Actividad 1_4

Ricardo Kaleb Flores Alfonso

2024-09-26

1) Ejercicio 1

Hallar el procedimiento para el cálculo de probabilidad de que $P(X_1 \leq 3, X_2 \leq 2)$ con X_1, X_2 se distribuyen normalmente con

$$\mu = \begin{bmatrix} 2.5 \\ 4 \end{bmatrix}$$

y

$$\sum = \begin{bmatrix} 1.2 & 0 \\ 0 & 2.3 \end{bmatrix}$$

.

```
upper_bound = c(3,2)
mu = c(2.5, 4)

sigma = matrix(c(1.2, 0, 0, 2.3), nrow = 2)

print("El resultado obtenido es: ")
```

```
## [1] "El resultado obtenido es: "
```

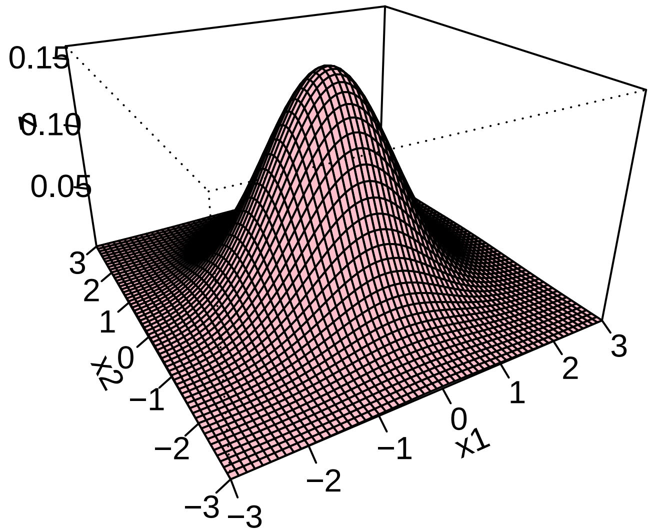
```
pmnorm(x = upper_bound, mean = mu, varcov = sigma)
```

```
## [1] 0.06328658
```

Ejercicio 2

Grafique la anterior distribución

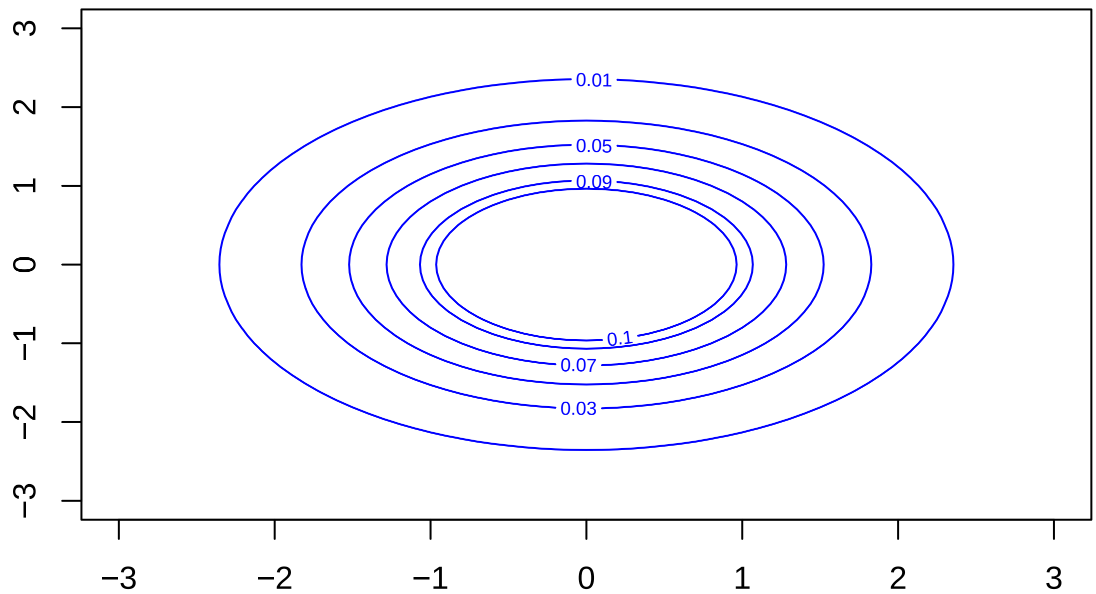
```
x1    <- seq(-3, 3, 0.1)
x2     <- seq(-3, 3, 0.1)
mu    <- c(0, 0)
sigma <- matrix(c(1, 0, 0, 1), nrow=2)
f     <- function(x1, x2) dmnorm(cbind(x1, x2), mu, sigma)
z     <- outer(x1, x2, f)
#create surface plot
persp(x1, x2, z, theta=-30, phi=25, expand=0.6, ticktype='detailed', col = "pink")
```



Ejercicio 3)

¿A qué altura crees que esté el contorno con el diámetro más pequeño? El contorno más pequeño esta cerca de la punta, entonces yo creo que en 0.12 ¿y el más grande? Es el contorno que esta más cerca de la base en 0.01

```
contour(x1, x2, z, col = "blue", levels = c(0.01, 0.03, 0.05, 0.07, 0.09, 0.1))
```



¿Cómo se relaciona el resultado del primer problema con el segundo?

En el primer problema se obtuvo la probabilidad de que dos variables aleatorias normalmente distribuidas caigan dentro de un rectángulo utilizando una distribución normal multivariada. En este problema, se visualiza la función de densidad conjunta de estas dos variables. El área bajo la superficie en el gráfico hasta el punto $(3, 2)$ es la probabilidad que se calculó en el primer problema.

¿Cómo se relacionan los gráficos de los incisos 2 y 3?

El gráfico del problema 2 es una representación de la función de densidad conjunta de las dos variables, donde la altura representa la densidad en ese punto. El gráfico del problema 3 muestra las líneas de contorno de la función de densidad para diferentes niveles de altura. Las líneas de contorno en el gráfico representan la proyeccion del gráfico sobre el plano.

¿Cómo se relaciona la altura del contorno (pregunta 3) con el porcentaje de datos que abarcaría el contorno?

La altura del contorno en una distribución normal multivariada está relacionada inversamente con el porcentaje de datos que abarca. Contornos de mayor altura representan regiones de mayor densidad de probabilidad, estas abarcan un porcentaje menor de los datos totales. Por otro lado, contornos más abajo representan regiones de menor densidad de probabilidad, que abarcan un porcentaje mayor de los datos.

Ejercicio 4)

Dado que las variables x_1 y x_2 son normales e independientes, con distribución univariada N(0,1). Dadas las distribuciones

$$Y_1 = 3 + 2X_1 - X_2$$

y

$$Y_2 = 5 + X_1 + X_2$$

Encuentra la función de densidad bivariante de $Y1$ y $Y2$ Primero se calcula el vector de medias, dado que las medias son 0 para X_1 e X_2

$$\mu = \begin{bmatrix} E(Y_1) \\ E(Y_2) \end{bmatrix}$$

$$E(Y_1) = E(3)$$

y

$$E(Y_2) = E(5)$$

Entonces

$$\mu = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

$$\sum = \begin{bmatrix} cov(Y_1, Y_1) & cov(Y_1, Y_2) \\ cov(Y_2, Y_1) & cov(Y_2, Y_2) \end{bmatrix}$$

$$cov(Y_1, Y_1) = var(Y_1) = var(3) + 2^2 var(X_1) + (-1)^2 var(X_2)$$

Por lo tanto

$$cov(Y_1, Y_1) = 5$$

$$cov(Y_2, Y_2) = var(Y_2) = var(5) + 1^2 var(X_1) + 1^2 var(X_2)$$

De igual manera

$$cov(Y_2, Y_2) = 2$$

$$cov(Y_1, Y_2) = cov(Y_2, Y_1) = cov(2X_1 - X_2, X_1 + X_2)$$

$$cov(Y_1, Y_2) = 2var(X_1) - 1var(X_2) = 1$$

$$cov(Y_1, Y_2) = 2 + 0 + 0 - 1 = 1$$

Por lo que la matriz de covarianza es

$$\sum = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$$

Dado que Y_1 e Y_2 se unen normalmente, su función de densidad es

$$(X, Y)\ N_2 \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$$

Parte 2

Dada z $Z = 3Y_1 + 4Y_2 - 1$ ¿Cómo se distribuye la variable? Primero hay que encontrar su promedio

$$E(Z) = 3E(Y_1) + 4E(Y_2) - 1$$

$$E(Z) = 3(3) + 3(5) - 1 = 28$$

Debido a que Z es una combinación lineal la distribución de Z también será una distribución normal con:

$$\sum = 3^2 var(Y_1) + 4^2 var(Y_2) + 24cov(Y_1, Y_2)$$

$$\sum = 9(5) + 16(2) + 24 = 101$$

De esta manera

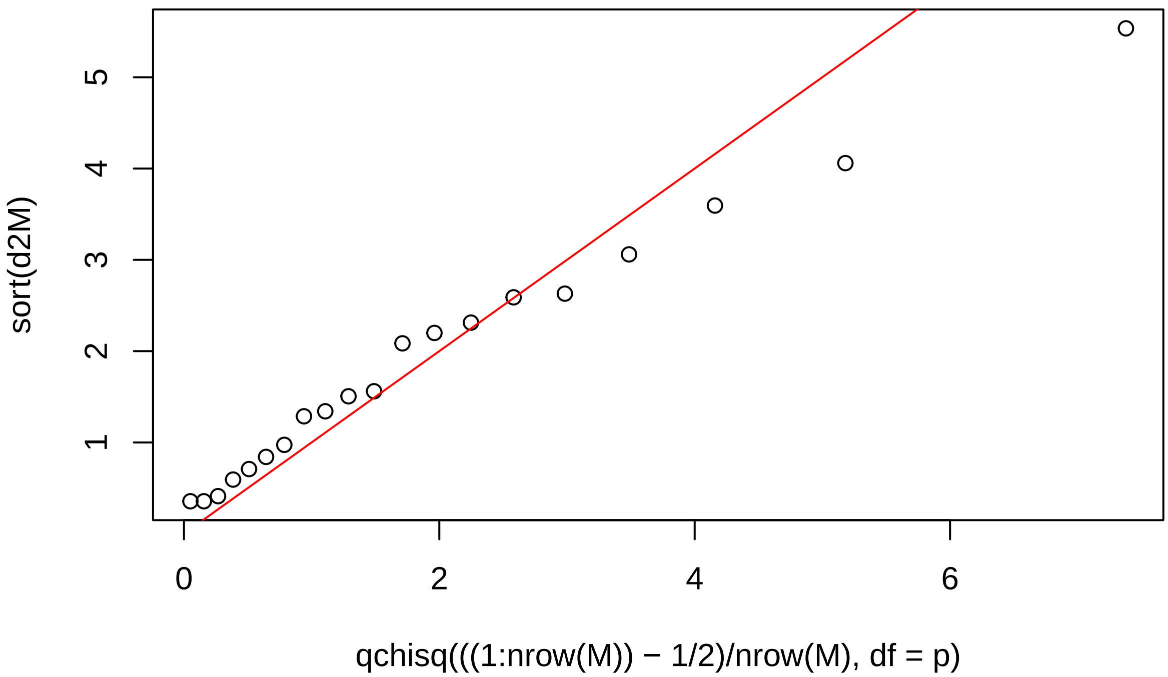
$$\sum N(28, 101)$$

Ejercicio 5)

```
M = read.csv("datos-A2.csv")

p = 2
# 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 ) )
abline(a=0, b=1,col="red")
```




```r
## Test de Multinomalidad: 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
```

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

En el caso univariado, se tiene una gráfica que compara los cuartiles obtenidos con los cuartiles teóricos. Mientras más se apeguen a la linea, más confianza se tiene de que los datos sean distribuidos normalmente. Para el caso multivariado, la diferencia es que ahora se compara la distancia de Mahalanobis con un valor de Chi cuadrada, la linea siendo los valores esperados en el caso de ser una distribución normal.

B. Interpreta los valores p de los resultados correspondientes a Mardia Skewness y Mardia Kurtosis. Recuerde que para el caso de Normalidad. Observa las significancias univariadas y multivariadas.

El sesgo de los datos tiene un valor de p de 0.463, por lo que no se tiene evidencia para rechazar la hipótesis nula. Por lo que no se puede negar que el sesgo no correspinda a una distribución normal. La curtosis obtuvo un valor de p de 0.151, por lo que no se puede rechazar la hipótesis nula, de esta manera la curtosis es la que tendria una distribución normal. Estos resultados indican que la distribución parece normal.

Al realizar el analisis univariado se obtine que el valor de p para la variable x es de 0.0024. De esta manera se rechaza la hipótesis nula y se concluye que la distribución univariada de x no es univariada. Para la variable y el valor de p es de 0.7257, por lo que no se tiene evidencia para rechazar la hipótesis nula. De esta manera se concluye que la distribución univariada de y es normal.

C. Concluye sobre la normalidad multivariada de la matriz X y la normalidad univariada de cada una de las variables que componen la matriz X.

La matriz X muestra evidencia de seguir una distribución normal multivariada, dadas las pruebas de Mardia para asimetría y curtosis. Por otro lado cuando se examina cada variable, se observa que la variable x no sigue una distribución normal, mientras que la variable y sí. Es por esto que aunque los datos pueden ser ser normales de manera conjunta, existe una falta de normalidad en la variable x. Como los modelos noramales multivariados suponen que sus variables serán normales, se puede concluir que la distribución multivariada obtenida no es normal.