuestas en el apartado anterior al caso de más de dos muestras. En el caso del test de Levene habrá que utilizar el test MANOVA con cualquiera de los cuatro tests propuestos. En el caso del test de Van Valen, se pueden comparar las distancias de varios grupos con un ANOVA.

6.7. Contrastes de normalidad multivariante

En primer lugar hay que recordar que si existe normalidad multivariante, también las distribuciones marginales deben ser normales. De modo que el primer procedimiento debe ser comprobar la normalidad de cada variable de forma gráfica y con algún test como el de Shapiro-Wilk. Si alguna de las variables se desvía de la normalidad, tampoco habrá normalidad multivariante. Sin embargo, esta condición es necesaria, pero no suficiente.

Por otra parte, la normalidad univariante de las componentes principales sí asegura la normalidad multivariante (ver Cuadras[1] pág. 71).

Aunque en **R** hay varias funciones para contrastar la normalidad univariante, como `shapiro.test()`, aquí vamos a utilizar el paquete **MVN** y su función `mvn()` que permite calcular tests univariantes, multivariantes y gráficos adecuados. La función `mvn()` proporciona varias pruebas de normalidad univariante ampliamente utilizadas, incluido "SW" para la prueba de Shapiro-Wilk (no debe aplicarse si el conjunto de datos incluye más de 5000 casos o menos de 3 casos), "CVM" para el test de Cramer-von Mises, "Lillie" para la prueba de Lilliefors, "SF" para la prueba de Shapiro-Francia y "AD" para la prueba de Anderson-Darling.

Ejemplo 6.8 Con los datos de los gorrones podemos calcular la normalidad univariante para cada variable dentro de cada grupo.

```
> library(MVN)
> result.S <- mvn(gor.estandar[alive=="S",1:5])
> result.S$univariateNormality
```

	Test	Variable	Statistic	p value	Normality
1	Shapiro-Wilk	length	0.9340	0.1653	YES
2	Shapiro-Wilk	wing	0.9189	0.0824	YES
3	Shapiro-Wilk	head	0.9522	0.3748	YES
4	Shapiro-Wilk	humerus	0.9653	0.6284	YES
5	Shapiro-Wilk	sternum	0.9364	0.1845	YES

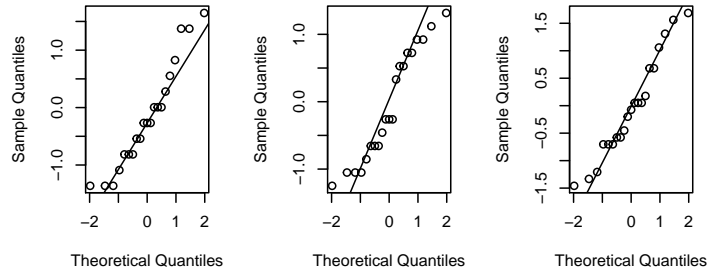
```
> result.N <- mvn(gor.estandar[alive=="N",1:5])
> result.N$univariateNormality
```

	Test	Variable	Statistic	p value	Normality
1	Shapiro-Wilk	length	0.9384	0.1004	YES
2	Shapiro-Wilk	wing	0.9745	0.7061	YES
3	Shapiro-Wilk	head	0.9741	0.6947	YES
4	Shapiro-Wilk	humerus	0.9760	0.7474	YES
5	Shapiro-Wilk	sternum	0.9790	0.8254	YES

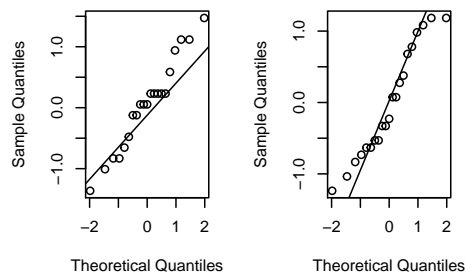
Con el parámetro `univariatePlot = "qqplot"`, la función `mvn()` dibuja los gráficos univariantes Q-Q de ajuste a la normal.

```
> result.S <- mvn(gor.estandar[alive=="S",1:5],
+               univariatePlot = "qqplot")
```

Normal Q-Q Plot (lengt Normal Q-Q Plot (wing Normal Q-Q Plot (heac



Normal Q-Q Plot (humer Normal Q-Q Plot (sternu

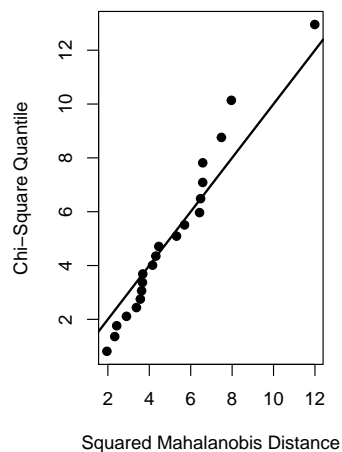


Antes de mostrar algunos contrastes multivariantes, recordemos la propiedad 2.2.8 de las distribuciones normales multivariantes, por la que la distancia de Mahalanobis al cuadrado entre una observación y la media sigue la distribución χ^2 . Luego un gráfico con las distancias D^2 es muy práctico para descartar o no la normalidad e incluso destacar posibles observaciones atípicas.

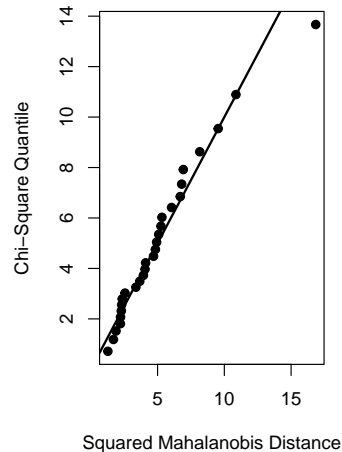
Ejemplo 6.9 Con los datos de los gorrones, los dos conjuntos de distancias al cuadrado se muestran así:

```
> oldpar <- par(mfrow=c(1,2))
> result.S <- mvn(gor.estandar[alive=="S",1:5], multivariatePlot = "qq")
> result.N <- mvn(gor.estandar[alive=="N",1:5], multivariatePlot = "qq")
> par(oldpar)
```

Chi-Square Q-Q Plot



Chi-Square Q-Q Plot



En el primer gráfico (gorriones supervivientes) se ven algunas distancias que se apartan de la recta pero no parece nada grave. En el segundo gráfico el ajuste es mejor, excepto por una observación claramente atípica.

En cuanto a los contrastes, la función `mvn()` dispone de los contrastes de Mardia, Royston, Henze-Zirkler y Energy. La prueba de Mardia se basa en las distribuciones asintóticas de los estimadores multivariantes de asimetría y kurtosis (ver Mardia[3]).

Ejemplo 6.10 Con los datos de los gorriones, por defecto se obtiene el test de Mardia.

```
> result.S$mmultivariateNormality
```

	Test	Statistic	p value	Result
1	Mardia Skewness	22.557061342527	0.948452434424016	YES
2	Mardia Kurtosis	-1.27021254209566	0.20400893124091	YES
3	MVN	<NA>	<NA>	YES

```
> result.N$mmultivariateNormality
```

	Test	Statistic	p value	Result
1	Mardia Skewness	50.3843124692158	0.0445863656184518	NO
2	Mardia Kurtosis	0.24100488890806	0.809551327393527	YES
3	MVN	<NA>	<NA>	NO

El resultado indica la normalidad multivariante del grupo de supervivientes. En cambio, se rechaza la normalidad multivariante del grupo de no supervivientes por el test de asimetría. Se puede analizar con más detalle y descubrir que el culpable es el gorrion 37 de la muestra conjunta (16 en la de no supervivientes) que tiene unas observaciones atípicas como vimos en el gráfico de las distancias de Mahalanobis al cuadrado.

6.8. Pruebas de esfericidad

Antes de aplicar el Análisis de Componentes Principales (PCA) a un conjunto de datos multivariantes, debe comprobarse si será útil, es decir, si la correlación entre las variables analizadas es lo suficientemente grande como para justificar la descomposición de la matriz de coeficientes de correlación. Esta comprobación puede hacerse mediante el test de Bartlett (1950), que parte de la hipótesis nula de que la matriz de correlaciones no es significativamente distinta de la matriz identidad $H_0 : \mathbf{R} = \mathbf{I}$. Bartlett calcula un estadístico basado en la razón de verosimilitudes del siguiente modo:

$$-[n - 1 - (2m + 5)/6] \log |\mathbf{R}| \sim \chi^2_{(m^2 - m)/2}$$

donde \mathbf{R} es la matriz de correlaciones muestral.

Ejemplo 6.11 Con los datos de los gorriones, el test de esfericidad de Bartlett es

```
> # Bartlett's sphericity test
> bartlett.sphere <- function(data){
+   n <- dim(data)[1]
+   k <- dim(data)[2]
+   R <- cor(data, use='pairwise.complete.obs')
+   chi.square <- -(n-1-(2*k+5)/6) * log(det(R))
+   cat("\n", "Bartlett's sphericity test", "\n", "\n")
+   cat('chi.square = ', round(chi.square, 3) ,
+       ', df = ', (k^2-k)/2,
+       ', p-value = ', 1-pchisq(chi.square, (k^2-k)/2))
+ }
```

```
+ }  
> bartlett.sphere(gor.estandar[,1:5])  
  
Bartlett's sphericity test  
  
chi.square = 150.193 , df = 10 , p-value = 0
```

En este caso, el PCA está justificado ya que el estadístico es significativo.

El índice de Kaiser-Meyer-Olkin o medida de adecuación muestral KMO tiene el mismo objetivo que el test de Bartlett, se trata de saber si podemos factorizar las variables originales de forma eficiente.

Referencias

- [1] C.M. Cuadras, *Métodos de Análisis Multivariante*, EUNIBAR, 1981.
- [2] C.M. Cuadras, *Nuevos métodos de análisis multivariante*, CMC Editions, 2018.
- [3] Mardia, K. V., J. T. Kent, and J. M. Bibby *Multivariate Analysis*, London: Academic Press, 1979.
- [4] Manly, Bryan F.J. and Navarro Alberto, Jorge A., *Multivariate Statistical Methods: A Primer*, Fourth edition, CRC, 2017.