Matrices y Vectores Aleatorios

Andrea Piñeiro

2022-10-03

Matrices y Vectores Aleatorios

1.Matriz de Datos

Encontrar b'X y c'X

```
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) Media, varianza y covarianza

```r
bX_mu = mean(bX)
cX_mu = mean(cX)

bX_var = var(bX[,1])
cX_var = var(cX)

cat("Media de b'X:", bX_mu, "\n")
```

```
## Media de b'X: 12
```

```r
cat("Media de c'X:", cX_mu, "\n", "\n")
```

```
## Media de c'X: -1
##
```

```r
cat("Varianza de b'X:", bX_var, "\n")
```

```
## Varianza de b'X: 12
```

```r
cat("Varianza de c'X:", cX_var, "\n", "\n")
```

```
## Varianza de c'X: 43
##
```

c) Matriz de varianzas-covarianzas

```r
cov_A = cov(A)
cov_A
```

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

b) Determinante de S

```r
bX_det = det(cov_A)
cat("Determinante de cov_A(S):", bX_det, "\n")
```

```
## Determinante de cov_A(S): 507
```

d) Valores y Vectores de S

```
e = eigen(cov_A)
```

$Y_1 = -0.09544671 * x_1 + 0.99543454 * x_2$ $Y_1 = -0.99543454 * x_1 - 0.09544671 * x_2$

Por los valores de Eigen, tomaremos Y1

e) ¿b'X y c'X son independientes?

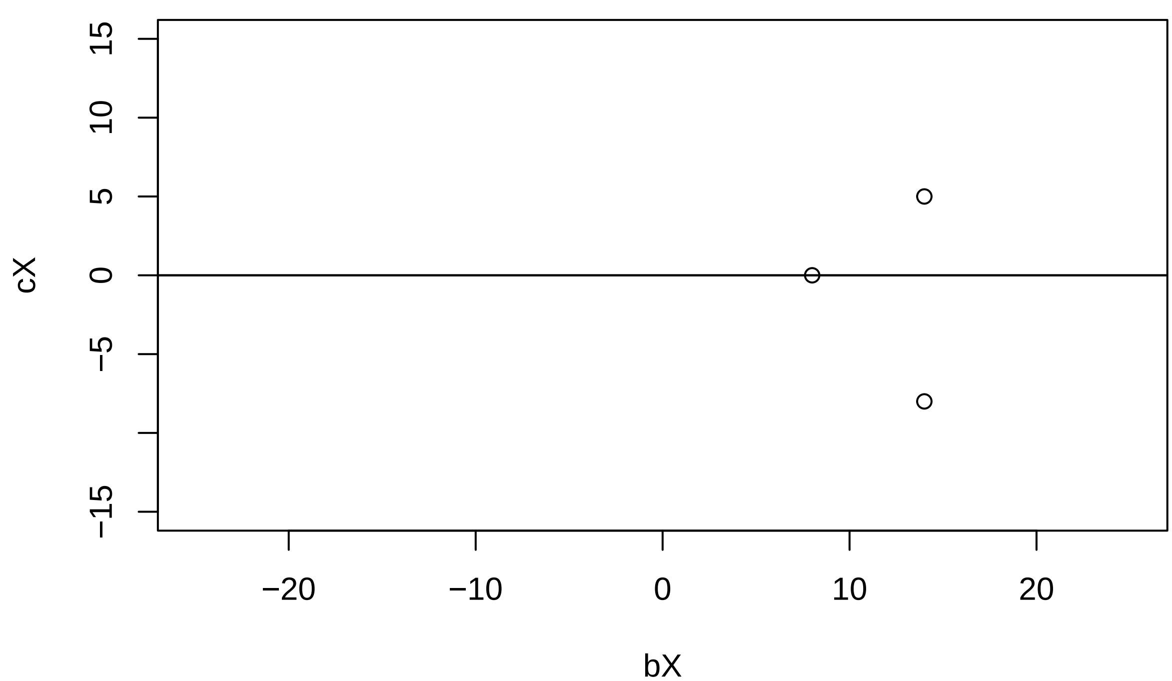
Los tamaños de los coeficientes de ambas variables son lo suficientemente grandes por lo que ambas contribuyen en la interpretación de Y1.

La variable X2 tene mayor peso en el componente Y.

Dado que ambas variables contribuyen para obtener Y, podemos afirmar que b'X y c'X son independientes

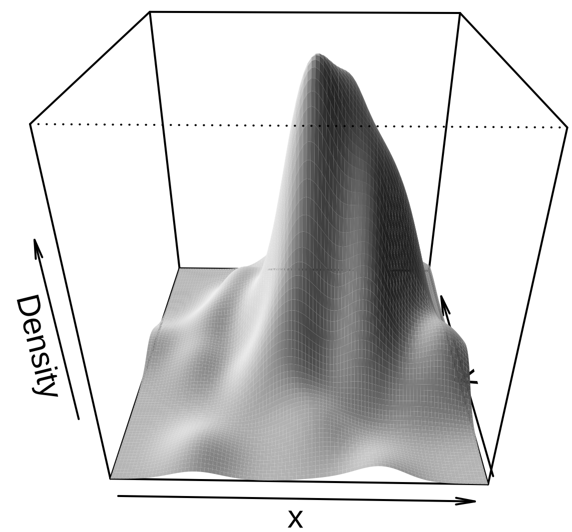
f) Varianza Generalizada

```
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)
```



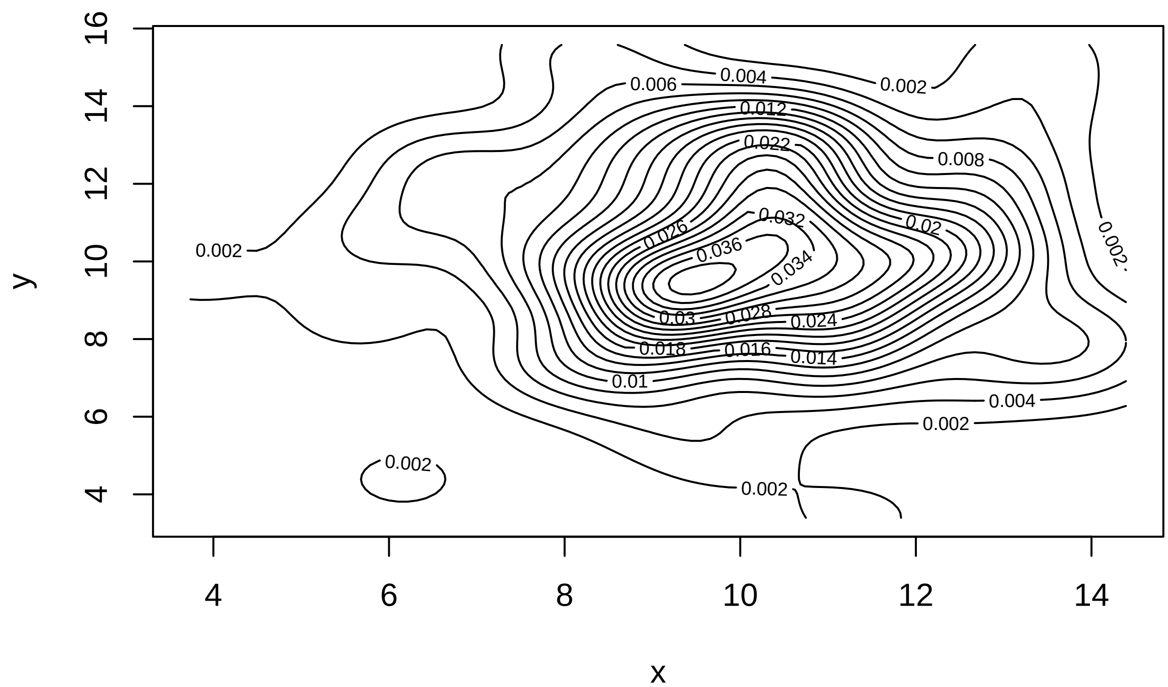
2. Explorar resultados

```
library(MVN)
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 0.5307613 0.5297762 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.4150    0.3286    YES
## 2 Anderson-Darling     y        0.2949    0.5909    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th       Skew
## x 100 10.10803 1.964173 10.27522 3.738664 14.39228 8.995232 11.30185 -0.3560865
## y 100 10.16872 2.256678 10.15635 3.395850 15.57802 8.871966 11.67016 -0.1358233
##    Kurtosis
## x 0.3628485
## y 0.3005437
```

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




```
## $multivariateNormality
##            Test        HZ   p value MVN
## 1 Henze-Zirkler 0.5307613 0.5297762 YES
##
## $univariateNormality
##               Test  Variable Statistic   p value Normality
## 1 Anderson-Darling     x        0.4150    0.3286    YES
## 2 Anderson-Darling     y        0.2949    0.5909    YES
##
## $Descriptives
##     n     Mean  Std.Dev   Median      Min      Max     25th     75th       Skew
## x 100 10.10803 1.964173 10.27522 3.738664 14.39228 8.995232 11.30185 -0.3560865
## y 100 10.16872 2.256678 10.15635 3.395850 15.57802 8.871966 11.67016 -0.1358233
##    Kurtosis
## x 0.3628485
## y 0.3005437
```

La gráfica de contorno muestra la normalidad y correlación de los datos. Existe una correlacón positiva entre nuestra variable x y y.

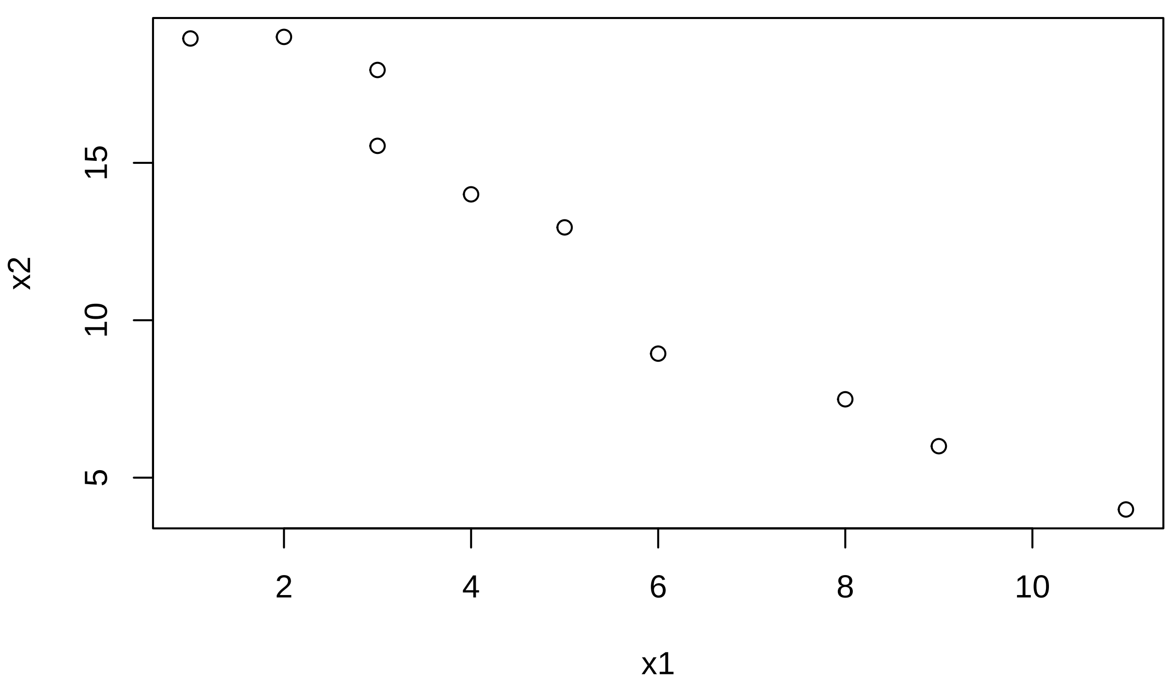
3. Un periódico matutino enumera los siguientes precios de autos usados para un compacto extranjero con edad medida en años y precio en venta medido en miles de dólares.

```
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)
data = data.frame(x1 = x1, x2=x2)
head(data)
```

```
##   x1    x2
## 1  1 18.95
## 2  2 19.00
## 3  3 17.95
## 4  3 15.54
## 5  4 14.00
## 6  5 12.95
```

a) Diagrama de dispersión

```
plot(data)
```



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

Debido a que conforme x1 aumenta, x2 decrece; por lo que la covarianza entre ambos será negativa.

```
cov(x1, x2)
```

```
## [1] -17.71022
```

c) Calcular el cuadrado de las distancias estadísticas

```
maha = mahalanobis(data, colMeans(data), cov(data))
maha
```

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

d) Proporción de las observaciones que caen dentro del contorno de probabilidad estimado del 50% de una distribución normal bivariada.

```
pvalue = pchisq(maha, df = 1)
pvalue
```

```
##  [1] 0.8291312 0.8447942 0.9114704 0.6088184 0.4226382 0.1055900 0.9466493
##  [8] 0.6338063 0.7591031 0.9599385
```

```
inn = pvalue > 0.5
inn
```

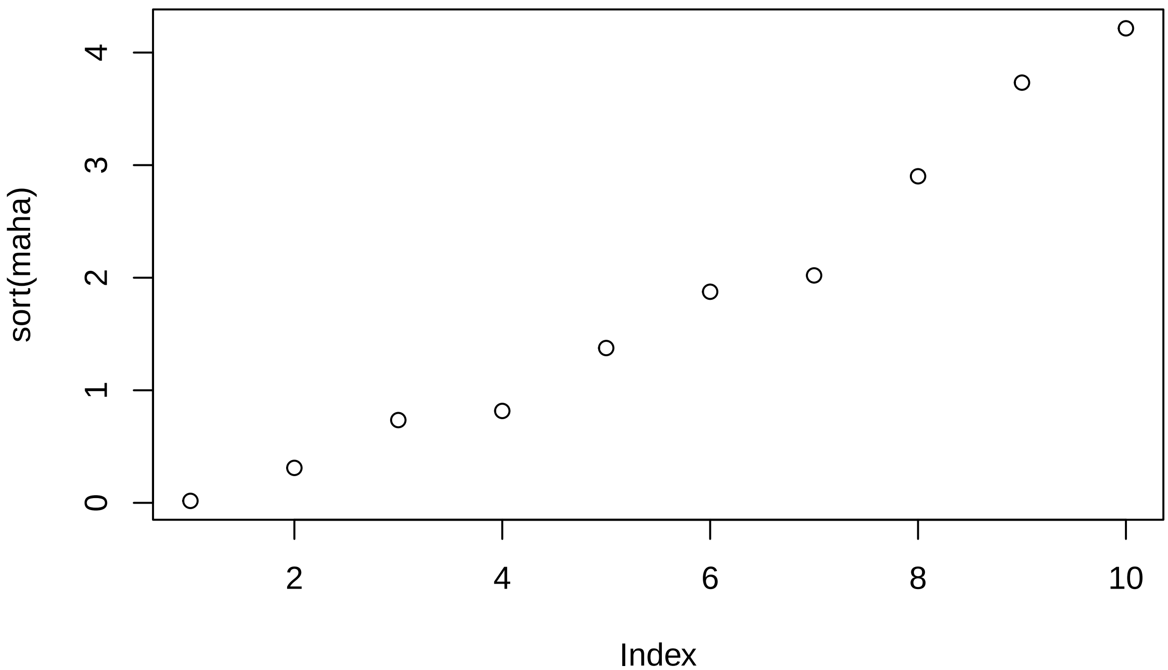
```
##  [1]  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
```

```
table(inn)
```

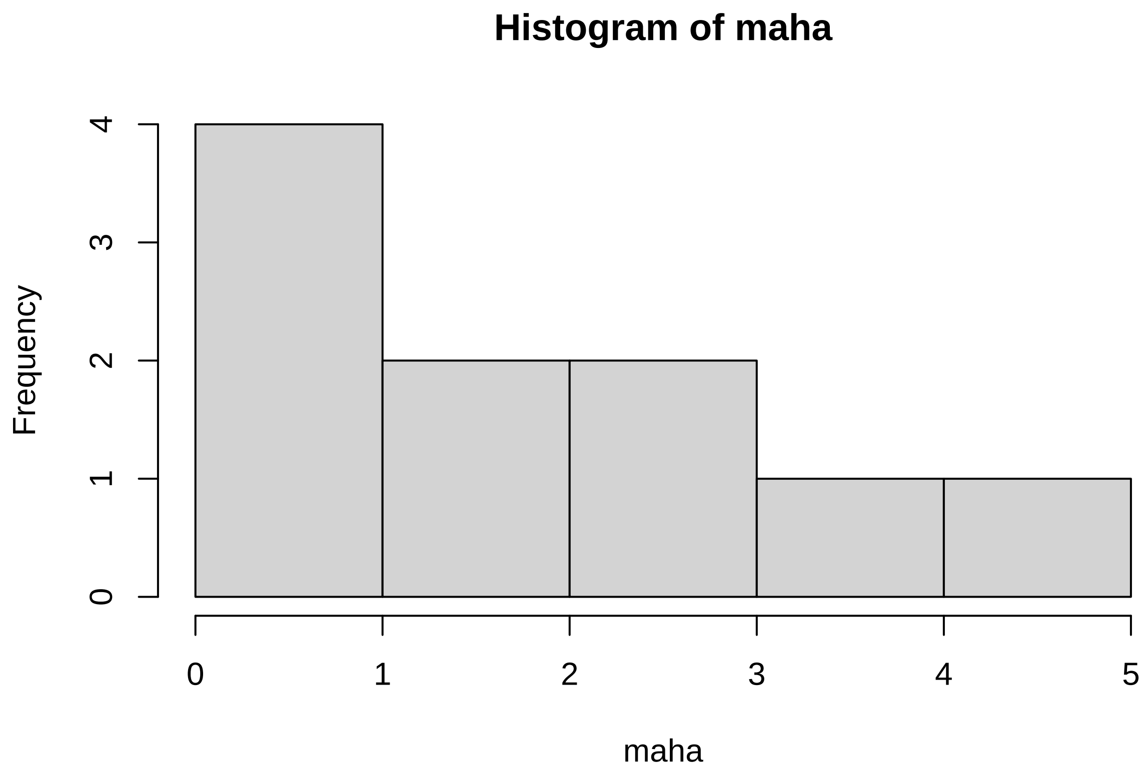
```
## inn
## FALSE  TRUE
##     2     8
```

e) Diagrama chi-cuadrado

```
plot(sort(maha))
```




```
hist(maha)
```



f) ¿Serían argumentos para decir que los datos son aproximadamente normales bivariados?

Si, concluimos que estos datos son aproximadamente normales bivariados.