Act2_A01251534

Eduardo Alvarado Gómez

2022-10-05

Matrices y Vectores Aleatorios

Ejercicio 1

```
X = matrix(c(1, 6, 8, 4, 2, 3, 3, 6, 3), ncol = 3)
X

##      [,1] [,2] [,3]
## [1,]    1    4    3
## [2,]    6    2    6
## [3,]    8    3    3

bX = X %*% c(1, 1, 1)
bX

##      [,1]
## [1,]    8
## [2,]   14
## [3,]   14

cX = X %*% c(1, 2, -3)
cX

##      [,1]
## [1,]    0
## [2,]   -8
## [3,]    5

A = cbind(bX, cX)
A

##      [,1] [,2]
## [1,]    8    0
## [2,]   14   -8
## [3,]   14    5
```

a) Hallar la media, varianza y covarianza de b’X y c’X

```
mean(bX) # Media

## [1] 12

mean(cX)
```

```
## [1] -1

var(bX[,1]) # Varianza

## [1] 12

var(cX)

##      [,1]
## [1,]   43

cov(bX[,1], cX[,1]) # Covarianza

## [1] -3
```

b) Hallar el determinante de S (matriz de var-covarianzas de X)

```
det(cov(A))

## [1] 507
```

c) Hallar la matriz de varianzas-covarianzas (o porqué no se puede hallar)

```
cov(A)

##      [,1] [,2]
## [1,]   12   -3
## [2,]   -3   43
```

d) Hallar los valores y vectores propios de S

```
e = eigen(cov(A))
e

## eigen() decomposition
## $values
## [1] 43.28765 11.71235
##
## $vectors
##             [,1]        [,2]
## [1,] -0.09544671 -0.99543454
## [2,]  0.99543454 -0.09544671
```

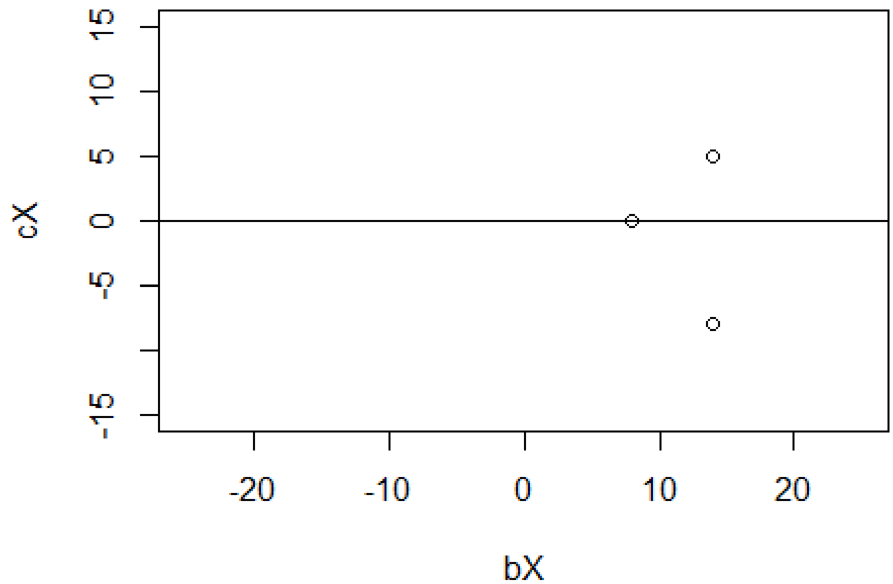
e) Argumentar si b'X y c'X son independientes o no.

Se puede decir que ambos coeficientes son lo suficientemente grandes para contribuir a Y1. Por esto, podemos concluir que b'X y c'X son independientes.

f)Hallar la varianza generalizada. Explicar el comportamiento de los datos de X basándose en los la variable generalizada, en los valores y vectores propios.

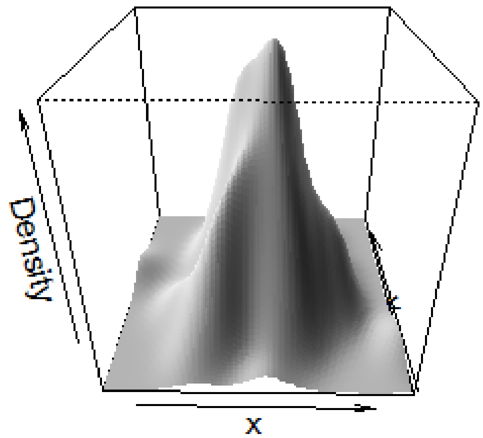
```
plot(bX, cX, xlim = c(-25,25), ylim= c(-15,15))
x11 = seq(0, 100, 100)
x12 = e$vectors[1,1] / e$vectors[2,1] * x11
x21 = seq(0,100,100)
```

```
x22 = e$vectors[1,1] / e$vectors[1,2] * x21
abline(x11,x12)
abline(x21,x22)
```



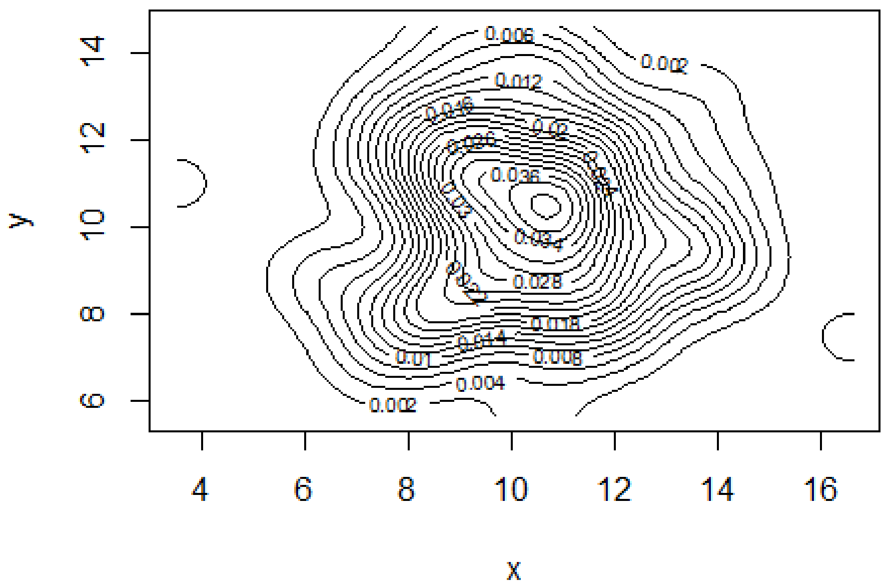
Ejercicio 2

```
library(MVN)

## Warning: package 'MVN' was built under R version 4.0.5

x = rnorm(100, 10, 2)
y = rnorm(100, 10, 2)
datos = data.frame(x,y)
mvn(datos, mvnTest = "hz", multivariatePlot = "persp")
```




```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.5169359 0.5598496 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.3674    0.4249    YES
## 2 Anderson-Darling     y        0.3179    0.5319    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
Skew
## x 100 10.27934 1.980738 10.36636 3.510938 16.62401 8.895845 11.48731
0.04254431
## y 100 10.19276 1.813701 10.24006 5.662811 14.59486 8.843848 11.46476
0.08875697
##     Kurtosis
## x  1.0105541
## y -0.5532237
```

```
mvn(datos, mvnTest = "hz", multivariatePlot = "contour")
```




```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.5169359 0.5598496 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.3674    0.4249    YES
## 2 Anderson-Darling     y        0.3179    0.5319    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th
Skew
## x 100 10.27934 1.980738 10.36636 3.510938 16.62401 8.895845 11.48731
0.04254431
## y 100 10.19276 1.813701 10.24006 5.662811 14.59486 8.843848 11.46476
0.08875697
##     Kurtosis
## x  1.0105541
## y -0.5532237
```

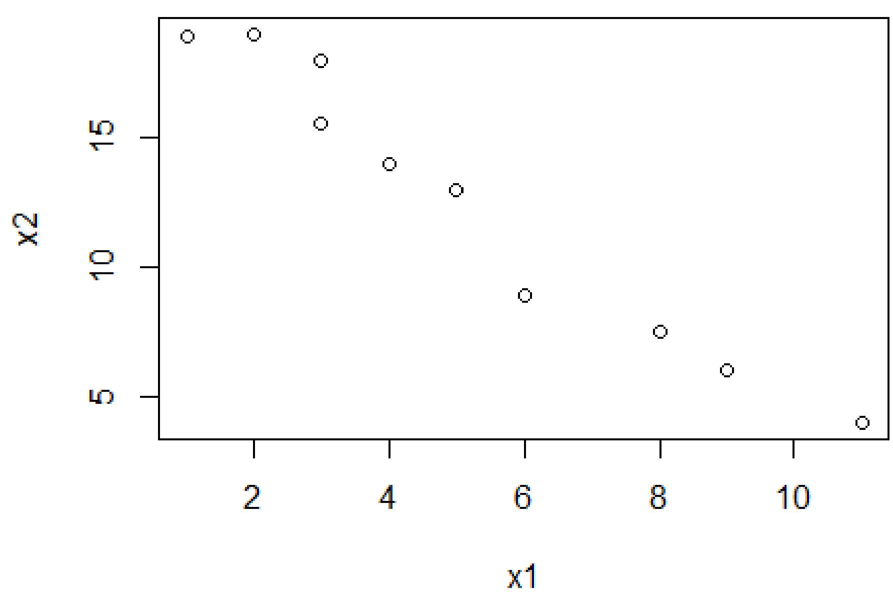
Tras observar las gráficas obtenidas, se pudieron encontrar varias cosas. Primeramente, el diagrama de perspectiva presenta una parte de la curva de distribución de probabilidad univariada, en una distribución de tres dimensiones. Con esto, se puede obtener información de la relación de las variables dentro del set de datos. Las dos dimensiones representan dos variables, mientras que la tercera es la densidad de probabilidad multivariable. Siguiendo con las gráficas, la gráfica de contorno nos muestra un espacio en dos dimensiones con lo que podemos verificar la supocisión de normalidad multivariado. Con esto, se observa la información de normalidad y correlación. En nuestro caso, se puede concluir que hay una correlación positiva entre nuestras variables, hasta cierto nivel. Mientras la correlación se acerque a 0, las líneas en la gráfica se volverán circulares.

Ejercicio 3

```
n = 10
x1 = c(1,2,3,3,4,5,6,8,9,11)
x2 = c(18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99)
X = cbind(x1,x2)
Xbar = colMeans(X)
S = cov(X)
Sinv = solve(S)
```

a)Construya un diagrama de dispersión

```
plot(x = x1, y = x2)
```



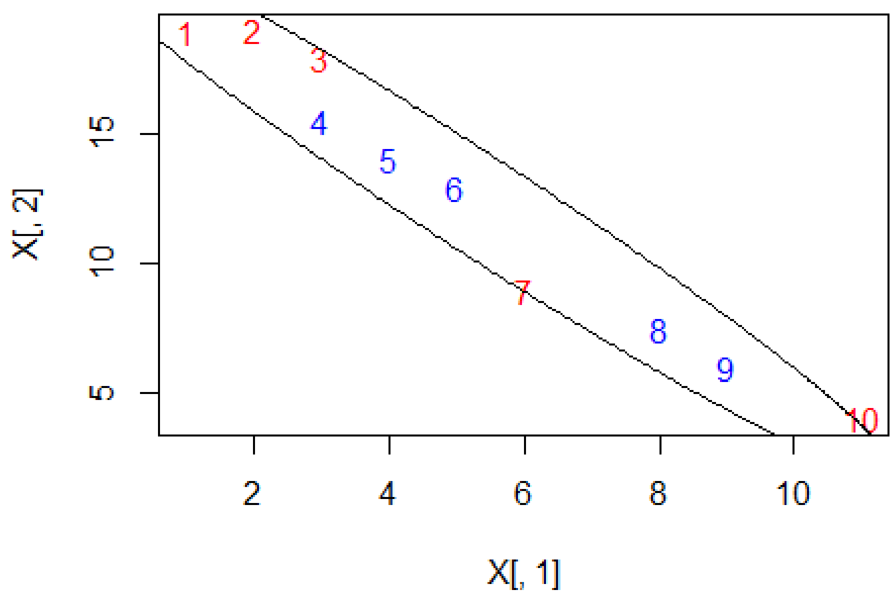
b) Inferir el signo de la covarianza muestral a partir del gráfico.

c) Calcular el cuadrado de las distancias estadísticas

```
d = mahalanobis(X, colMeans(X), cov(X))
d

## [1] 1.8753045 2.0203262 2.9009088 0.7352659 0.3105192 0.0176162 3.7329012
## [8] 0.8165401 1.3753379 4.2152799
```

d) Usando las anteriores distancias, determine la proporción de las observaciones que caen dentro del contorno de probabilidad estimado del 50% de una distribución normal bivariada

```
library(ellipse)

##
## Attaching package: 'ellipse'

## The following object is masked from 'package:graphics':
##
##     pairs

p = 2
elps = t(t(ellipse(S, level=0.85, npoints=1000))+Xbar)
plot(X[,1],X[,2],type="n")
index = d < qchisq(0.5,df=p)
text(X[,1][index],X[,2][index],(1:n)[index],col="blue")
text(X[,1][!index],X[,2][!index],(1:n)[!index],col="red")
lines(elps,col="black")
```

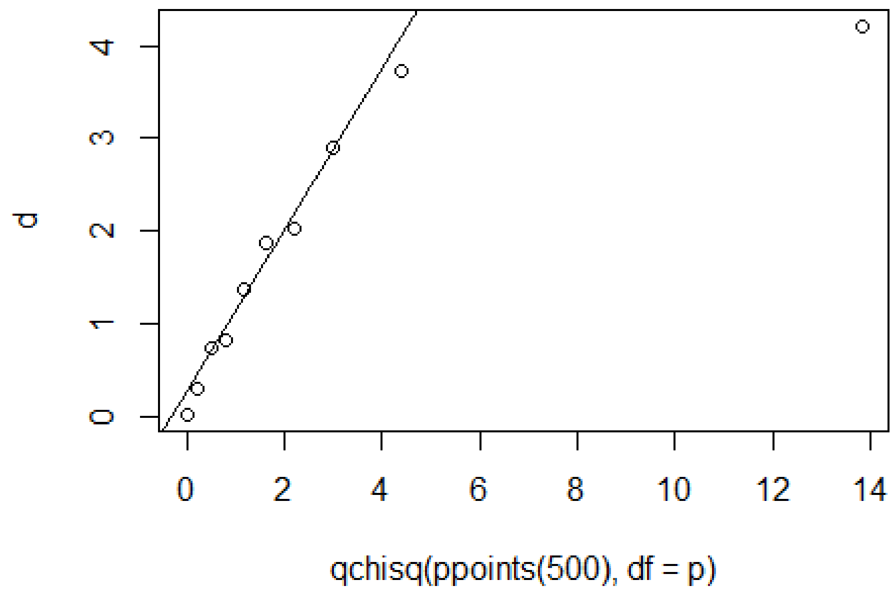


e) Ordene las distancias del inciso c y construya un diagrama chi-cuadrado

```
names(d) = 1:10
sort(d)

##         6         5         4         8         9         1         2 
3
## 0.0176162 0.3105192 0.7352659 0.8165401 1.3753379 1.8753045 2.0203262
```

```
2.9009088
##         7        10
## 3.7329012 4.2152799
```

```
qqplot(qchisq(ppoints(500),df=p), d)
qqline(d,distribution=function(x){qchisq(x,df=p)})
```



f) Dados los resultados anteriores, ¿serían argumentos para decir que los datos son aproximadamente normales bivariados?

Ya que la mayoría de los datos están cerca de la línea teórica, se puede concluir que estos datos son normales bivariados, aproximadamente.