Universitat de Barcelona

Treball

Tercer Semestre

Anàlisi de Dades i Introducció a la Probabilitat (ADIP)

Autor:
Mario Vilar

Professor:
Dra. Carme Florit

8 d'octubre de 2021

Universitat de Barcelona

Índex

Taula

Capítol 1

Capítol 2

I

Preliminars

1.1 Comandes

Definició 1.1.1 (Funció sumari). La funció `summary()` ens dona una sèrie de valors estadístics fonamentals. L'he escollit per raons evidents: té molt a veure amb el que hem estat parlant durant les primeres classes de l'assignatura.

```
X = 2:30
summary(X)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       2       9      16      16      23      30
```

Definició 1.1.2 (Outer). Si a i b són arrays numèrics, el seu producte *outer* és un array la dimensió del qual s'obté concatenant els seus vectors de dimensió, en ordre, i el vector de dades del qual s'obté formant tots els productes d'elements del vector d'a amb el vector b.

```
x <- 0:5
y <- 5:10
f <- function(x, y) cos(y)/(1 + x*x)
z <- outer(x, y, f)
z

##            [,1]       [,2]       [,3]         [,4]        [,5]        [,6]
## [1,] 0.28366219 0.96017029 0.75390225 -0.145500034 -0.91113026 -0.83907153
## [2,] 0.14183109 0.48008514 0.37695113 -0.072750017 -0.45556513 -0.41953576
## [3,] 0.05673244 0.19203406 0.15078045 -0.029100007 -0.18222605 -0.16781431
## [4,] 0.02836622 0.09601703 0.07539023 -0.014550003 -0.09111303 -0.08390715
## [5,] 0.01668601 0.05648061 0.04434719 -0.008558826 -0.05359590 -0.04935715
## [6,] 0.01091008 0.03692963 0.02899624 -0.005596155 -0.03504347 -0.03227198
```

Definició 1.1.3 (Transposada d'una matriu). La funció `aperm(a,perm)` es pot usar per a permutar un array `a`. L'argument `perm` ha de ser una permutació d'enters $\{1, \ldots, k\}$, on k és el nombre d'índexs en `a`. El resultat de la funció és un array.

```
A<-seq(7,35,by=3)
dim(A)<-c(5,2)
B<-t(A)
B
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    7   10   13   16   19
## [2,]   22   25   28   31   34
```

Definició 1.1.4 (Multiplicació de matrius)**.** Creem una funció `mult()` que s'inventarà dues matrius, les multiplicarà i retornarà el resultat.

```
m <- matrix(1:8, nrow=2)
n <- matrix(8:15, nrow=2)
r <- m*n
r

##      [,1] [,2] [,3] [,4]
## [1,]    8   30   60   98
## [2,]   18   44   78  120
```

Definició 1.1.5 (Descriptiva de dos vectors)**.** En el següent exemple ensenyarem com podem fabricar les nostres pròpies funcions utilitzant com a exemple el càlcul d'algunes dades d'estadística descriptiva important dels dos vectors.

```
x<-c(43,21,67,77,21,13)
y<-c(9,76,80,1,23,44)
desc <- function(x,y) { c(cor(x,y),var(x,y),sd(x),mean(x)); }
print(desc(x,y))

## [1]   -0.2053553 -184.5333333   26.6733325   40.3333333
```

1.2

Dibuix d'un gràfic

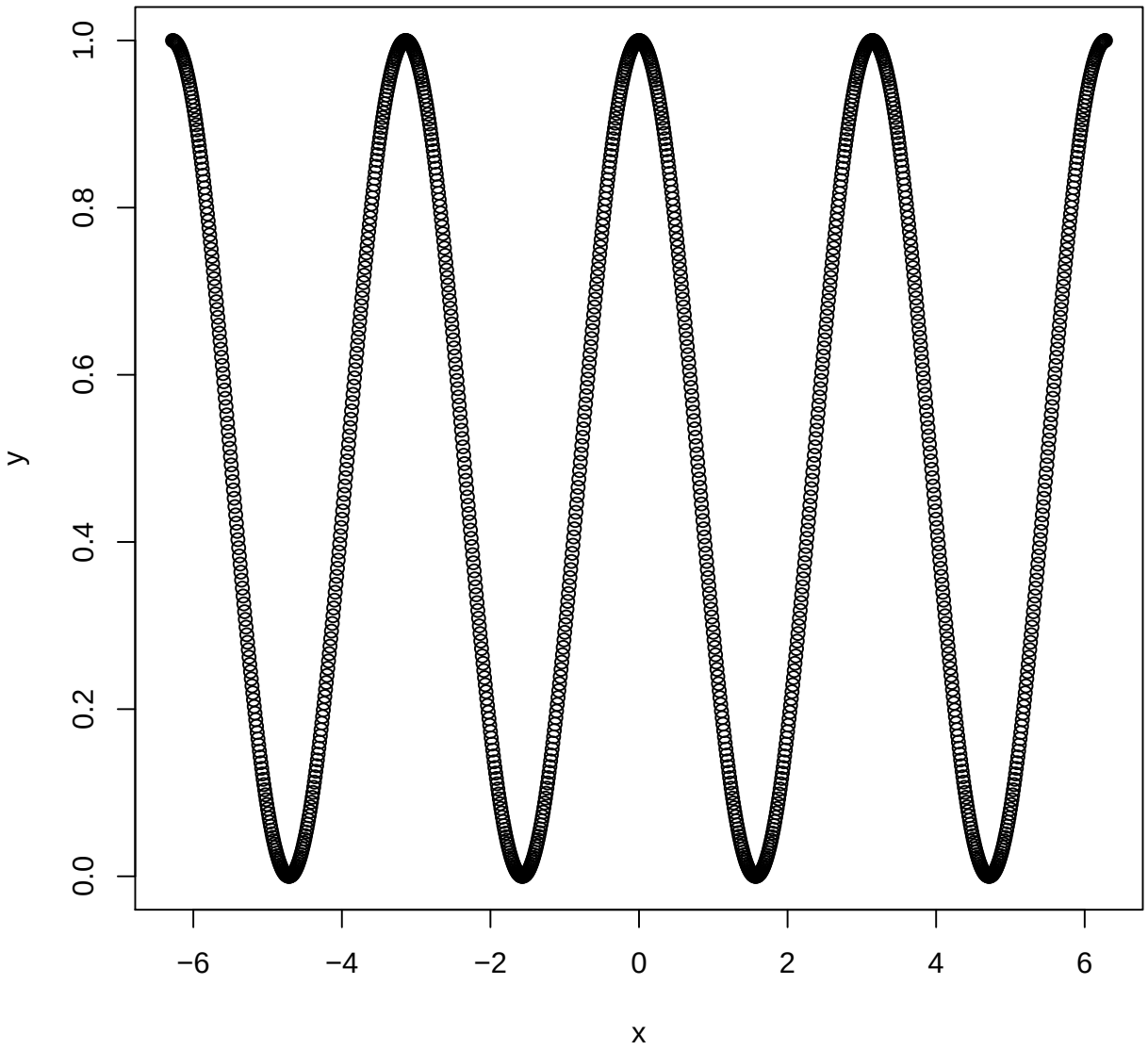
Suposem que volem dibuixar la gràfica del cosinus al quadrat. És molt senzill i ho aconseguim amb poques línies de codi.

Observació 1.2.1. Convé destacar la importància de `seq`, que recull l'interval que hi ha d'haver entre dos punts de la mostra, el qual ha hagut de ser força baix, per evitar la poca precisió i legibilitat de la segona il·lustració. Numèricament:

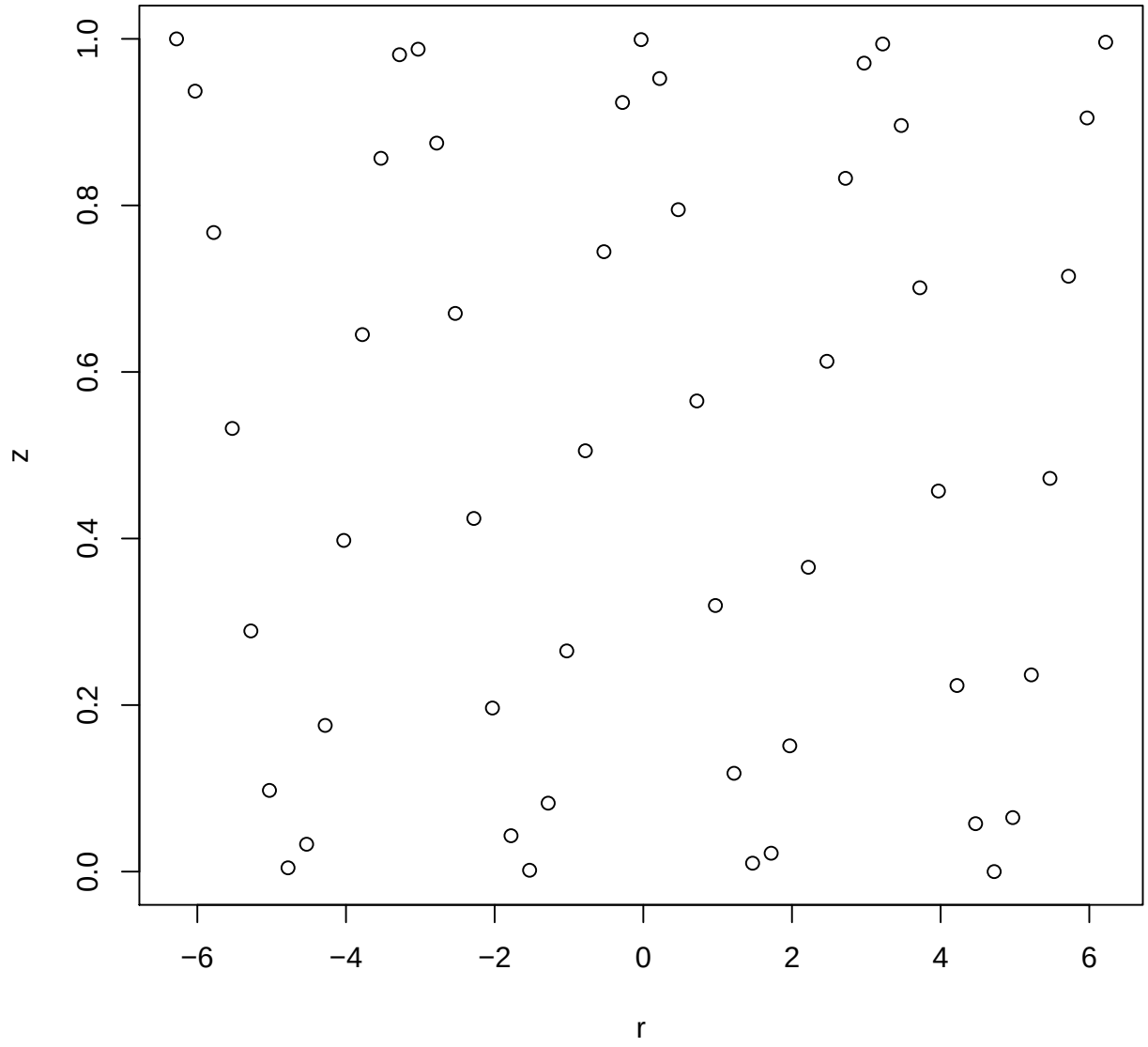
$$\frac{\frac{2\pi}{0.01}}{\frac{2\pi}{0.25}} = 2.5 \implies \times 2.5 \text{ més punts}$$

$$\begin{aligned} \text{Gràfica 1: } & \frac{2\pi}{0.01} \approx 62 \text{ punts}, \\ \text{Gràfica 2: } & \frac{2\pi}{0.25} \approx 25 \text{ punts} \end{aligned} \tag{1.2.1}$$

```
x<-seq(-6.28,6.28,by=0.01)
r<-seq(-6.28,6.28,by=0.25)
y<-cos(x)*cos(x)
z<-cos(r)*cos(r)
plot(x,y,type="p")
```




```
plot(r,z,type="p")
```


z
r
0.0
0.2
0.4
0.6
0.8
1.0
−6
−4
−2
0
2
4
6

II

Cos del treball: estudi d'una variable

2.1

Variable

En aquest capítol analitzarem el nombre de llibres que llegeix cada quinze dies la població, la qual contestarà l'enquesta en clau de mostreig aleatori simple.

Definició 2.1.1 (Variable qualitativa)**.** Es refereixen a una característica no numèrica de l'individu. En principi no s'expressen numèricament, però a la pràctica moltes vegades es codifiquen numèricament per facilitar-ne el tractament.

Definició 2.1.2 (Variable quantitativa)**.** Són les que es refereixen a característiques dels individus que s'expressen numèricament. Dintre d'aquestes en podem trobar de discretes (quan només poden prendre un nombre discret de valors) o de contínues (quan poden prendre qualsevol valor dins d'un interval).

Dit això, és evident que la variable estudiada és quantitativa discreta, ja que el nombre de llibres no pot ser no discret.

Com hem extret la variable? Simplement hem creat un *Google Forms* al qual es pot accedir des del següent LINK. Aquest solament consisteix d'una pregunta de resposta única i obligatòria. Una de les alternatives possibles és crear intervals discrets de mida variable, però ens complicaria lleugerament els càlculs i ens caldria treballar amb marques de classe. Ens inclinarem per treballar amb opcions que continguin un únic valor, excepte l'últim, que recollirà casos d'alta excepcionalitat (de fet, dintre del nostre estudi no hem trobat cap individu que arribi a tal valor).

2.2

Estudi descriptiu

2.2.1 Taula de freqüències

Hem de començar enumerant els diferents valors que pren la nostra variable.

Definició 2.2.1 (Freqüència absoluta)**.** La freqüència absoluta d'un valor és el nombre de repeticions d'aquest valor.

Definició 2.2.2 (Freqüència relativa)**.** És la freqüència absoluta dividida entre la grandària de la mostra. En altres paraules, la proporció d'observacions de cada valor.

Les freqüències absoluta i relativa acumulades es definirien d'una manera totalment anàloga, però tenint en consideració que no són més que això, un acumulat:

Definició 2.2.3 (Freqüència absoluta acumulada)**.** La freqüència absoluta acumulada N_i d'un valor x_i és la freqüència absoluta d'aquesta classe més la de totes les anteriors o, dit d'una altra manera, el nombre d'observacions amb valors menors o iguals a x_i. Tenim que $N_i = n_1 + \cdots + n_i$.

Definició 2.2.4 (Freqüència relativa acumulada)**.** La freqüència relativa acumulada d'un valor x_i és la freqüència relativa d'aquesta classe més la de totes les anteriors o, dit d'una altra manera, la proporció d'observacions amb valors menors o iguals a x_i. Tenim, per tant, $F_i = \frac{N_i}{n}$.

```
x<-read.table("data.csv")
vector<-x[[1]]
median<-median(vector) # mediana
mean<-mean(vector) # mitjana aritmètica
rang<-range(vector) # rang
abs<-table(vector) # taula de freqüència absoluta
absacum<-cumsum(abs) # absoluta acumulada
relat<-prop.table(abs) # relativa
relatacum<-cumsum(relat) # relativa acumulada
variancia<-var(vector) # variància mostral corregida
varem<-function(vector){v<-var(vector); n<-length(vector);v*(n-1)/n}
varem<-varem(vector)
desviacio<-sd(vector) # desviació típica
skew<-2.333928 # s'ha de descarregar la llibreria moments() i executar
               # skewness(vector).
kurt<-10.16807 # curtosi, descarregar moments() i usar kurtosis(vector)
print(c(median,mean,varem,variancia,desviacio,skew,kurt))
```

```
## [1]  0.0000000  0.6790123  0.9586953  0.9706790  0.9852304  2.3339280 10.1680700
```

```
# i ara la taula de freqüència, donada pels vectors columna següents
r<-c(c(0,1,2,3,5),abs,absacum,relat,relatacum)
print(matrix(r,nrow=5,ncol=5))
```

```
##      [,1] [,2] [,3]       [,4]      [,5]
## [1,]    0   43   43 0.53086420 0.5308642
## [2,]    1   28   71 0.34567901 0.8765432
## [3,]    2    7   78 0.08641975 0.9629630
## [4,]    3    1   79 0.01234568 0.9753086
## [5,]    5    2   81 0.02469136 1.0000000
```

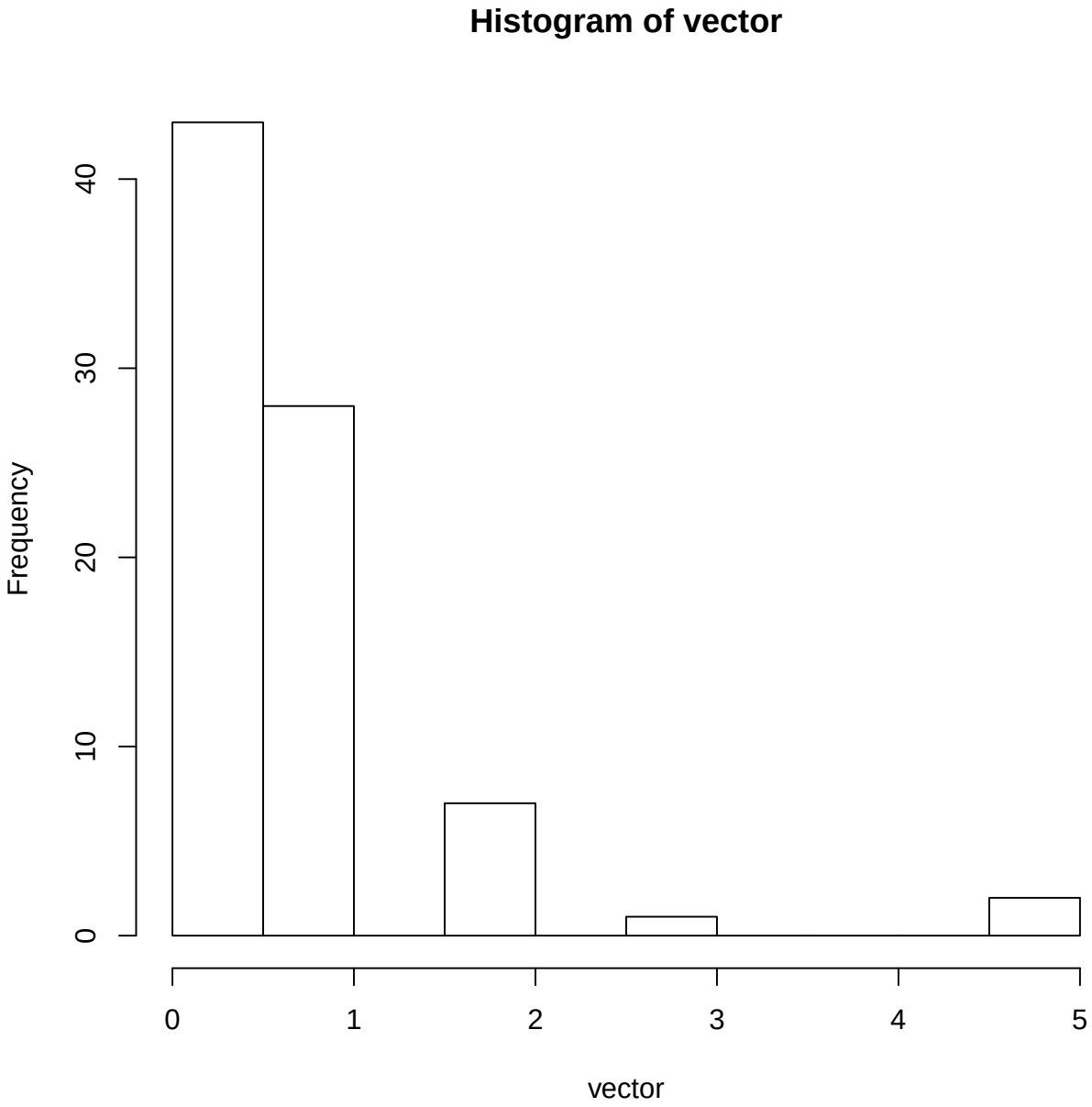
```
# notar que la mitjana harmònica no es pot calcular donat que no tots els
# valors de la mostra són positius.
# finalment, un parell de sumaris i un histograma
summary(vector)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   0.679   1.000   5.000
```

```
fivenum(vector) # Tukey's Five-number Summary
```

```
## [1] 0 0 0 1 5
```

```
hist(vector)
```


Histogram of vector
Frequency
0
10
20
30
40
0
1
2
3
4
5
vector

Bibliografia

[VST+09] William N Venables, David M Smith, R Development Core Team et al. *An introduction to R*. 2009.

[Dav17] Jefferson Davis. "Introduction to R". A: Indiana University Workshop in Methods. 2017.

[Tan17] Teck Kiang Tan. "Introduction to R". A: *Doubly Classified Model with R*. Springer, 2017, pàg. 1 - 19.

[Woo17] Simon N Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.