

Grau interuniversitari (UB-UPC) d'Estadística
Software Estadístic: Solució del Test 1 amb R

Exercici 1 (0.75 + 0.75 + 0.75 + 0.75 + 1 + 0.5 + 0.5 = 5 punts)

- a) Obteniu els següents tres vectors, de 20 components cadascun:

```
 $\vec{u} = (2, 7, 12, 17, \dots, 92, 97)$ 
> u <- seq(from = 2, to = 97, by = 5)
 $\vec{v} = (5, -1, 5, -1, \dots, 5, -1)$ 
> v <- c(5, -1) * rep(x = 1, times = 20)
 $\vec{w} = (0, 1, 0, 2, 0, 3, \dots, 0, 10)$ 
> w <- c(0, 1) * rep(x = 1:10, each = 2)
```

- b) Eleveu al quadrat les components de \vec{u} i obteniu el valor més proper a 2460.

```
> u2 <- u^2
> u2[which.min(abs(u2 - 2460))]
[1] 2704
```

- c) Construïu una matriu diagonal $\mathbf{D}_{10 \times 10}$ en què a la seva diagonal principal s'ubiquin els components no nuls del vector \vec{w} (els valors de la diagonal principal d'aquesta matriu s'obtidran partint del vector \vec{w} , tot respectant l'ordre d'aparició dels seus components).

```
> D <- diag(w[w != 0])
> D
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1	0	0	0	0	0	0	0	0	0
[2,]	0	2	0	0	0	0	0	0	0	0
[3,]	0	0	3	0	0	0	0	0	0	0
[4,]	0	0	0	4	0	0	0	0	0	0
[5,]	0	0	0	0	5	0	0	0	0	0
[6,]	0	0	0	0	0	6	0	0	0	0
[7,]	0	0	0	0	0	0	7	0	0	0
[8,]	0	0	0	0	0	0	0	8	0	0
[9,]	0	0	0	0	0	0	0	0	9	0
[10,]	0	0	0	0	0	0	0	0	0	10

- d) Obtenir la matriu $\mathbf{B}_{20 \times 20}$, amb la següent estructura en blocs:

$$\mathbf{B}_{20 \times 20} = \begin{pmatrix} \mathbf{D}_{10 \times 10} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

```
> N <- matrix(rep(0, 10 * 10), nc = 10) # Opció 1
> B <- cbind(rbind(D, N), rbind(N, N))

> B <- matrix(rep(0, 20 * 20), nc = 20) # Opció 2
> diag(B)[1:10] <- w[w != 0]
```

- e) Considereu ara la següent estructura de matriu tridiagonal

$$\mathbf{T}_{n \times n} = \begin{pmatrix} t_{11} & 1 & & 0 \\ & 1 & \ddots & \ddots \\ & & \ddots & \ddots & 1 \\ 0 & & & 1 & t_{nn} \end{pmatrix},$$

on els elements (t_{11}, \dots, t_{nn}) són realitzacions aleatòries d'una distribució normal de mitjana 3 y variància 5, amb els valors arrodonits sense cap decimal.

Obteniu una proposta de matriu $\mathbf{T}_{n \times n}$ així definida pel cas $n = 20$.

```
> T <- matrix(rep(0, 20 * 20), nc = 20)      # Matriu de zeros 20x20
> set.seed(2)
> diag(T) <- round(rnorm(20, 3, sqrt(5)))     # Diagonal principal
> T[abs(col(T) - row(T)) == 1] <- 1          # Diagonals secundàries
```

- f) Obteniu l'índex de la fila i de la columna on es situa el màxim valor de la vostra matriu $\mathbf{T}_{20 \times 20}$. És invertible aquesta matriu?

```
> which(T == max(T), arr.ind = TRUE)

      row col
[1,]    3   3
[2,]    9   9
[3,]   15  15

> # 3 opcions vàlides per saber si la matriu T és invertible:
> det(T) != 0                                # TRUE: El determinant és no nul
> qr(D)$rank == nrow(T)                      # TRUE: La matriu és de rang màxim
> sum(eigen(T)$values != 0) == ncol(T)       # TRUE: 20 valors propis no nuls
```

- g) Fent servir $\mathbf{T}_{20 \times 20}$, comproveu si és certa la relació

$$\det(\mathbf{T}_{n \times n}^2 + \mathbf{I}_n) \geq [\text{Tr}(\mathbf{T}_{n \times n}) - \det(\mathbf{T}_{n \times n})]^2,$$

on:

$\mathbf{T}_{n \times n}^2$ és el producte matricial del tipus $\mathbf{T}_{n \times n} \cdot \mathbf{T}_{n \times n}$,

\mathbf{I}_n és la matriu identitat de dimensions $n \times n$,

Tr és la funció traça (suma dels elements ubicats a la diagonal principal),

\det és la funció determinant.

```
> # Es planteja la desigualtat en termes de codi:
> det(T%*%T + diag(20)) >= (sum(diag(T)) - det(T))^2

[1] TRUE
```

Exercici 2 (0.5 + 0.6 + 0.6 + 0.5 + 0.6 + 0.5 + 0.4 + 0.8 + 0.5 = 5 punts)

Abans de començar, esborreu tots els elements de l'àrea de treball mitjançant la instrucció:

```
> rm(list = ls(all = TRUE))
```

- a) Carregueu l'àrea de treball `RTest1.RData` i mireu quins objectes n'hi ha. Feu una descripció numèrica d'aquests objectes.

```
> load("RTest1.RData")
> ls()
[1] "bmi"      "filename" "wei"

> summary(bmi)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 22.00  23.70   24.40   24.34  25.50   27.10

> summary(wei)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 53.00  64.00   74.00   76.52  91.00  105.00
```

- b) L'índex de massa corporal (*body mass index*; vector `bmi`) de les 25 persones fictícies s'ha calculat mitjançant la següent fórmula:

$$\text{IMC} = \frac{\text{Pes [kg]}}{\text{Altura [m]}^2}.$$

Donats l'índex de massa corporal i el pes (vector `wei`) calculeu les altures (en cm i sense decimals) de les 25 persones. Guardeu els valors en un vector amb nom `hei`.

```
> (hei <- round(sqrt(wei/bmi)*100))
[1] 154 168 162 163 172 180 192 165 192 161 181 202 174 167 174 194 161 189 189
[20] 178 174 169 170 181 197
```

- c) Utilitzeu la funció `sample` juntament amb l'argument `prob` per crear un vector aleatori anomenat `city` amb les ciutats de residència d'aquestes 25 persones. Les probabilitats que les persones siguin de Barcelona, Badalona, Granollers i Sitges són 0.4, 0.3, 0.2 i 0.1, respectivament.

```
> set.seed(610)
> city <- factor(sample(c("Barcelona", "Badalona", "Granollers", "Sitges"),
+                       25, prob = 4:1, replace = T))
> city
[1] Granollers Granollers Badalona  Badalona  Badalona  Badalona
[7] Badalona  Granollers Badalona  Granollers Badalona  Badalona
[13] Barcelona Sitges      Barcelona Granollers Granollers Barcelona
[19] Barcelona Badalona  Badalona  Barcelona Barcelona Badalona
[25] Barcelona
Levels: Badalona Barcelona Granollers Sitges
```

- d) Quantes persones venen de cadascuna de les 4 ciutats?

```
> table(city)

city
Badalona  Barcelona Granollers  Sitges
      11         7         6         1
```

- e) Quin és l'índex de massa corporal més gran i de quina ciutat és la persona que ho té?

```
> max(bmi)
[1] 27.1

> city[which(bmi == max(bmi))]
[1] Barcelona
Levels: Badalona Barcelona Granollers Sitges
```

- f) Quina és la mediana del pes de les persones que viuen a Barcelona o a Sitges?

```
> median(wei[city %in% c("Barcelona", "Sitges")])
[1] 73

> median(wei[as.numeric(city) %in% c(2, 4)])
[1] 73
```

- g) Creeu un *data frame* amb nom `exer2` que contingui les variables `hei`, `wei`, `bmi` i `city`.

```
> (exer2 <- data.frame(hei, wei, bmi, city))

   hei wei  bmi   city
1  154  53 22.3 Granollers
2  168  62 22.0 Granollers
3  162  59 22.5  Badalona
4  163  64 24.1  Badalona
5  172  71 24.0  Badalona
6  180  79 24.4  Badalona
7  192  91 24.7  Badalona
8  165  60 22.0 Granollers
9  192  91 24.7  Badalona
10 161  64 24.7 Granollers
11 181  84 25.6  Badalona
12 202 104 25.5  Badalona
13 174  72 23.8  Barcelona
14 167  66 23.7    Sitges
15 174  69 22.8  Barcelona
16 194  94 25.0 Granollers
17 161  62 23.9 Granollers
18 189  96 26.9  Barcelona
19 189  93 26.0  Barcelona
20 178  76 24.0  Badalona
21 174  75 24.8  Badalona
22 169  64 22.4  Barcelona
23 170  74 25.6  Barcelona
24 181  85 25.9  Badalona
25 197 105 27.1  Barcelona
```

- h) Quines són les mitjanes de totes les variables numèriques d'`exer2` de les persones que són de Barcelona?

```
> colMeans(subset(exer2, city == "Barcelona")[1:3])
```

```
      hei      wei      bmi  
180.28571  81.85714  24.94286
```

- i) Guardeu (només) el *data frame* en un àrea de treball nova a la vostra carpeta de treball.

```
> save(exer2, file = "SE_Test1.RData")
```