UNIDAD 1: Práctica 03 - Tipos de objetos: factores, listas y hojas de datos, operadores y funciones que operan sobre ellos

Caterine Melissa Guerrero España

5/8/2022

1. FACTORES NOMINALES Y ORDINALES.

Un factor es un vector utilizado para especificar una clasificación discreta de los elementos de otro vector de igual longitud. En R existen factores nominales y factores ordinales. Los factores son útiles a la hora de querer hacer contrastes o de calcular medidas de resúmenes para variables numéricas en distintos niveles de una segunda variable la cual es no numérica.

FACTORES NOMINALES.

- Ejemplo 1: Variables sexo (categórica) y edad en una muestra de 7 alumnos del curso.

Supongamos que se obtuvieron los siguientes datos:

```
sexo <- c("M", "F", "F", "M", "F", "F", "M")
sexo
```

```
## [1] "M" "F" "F" "M" "F" "F" "M"
```

```
edad <- c(19, 20, 19, 22, 20, 21, 19)
edad
```

```
## [1] 19 20 19 22 20 21 19
```

Podemos construir un factor con los niveles o categorias de sexo

```
FactorSexo = factor(sexo)
FactorSexo
```

```
## [1] M F F M F F M
## Levels: F M
```

Se pueden ver los niveles o categorías del factor con: levels(FactorSexo)

```
levels(FactorSexo)
```

```
## [1] "F" "M"
```

Crear una tabla que contenga la media muestral por categoría de sexo (nivel del factor):

```
mediaEdad<-tapply(edad, FactorSexo, mean)
mediaEdad
```

```
##  F  M
## 20 20
```

Note que el primer argumento debe ser un vector, que es del cual se encontrarán las medidas de resumen; el segundo es el factor que se está considerando, mientras que en el tercero se especifica la medida de interés,

solamente puede hacerse una medida a la vez.

¿Qué tipo de objeto es la variable mediaEdad?

```
is.vector(mediaEdad)
```

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

```
is.matrix(mediaEdad)
```

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

```
is.list(mediaEdad)
```

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

```
is.table(mediaEdad)
```

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

```
is.array(mediaEdad)
```

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

FACTORES ORDINALES

Los niveles de los factores se almacenan en orden alfabético, o en el orden en que se especificaron en la función factor() si ello se hizo explícitamente.

A veces existe una ordenación natural en los niveles de un factor, orden que deseamos tener en cuenta en los análisis estadísticos. La función ordered() crea este tipo de factores y su uso es idéntico al de la función factor(). Los factores creados por la función factor() los denominaremos nominales o simplemente factores cuando no haya lugar a confusión, y los creados por la función ordered() los denominaremos ordinales. En la mayoría de los casos la única diferencia entre ambos tipos de factores consiste en que los ordinales se imprimen indicando el orden de los niveles. Sin embargo, los contrastes generados por los dos tipos de factores al ajustar Modelos lineales, son diferentes.

2. CREACIÓN Y MANEJO DE LISTAS

Una lista es un objeto que contiene una colección ordenada de objetos de diferente tipo (vector, matriz, arreglo, función, o lista), conocidos como componentes. Se construye con la función list(), que tiene la forma general siguiente:

Lista <- list(nombre1 = objeto1, nombre2 = objeto2, . . . , nombren = objeton)

Si omite los nombres, las componentes sólo estarán numeradas. Las componentes pueden accederse por su número o posición, ya que siempre están numeradas, o también pueden referirse por su nombre, si lo tienen.

- Ejemplo 1: Crear una Lista con cuatro componentes.

```
lista1<-list(padre="Pedro", madre="María", no.hijos=3, edad.hijos=c(4,7,9))

lista1
```

```
## $padre
## [1] "Pedro"
##
## $madre
## [1] "María"
##
```

```
## $no.hijos
## [1] 3
##
## $edad.hijos
## [1] 4 7 9
```

Revise algunos tipos como:

```
is.matrix(lista1)
```

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

```
is.vector(lista1$edad.hijos)
```

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

- Ejemplo 2: Acceso a las componentes de una lista:

```
lista1[1] #accede a la componente como una lista (con etiqueta y valor)
```

```
## $padre
## [1] "Pedro"
```

```
lista1["padre"] #el acceso es igual que con lista1[1]
```

```
## $padre
## [1] "Pedro"
```

```
lista1[[2]] #accede al valor o valores de la componente segunda pero no muestra el nombre de la compone
```

```
## [1] "María"
```

```
lista1["madre"] #el acceso es igual que con lista1[[1]]
```

```
## $madre
## [1] "María"
```

- Ejemplo 3: Acceso a los elementos de la cuarta componente:

```
lista1[[4]][2] #se indica el elemento a ingresar en el segundo corchete)
```

```
## [1] 7
```

- Ejemplo 4: Acceso de las componentes de una lista por su nombre:

```
lista1$padre #similar a lista1["padre"].
```

```
## [1] "Pedro"
```

Forma general: **Nombre_de_lista$nombre_de_componente ** Por ejemplo : *lista1padre equivale a lista1[[1]]; y lista1$edad.hijos[2] equivale a lista1[[4]][2]*

- Ejemplo 5: Utilizar el nombre de la componente como índice:

```
lista1[["padre"]]
```

```
## [1] "Pedro"
```

```
lista1
```

```
## $padre
## [1] "Pedro"
##
## $madre
## [1] "María"
```

```
##
## $no.hijos
## [1] 3
##
## $edad.hijos
## [1] 4 7 9
```

se puede ver que equivale a lista1$nombre También es útil la forma:

```
x<-"padre"
lista1[x]
```

```
## $padre
## [1] "Pedro"
```

- Ejemplo 6: Creación de una sublista de una lista existente:

```
subLista <- lista1[4]
subLista
```

```
## $edad.hijos
## [1] 4 7 9
```

- Ejemplo 7: Ampliación de una lista: por ejemplo, la lista lista1 tiene 4 componentes y se le puede agregár una quinta componente con:

```
lista1[5]<-list(sexo.hijos=c("F", "M", "F"))
lista1
```

```
## $padre
## [1] "Pedro"
##
## $madre
## [1] "María"
##
## $no.hijos
## [1] 3
##
## $edad.hijos
## [1] 4 7 9
##
## [[5]]
## [1] "F" "M" "F"
```

Observe que no aparece el nombre del objeto agregado, pero usted puede modificar la estructura de la lista lista1 con:

```
#lista1<-edit(lista1)
```

- Ejemplo 8: Funciones que devuelven una lista. Las funciones y expresiones de R devuelven un objeto como resultado, por tanto, si deben devolver varios objetos, previsiblemente de diferentes tipos, la forma usual es una lista con nombres. Por ejemplo, la función eigen() que calcula los autovalores y autovectores de una matriz simétrica.

Ejecute las siguientes instrucciones:

```
S<-matrix(c(3, -sqrt(2), -sqrt(2), 2), nrow=2, ncol=2)
S
```

```
##           [,1]      [,2]
## [1,]  3.000000 -1.414214
```

```
## [2,] -1.414214  2.000000
```

```
autovS<-eigen(S)
autovS
```

```
## eigen() decomposition
## $values
## [1] 4 1
##
## $vectors
##            [,1]       [,2]
## [1,] -0.8164966 -0.5773503
## [2,]  0.5773503 -0.8164966
```

Observe que la función eigen() retorna una lista de dos componentes, donde la componente *autovSvalueseselvectordeautovaloresdeSylacomponenteautovS*vectors es la matriz de los correspondientes autovectores. Si quisiéramos almacenar sólo los autovalores de S, podemos hacer lo siguiente:

```
evals<-eigen(S)$values
evals
```

```
## [1] 4 1
```

- Ejemplo 9: Crear una matriz dando nombres a las filas y columnas

```
Notas<-matrix(c(2,5,7,6,8,2,4,9,10), ncol=3, dimnames=list(c("Matemática","Álgebra","Geometría"),c("Jua
Notas # Los nombres se dan primero para filas y luego para columnas.
```

```
##            Juan José René
## Matemática    2    6    4
## Álgebra       5    8    9
## Geometría     7    2   10
```

3. CREACIÓN Y MANEJO DE HOJAS DE DATOS (DATA FRAME).

Una hoja de datos (data frame) es una lista que pertenece a la clase "data.frame". Un data.frame puede considerarse como una matriz de datos.

- Ejemplo 1: Creación de un data frame teniendo como columnas tres vectores:

En primer lugar generamos los tres vectores El primer vector tendrá 20 elementos que se obtienen con reemplazamiento de una muestra aleatoria de valores lógicos.

```
log <- sample(c(TRUE, FALSE), size = 20, replace = T)
log #Note que puede usar T en lugar de TRUE y F en lugar de FALSE.
```

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

El segundo vector tendrá 20 elementos de valores complejos cuya parte real proviene de una distribución Normal estándar y cuya parte imaginaria lo hace de una distribución Uniforme(0,1)

```
comp<-rnorm(20)+runif(20)*(1i)
comp
```

```
##  [1] -0.3776473+0.4620549i  0.9616472+0.9273167i  0.7915350+0.4829094i
##  [4]  1.6470198+0.8462496i  0.5841658+0.7463826i -0.4297406+0.3624569i
##  [7] -0.5715337+0.2422898i  0.1577112+0.2249983i -0.5903280+0.7787532i
```

```
## [10]  1.6076657+0.4699216i -0.3086783+0.4388814i -0.3499365+0.2725107i
## [13]  0.2852602+0.5948103i  0.2138867+0.7686664i -1.5692128+0.3036702i
## [16] -0.0082473+0.2114709i  0.6935598+0.8385299i  0.3676007+0.0367896i
## [19]  1.0689762+0.5272950i  0.3010246+0.8932596i
```

El tercer vector tendrá 20 elementos de una distribución Normal estándar

```
num <- rnorm(20, mean=0, sd=1)
num
```

```
##  [1] -1.30776504  0.59014302  0.62155932 -0.20895158  1.89437370 -0.02167945
##  [7]  1.20676401 -0.15841130 -1.14325136  0.60891254  1.50417672 -1.48588464
## [13]  0.73271926  1.23394354 -0.81426194 -1.38061975 -1.37004399  0.18539551
## [19] -1.25639920 -1.13201664
```

Crear un data frame compuesto por los tres vectores anteriores

```
df1 <- data.frame(log, comp, num)
df1
```

```
##      log                  comp         num
## 1   TRUE -0.3776473+0.4620549i -1.30776504
## 2  FALSE  0.9616472+0.9273167i  0.59014302
## 3   TRUE  0.7915350+0.4829094i  0.62155932
## 4  FALSE  1.6470198+0.8462496i -0.20895158
## 5   TRUE  0.5841658+0.7463826i  1.89437370
## 6   TRUE -0.4297406+0.3624569i -0.02167945
## 7  FALSE -0.5715337+0.2422898i  1.20676401
## 8   TRUE  0.1577112+0.2249983i -0.15841130
## 9  FALSE -0.5903280+0.7787532i -1.14325136
## 10  TRUE  1.6076657+0.4699216i  0.60891254
## 11 FALSE -0.3086783+0.4388814i  1.50417672
## 12  TRUE -0.3499365+0.2725107i -1.48588464
## 13  TRUE  0.2852602+0.5948103i  0.73271926
## 14  TRUE  0.2138867+0.7686664i  1.23394354
## 15  TRUE -1.5692128+0.3036702i -0.81426194
## 16 FALSE -0.0082473+0.2114709i -1.38061975
## 17  TRUE  0.6935598+0.8385299i -1.37004399
## 18 FALSE  0.3676007+0.0367896i  0.18539551
## 19 FALSE  1.0689762+0.5272950i -1.25639920
## 20  TRUE  0.3010246+0.8932596i -1.13201664
```

Crear un vector de nombres de los tres vectores anteriores

```
nombres<-c("logico", "complejo", "numerico")
```

Define los nombres de las columnas del data frame asignándoles el vector nombres

```
names(df1) <- nombres
df1
```

```
##    logico              complejo    numerico
## 1    TRUE -0.3776473+0.4620549i -1.30776504
## 2   FALSE  0.9616472+0.9273167i  0.59014302
## 3    TRUE  0.7915350+0.4829094i  0.62155932
## 4   FALSE  1.6470198+0.8462496i -0.20895158
## 5    TRUE  0.5841658+0.7463826i  1.89437370
## 6    TRUE -0.4297406+0.3624569i -0.02167945
## 7   FALSE -0.5715337+0.2422898i  1.20676401
```

```
## 8    TRUE  0.1577112+0.2249983i -0.15841130
## 9   FALSE -0.5903280+0.7787532i -1.14325136
## 10   TRUE  1.6076657+0.4699216i  0.60891254
## 11  FALSE -0.3086783+0.4388814i  1.50417672
## 12   TRUE -0.3499365+0.2725107i -1.48588464
## 13   TRUE  0.2852602+0.5948103i  0.73271926
## 14   TRUE  0.2138867+0.7686664i  1.23394354
## 15   TRUE -1.5692128+0.3036702i -0.81426194
## 16  FALSE -0.0082473+0.2114709i -1.38061975
## 17   TRUE  0.6935598+0.8385299i -1.37004399
## 18  FALSE  0.3676007+0.0367896i  0.18539551
## 19  FALSE  1.0689762+0.5272950i -1.25639920
## 20   TRUE  0.3010246+0.8932596i -1.13201664
```

Define los nombres de las filas del data frame asignándoles un vector de 20 elementos correspondientes a las 20 primeras letras del abecedario

```
row.names(df1) <- letters[1:20]
df1
```

```
##   logico              complejo    numerico
## a   TRUE -0.3776473+0.4620549i -1.30776504
## b  FALSE  0.9616472+0.9273167i  0.59014302
## c   TRUE  0.7915350+0.4829094i  0.62155932
## d  FALSE  1.6470198+0.8462496i -0.20895158
## e   TRUE  0.5841658+0.7463826i  1.89437370
## f   TRUE -0.4297406+0.3624569i -0.02167945
## g  FALSE -0.5715337+0.2422898i  1.20676401
## h   TRUE  0.1577112+0.2249983i -0.15841130
## i  FALSE -0.5903280+0.7787532i -1.14325136
## j   TRUE  1.6076657+0.4699216i  0.60891254
## k  FALSE -0.3086783+0.4388814i  1.50417672
## l   TRUE -0.3499365+0.2725107i -1.48588464
## m   TRUE  0.2852602+0.5948103i  0.73271926
## n   TRUE  0.2138867+0.7686664i  1.23394354
## o   TRUE -1.5692128+0.3036702i -0.81426194
## p  FALSE -0.0082473+0.2114709i -1.38061975
## q   TRUE  0.6935598+0.8385299i -1.37004399
## r  FALSE  0.3676007+0.0367896i  0.18539551
## s  FALSE  1.0689762+0.5272950i -1.25639920
## t   TRUE  0.3010246+0.8932596i -1.13201664
```

- Ejemplo 2: Vamos a crear la siguiente hoja de datos que tiene 4 variables o columnas:

```
edad <- c(18, 21, 45, 54)
edad
```

```
## [1] 18 21 45 54
```

```
datos <- matrix(c(150, 160, 180, 205, 65, 68, 65, 69), ncol=2, dimnames=list(c(), c("Estatura","Peso")))
datos
```

```
##      Estatura Peso
## [1,]      150   65
## [2,]      160   68
## [3,]      180   65
## [4,]      205   69
```

```r
sexo <- c("F", "M", "M", "M")
sexo
```

```
## [1] "F" "M" "M" "M"
```

```r
hoja1<-data.frame(Edad=edad, datos, Sexo=sexo)
hoja1
```

```
##   Edad Estatura Peso Sexo
## 1   18      150   65    F
## 2   21      160   68    M
## 3   45      180   65    M
## 4   54      205   69    M
```