

Actividad 1.4 La Normal Multivariada

Franco Mendoza Muraira A01383399

2023-11-08

```
library(mnormt)
library(MVN)
```

```
## Warning: package 'MVN' was built under R version 4.2.3
```

1. Calculo de probabilidad

```
miu <- c(2.5, 4)
sigma <- matrix(c(1.2, 0, 0, 2.3), nrow = 2)

prob <- pmnorm(c(3, 2), mean = miu, varcov = sigma)
```

```
## La probabilidad de que  $P(X_1 \leq 3, X_2 \leq 2)$  es: 0.06328658
```

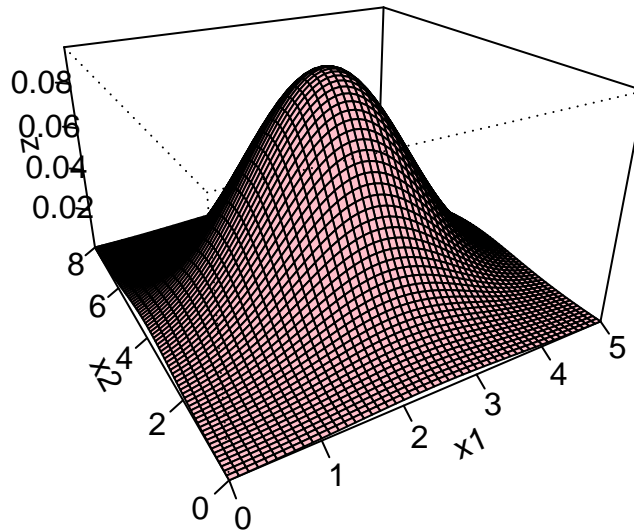
2. Graficación de la anterior distribución bivariada

```
x1 <- seq(0, 5, 0.1)
x2 <- seq(0, 8, 0.1)

mu <- c(2.5, 4)
sigma <- matrix(c(1.2, 0, 0, 2.3), nrow = 2)

f <- function(x1, x2) dmnorm(cbind(x1, x2), mu, sigma)
z <- outer(x1, x2, f)

persp(x1, x2, z, theta = -30, phi = 25, expand = 0.6, ticktype = 'detailed', col = "pink")
```



3. Contornos de la anterior distribución normal bivariada

Parece que el contorno con diámetro más pequeño usualmente son los de probabilidades más bajas, que en este caso serían los de 0.01 y 0.03, esto debido a que están más alejados del centro. En este caso el más pequeño sería el contorno de 0.01.

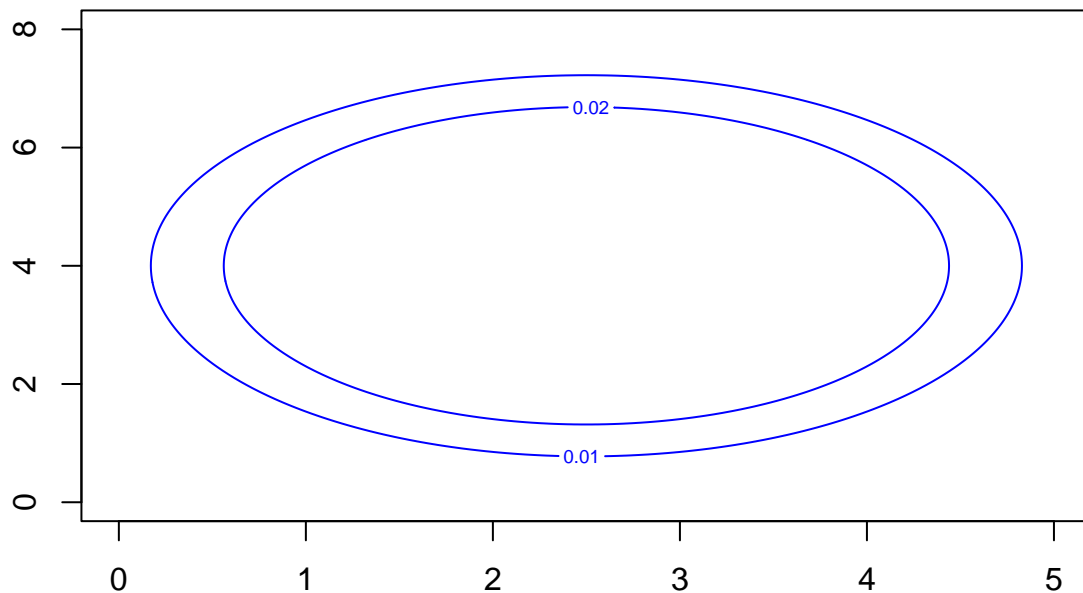
El contorno con el diámetro más grande serían los de 0.09 y 0.1 debido a que están más cerca del centro y representan áreas más comunes.

```
x1 <- seq(0, 5, 0.01)
x2 <- seq(0, 8, 0.01)

mu <- c(2.5, 4)
sigma <- matrix(c(1.2, 0, 0, 2.3), nrow = 2)

f <- function(x1, x2) dmnorm(cbind(x1, x2), mu, sigma)
z <- outer(x1, x2, f)

contour(x1, x2, z, col = "blue", levels = c(0.01, 0.02))
```



Relaciones entre los problemas y gráficos:

Relación entre el primer y segundo problema: El primer problema implica el cálculo de una probabilidad bivariada en una distribución normal, mientras que el segundo problema involucra la representación gráfica de contornos de una distribución normal bivariada. Ambos problemas están relacionados con el concepto de probabilidad en una distribución bivariada normal.

Relación entre los gráficos de los incisos 2 y 3: Ambos problemas presentan gráficos tridimensionales. El inciso 2 muestra una superficie que representa la función de densidad de probabilidad de una distribución normal bivariada, mientras que el inciso 3 muestra contornos de esa misma distribución, resaltando áreas con diferentes niveles de probabilidad.

Relación entre la altura del contorno y el porcentaje de datos abarcados: La altura del contorno en el gráfico representa la probabilidad de que los datos estén dentro de ese contorno. Por lo general, a mayor altura del contorno, mayor será el porcentaje de datos que abarca ese contorno, ya que representa una probabilidad más alta de ocurrencia en la distribución normal bivariada.

4. Combinaciones lineales

a) Las variables X_1 y X_2 son conjuntamente normales e independientes con distribución univariada $N(0,1)$. Dadas las combinaciones lineales:

$$Y_1 = 3 + 2X_1 - X_2 \text{ y } Y_2 = 5 + X_1 + X_2$$

Tomando en cuenta que el vector de medias de X es $[0,0]$ y la matriz sigma es una de identidad de 2×2 , sacamos la función de densidad con las siguientes formulas:

$$N(A\mu + d, A\Sigma_x A^t)$$

```
miux = matrix(c(0,0),nrow = 2,ncol = 1)
A = matrix(c(2,-1,1,1),nrow=2,ncol = 2,byrow = TRUE)
d = matrix(c(3,5),nrow=2,ncol=1)
miuy = A%%miux+d
miuy
```

```
##      [,1]
## [1,]    3
## [2,]    5
```

Aquí sacamos la primer parte de la función de densidad bivalente de Y_1 y Y_2 , la cuál es la parte de la media. Ahora sacamos la matriz de varianza:

```
sigx = diag(2)
sigy = A%%sigx%%t(A)
sigy
```

```
##      [,1] [,2]
## [1,]    5    1
## [2,]    1    2
```

Ya con esto tenemos el inciso a del problema, con el vector de medias y la matriz de varianza de Y_1 y Y_2 .

b) ¿Cómo se distribuye la variable $Z = 3Y_1 + 4Y_2 - 1$?

Ahora para sacar la distribución de la variable Z , se hará lo mismo pero con los valores de las Y .

```
B = matrix(c(3,4),nrow = 1,ncol=2)
e = -1
miuz = B%%miuy+e
miuz
```

```
##      [,1]
## [1,]   28
```

Aquí conseguimos la primer parte de la Z , la media de 28, ahora se sacará la varianza:

```
sigz = B%%sigy%%t(B)
sigz
```

```
##      [,1]
## [1,]  101
```

Con esto se consigue el sigma de la función de la variable Z , y tenemos que $Z \sim N(28, 101)$, se distribuye de forma normal.

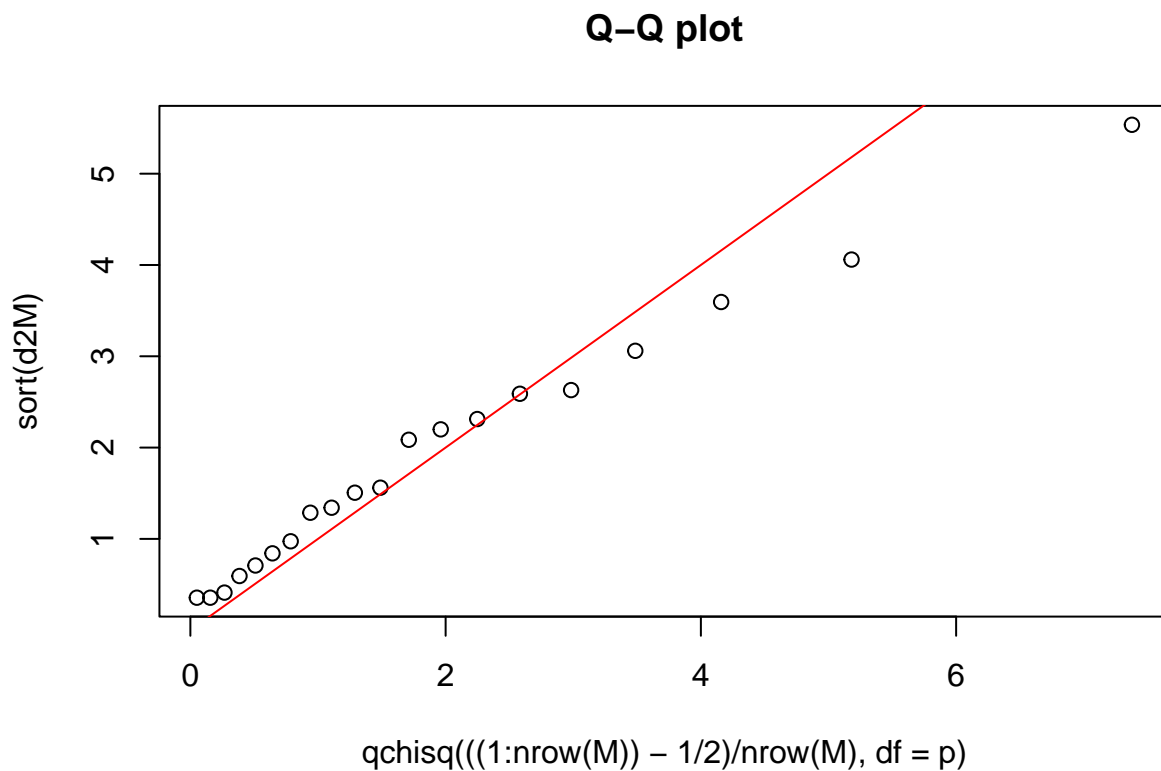
5. Prueba de normalidad bivariada

Nivel de Significancia: 0.05

```
M = read.csv("datosA2.csv")

p = 2          #indica que se trata de dos variables
# Vector de medias
X = colMeans(M)
#Matriz de covarianza
S = cov(M)
#Distancia de Mahalanobis
d2M = mahalanobis(M,X,S)

#Multinormalidad Test gráfico Q-Q Plot
plot(qchisq(((1:nrow(M)) - 1/2)/nrow(M),df=p),sort( d2M ),main="Q-Q plot" )
abline(a=0, b=1,col="red")
```



```
## Test de Multinormalidad: Método Sesgo y kurtosis de Mardia
mvn(M,subset = NULL,mvn = "mardia", covariance = FALSE,showOutliers = FALSE)
```

```
## $multivariateNormality
##           Test      Statistic      p value Result
## 1 Mardia Skewness 3.59823747819632 0.46309914697164   YES
## 2 Mardia Kurtosis -1.43530997731026 0.151198785877334   YES
```

```
## 3          MVN          <NA>          <NA>      YES
##
## $univariateNormality
##          Test Variable Statistic    p value Normality
## 1 Anderson-Darling      x      1.2355    0.0024      NO
## 2 Anderson-Darling      y      0.2451    0.7257      YES
##
## $Descriptives
##      n Mean   Std.Dev Median Min Max 25th 75th      Skew   Kurtosis
## x 20 0.18 0.1361114    0.1 0.0 0.5 0.10 0.225 0.8185140 -0.3698838
## y 20 5.04 1.0054588    5.0 3.3 6.7 4.35 5.850 0.1357527 -1.2067384
```

Interpreta el gráfico de QQ-Plot. Indica: ¿qué semejanzas hay entre este gráfico y el caso univariado? ¿qué diferencias?

Los gráficos Q-Q Plot se interpretan de manera similar a la normalidad univariada, comparando la distribución de los datos con la distribución teórica de una normal. Similar al análisis univariado, muestra la forma y dispersión de los datos, aunque en lugar de enfocarse en una única distribución, compara dos distribuciones diferentes, evidenciando si los datos se ajustan a la distribución teórica propuesta.

Interpreta los valores p de los resultados correspondientes a Mardia Skewness y Mardia Kurtosis. Observa las significancias univariadas y multivariadas.

H_0 : Los datos se distribuyen normalmente.

H_1 : Los datos no se distribuyen normalmente.

Mardia Skewness: Este mide la asimetría multivariada, en este caso tenemos un sesgo de 3.598 y un valor p de 0.463, por lo que no hay suficiente evidencia para rechazar la hipótesis nula, y con esto la normalidad de la distribución en cuanto a la asimetría multivariada.

Mardia Kurtosis: Esta mide la curtosis multivariada, en este estudio vemos que tenemos una curtosis de -1.435, y un valor p de 0.1512, por lo que al igual que en el sesgo, no podemos rechazar la hipótesis nula, y se concluye que los datos se distribuyen normalmente en cuanto a la curtosis multivariada.

Con esto también podemos ver que sí se presenta una normalidad multivariada debido al pase de estas 2 pruebas, teniendo pvalues mayores al nivel de significancia de 0.05.

Pruebas Univariadas: En las pruebas de normalidad para las 2 variables “x” y “y”, se usó la prueba Anderson-Darling. Para la variable x, se tiene un valor de la estadística de 1.2355 y un valor p de 0.0024, lo que hace que rechazemos la hipótesis nula, y nos concluye que no tiene una distribución normal en sus datos.

Para la variable y, se tiene un valor de estadística de 0.2451 y un valor p de 0.7257, por lo que con esta variable no tenemos suficiente evidencia para rechazar la hipótesis nula, y concluimos que los datos de la variable “y” sí se distribuyen de forma normal.

Conclusiones:

En la normalidad multivariada, las pruebas de Mardia Skewness y Kurtosis indican que la matriz de datos en su conjunto sigue una distribución normal multivariada, ya que los p-valores son mayores que el nivel de significancia (0.05).

En la normalidad univariada, la variable 'x' no sigue una distribución normal, ya que su valor p en la prueba Anderson-Darling es menor que 0.05. Sin embargo, la variable 'y' parece seguir una distribución normal ya que su p-valor es mayor a 0.05.

En resumen, la matriz X parece seguir una distribución normal multivariada, pero las variables individuales que la componen pueden no seguir una distribución normal en su totalidad, especialmente la variable 'x'.