

Guía nº 7

Ejercicio 1:

a) La función **array** es similar a la función **matrix**. Toma los datos y las dimensiones adecuadas como argumentos y produce el arreglo. Cuando se pasan los datos al **array**, estos se combinan en un vector de manera que la primera dimensión se mueve más rápido, la segunda es la siguiente más rápida y así en adelante. Vea cómo ocurre esto en el siguiente ejemplo.

```
> array(c(1:8, 11:18, 111:118), dim=c(2, 4, 3))
```

```
, , 1
  [,1] [,2] [,3] [,4]
[1,]   1   3   5   7
[2,]   2   4   6   8
```

```
, , 2
  [,1] [,2] [,3] [,4]
[1,]  11  13  15  17
[2,]  12  14  16  18
```

```
, , 3
  [,1] [,2] [,3] [,4]
[1,] 111 113 115 117
[2,] 112 114 116 118
```

La primera dimensión, las filas, aumentan primero. (Observe que es el siguiente orden: (1,1,1) (2,1,1) (1,2,1) (2,2,1) (1,3,1)(2,3,1)(1,4,1)(2,4,1) etc.) Esto equivale a ubicar los valores columna por columna. La segunda dimensión (las columnas) se incrementa en segundo lugar. La tercera dimensión es aumentada llenando una matriz para cada nivel de la tercera dimensión.

c) Vea qué ocurre con la siguiente instrucción:

```
> array(c(rnorm(8), 11:18, 111:118), dim=c(2, 4, 3))
```

y cuando no se le proveen datos a la función array, o sea, por ejemplo:

```
> array(dim=c(2, 4, 3))
```

Ejercicio 2: Estudie el siguiente conjunto de datos. Es un ejemplo de arreglo.

Del **help** obtenemos.

Fisher's Iris Data

SUMMARY:

Array giving 4 measurements on 50 flowers from each of 3 species of iris. Sepal length and width, and petal length and width are measured in centimeters. Species are Setosa, Versicolor, and Virginica. R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", *Annals of Eugenics*, 7, Part II, 1936, pp. 179-188. Republished by permission of Cambridge University Press.

The data were collected by Edgar Anderson, "The irises of the Gaspé Peninsula", *Bulletin of the American Iris Society*, 59, 1935, pp. 2-5.

Un arreglo es una generalización de matrices a más de dos dimensiones. La primera dimensión corresponde a las filas, la segunda a las columnas y la tercera se llena con una matriz en cada nivel.

a) Veamos los atributos del objeto

```
> dim(iris)
50  4  3
```

a1) ¿Qué significan los valores 50, 4 y 3?

```
> dimnames(iris)
[[1]]:
character(0)

[[2]]:
[1] "Sepal L." "Sepal W." "Petal L." "Petal W."
```

a2) La segunda dimensión corresponde a las variables medidas. ¿Cuales son?

```
[[3]]:
[1] "Setosa"      "Versicolor" "Virginica"
```

La tercera dimensión corresponde a las especies.

a3) ¿A qué corresponde la primera dimensión?

b) Fijamos las filas 1:5, de un total de....? ¿Qué observamos?

```
> iris[1:5,,]
, , Setosa
   Sepal L. Sepal W. Petal L. Petal W.
[1,]    5.1     3.5     1.4     0.2
[2,]    4.9     3.0     1.4     0.2
[3,]    4.7     3.2     1.3     0.2
[4,]    4.6     3.1     1.5     0.2
[5,]    5.0     3.6     1.4     0.2
```

```
, , Versicolor
      Sepal L. Sepal W. Petal L. Petal W.
[1,]      7.0      3.2      4.7      1.4
[2,]      6.4      3.2      4.5      1.5
[3,]      6.9      3.1      4.9      1.5
[4,]      5.5      2.3      4.0      1.3
[5,]      6.5      2.8      4.6      1.5
```

```
, , Virginica
      Sepal L. Sepal W. Petal L. Petal W.
[1,]      6.3      3.3      6.0      2.5
[2,]      5.8      2.7      5.1      1.9
[3,]      7.1      3.0      5.9      2.1
[4,]      6.3      2.9      5.6      1.8
[5,]      6.5      3.0      5.8      2.2
```

c) Fijamos la segunda componente de la tercera dimensión, a qué corresponde?

```
> iris[1:5,,2]
      Sepal L. Sepal W. Petal L. Petal W.
[1,]      7.0      3.2      4.7      1.4
[2,]      6.4      3.2      4.5      1.5
[3,]      6.9      3.1      4.9      1.5
[4,]      5.5      2.3      4.0      1.3
[5,]      6.5      2.8      4.6      1.5
```

d) Fijamos la tercera componente de la tercera dimensión, a qué corresponde?

```
> iris[1:5,,3]
      Sepal L. Sepal W. Petal L. Petal W.
[1,]      6.3      3.3      6.0      2.5
[2,]      5.8      2.7      5.1      1.9
[3,]      7.1      3.0      5.9      2.1
[4,]      6.3      2.9      5.6      1.8
[5,]      6.5      3.0      5.8      2.2
```

Pruebe con diferentes valores en cada una de las dimensiones del arreglo.

f) ¿Qué obtenemos ahora?

```
> iris[48,,]
      Setosa Versicolor Virginica
Sepal L.      4.6          6.2          6.5
Sepal W.      3.2          2.9          3.0
Petal L.      1.4          4.3          5.2
Petal W.      0.2          1.3          2.0
```

Ejercicio 3: La función **apply** permite operar en un arreglo utilizando secciones sucesivamente. Los argumentos de esta función son:

- i) El nombre de un arreglo
- ii) Un vector con valores enteros, **MARGIN**, dando los índices de las dimensiones que al fijarlas definen las secciones a las cuales se la va a aplicar la función separadamente.
- iii) La función o el nombre de la función a ser aplicada separadamente a cada sección.
- iv) Cualquier argumento adicional necesario para la función.

a) Calcular la media de las observaciones dadas en **iris**, por especie

```
> apply(iris,3,mean)
      Setosa Versicolor Virginica
      2.5355      3.573      4.285
```

b) Ídem a), por variable

c) Ídem a) por variable y por especie

```
> apply(iris,c(2,3),mean)
      Setosa Versicolor Virginica
Sepal L.  5.006      5.936      6.588
Sepal W.  3.428      2.770      2.974
Petal L.  1.462      4.260      5.552
Petal W.  0.246      1.326      2.026
```

¿Qué obtiene con **> apply(iris,c(3,2),mean)**?

d) Cuando las dimensiones tienen nombre se pueden utilizar para especificar el margen:

```
> names(dimnames(iris))<-c("flor","variable","especie")
> names(dimnames(iris))
[1] "flor"      "variable"  "especie"
```

Las filas corresponden a las variables y las columnas a las especies.

```
> apply(iris,c("variable", "especie" ),mean)
```

```
      Setosa Versicolor Virginica
Sepal L.  5.006      5.936      6.588
Sepal W.  3.428      2.770      2.974
Petal L.  1.462      4.260      5.552
Petal W.  0.246      1.326      2.026
```

e) Fijar una variable es equivalente a fijar una fila en la salida de la función **apply**. Vea que obtiene con la siguientes instrucciones:

```
> data.class(apply(iris,c("variable", "especie" ),mean))
[1] "matrix"
> apply(iris,c("variable", "especie" ),mean)[1,]
      Setosa Versicolor Virginica
5.006      5.936      6.588
```

f) Compare la distribución de cada una de las variables en las distintas especies, obtenga boxplots por variable y por especie. Si es necesario utilice las transformaciones adecuadas para mejorar la comparación.

Vea en qué orden van apareciendo los boxplots que obtiene mediante las siguientes instrucciones e identifique las variables y especies graficadas.

```
> par(mfrow=c(1,3))
> apply(iris,c(3,2),boxplot,ask=T,ylim=c(-1,8))

> apply(iris[,1,],2,boxplot,ask=T,ylim=c(0,8))
> apply(iris[,2,],2,boxplot,ask=T,ylim=c(0,8))
> apply(iris[,3,],2,boxplot,ask=T,ylim=c(0,8))
> apply(liris[,4,],2,boxplot,ask=T,ylim=c(-1,0.5))
```