

DATA FRAMES

Ramon Ceballos

22/1/2021

Construcción de Data Frame

Se pueden crear data frames a través de la siguiente función y vectores de observaciones que generarán los vectores.

- `data.frame(vector_1,...,vector_n)`: para construir un data frame a partir de vectores introducidos en el orden en el que queremos disponer las columnas de la tabla
 - R considera del mismo tipo de datos todas las entradas de una columna de un data frame
 - Las variables tomarán los nombres de los vectores. Estos nombres se pueden especificar en el argumento de `data.frame` entrando una construcción de la forma `nombre_variable = vector`
 - `rownames`: para especificar los identificadores de las filas
 - También en esta función podemos hacer uso del parámetro `stringsAsFactors` para evitar la transformación de las columnas de tipo palabra en factores

Ejemplos de construcción de Data Frame

```
Algebra = c(1,2,0,5,4,6,7,5,5,8)
Analysis = c(3,3,2,7,9,5,6,8,5,6)
Statistics = c(4,5,4,8,8,9,6,7,9,10)
grades = data.frame(Alg = Algebra, An = Analysis, Stat = Statistics)
str(grades)
```

```
## 'data.frame':    10 obs. of  3 variables:
## $ Alg : num  1 2 0 5 4 6 7 5 5 8
## $ An : num  3 3 2 7 9 5 6 8 5 6
## $ Stat: num  4 5 4 8 8 9 6 7 9 10
```

```
genero=c("Hombre", "Hombre", "Hombre", "Hombre", "Mujer", "Mujer", "Mujer", "Mujer")
edad=c(12,34,63,33,22,27,90,15)
soltero=c("yes","no","no","yes","yes","no","no","yes")
citizens =data.frame(gender=genero,age=edad,single=soltero,stringsAsFactors = TRUE)
rownames(citizens)=c("p1","p2","p3","p4","p5","p6","p7","p8")
str(citizens)
```

```
## 'data.frame':    8 obs. of  3 variables:
## $ gender: Factor w/ 2 levels "Hombre","Mujer": 1 1 1 1 2 2 2 2
## $ age : num  12 34 63 33 22 27 90 15
## $ single: Factor w/ 2 levels "no","yes": 2 1 1 2 2 1 1 2
```

```
citizens
```

```
##      gender age single
## p1 Hombre  12    yes
## p2 Hombre  34    no
## p3 Hombre  63    no
## p4 Hombre  33    yes
## p5 Mujer   22    yes
## p6 Mujer   27    no
## p7 Mujer   90    no
## p8 Mujer   15    yes
```

Hay más funciones que pueden variar los data frames.

- `fix(d.f)`: para crear / editar un data frame con el editor de datos
- `names(d.f)`: para cambiar los nombres de las variables
- `rownames(d.f)`: para modificar los identificadores de las filas. Han de ser todos diferentes
- `dimnames(d.f)=list(vec_nom_fil, vec_nom_col)`: para modificar el nombre de las filas y de las columnas simultáneamente
- `d.f[núm_fila,] = c(...)`: para añadir una fila a un data frame
 - Las filas que añadimos de esta manera son vectores, y por tanto sus entradas han de ser todas del mismo tipo
 - Si no añadimos las filas inmediatamente siguientes a la última fila del data frame, los valores entre su última fila y las que añadimos quedarán no definidos y aparecerán como NA
 - Para evitar el problema anterior, vale más usar la función `rbind()` para concatenar el data frame con la nueva fila
- `d.f$new_var`: para añadir una nueva variable al data frame
 - Podemos concatenar columnas con un data frame existente mediante la función `cbind()`. De este modo se puede añadir la columna directamente sin necesidad de convertirla antes a data frame
 - Esta nueva variable ha de tener la misma longitud que el resto de columnas del data frame original. Si no, se añadirán valores NA a las variables del data frame original o a la nueva variable hasta completar la misma longitud

```
gender = c("H", "M", "M", "M", "H")
age = c( 23, 45, 20, 30, 18)
family = c( 2, 3, 4, 2, 5)
df5 = data.frame(genero = gender, edad = age, familia = family, stringsAsFactors = TRUE)
rownames(df5) = c("P1", "P2", "P3", "P4", "P5")
df5
```

```
##      genero edad familia
## P1        H   23       2
## P2        M   45       3
## P3        M   20       4
## P4        M   30       2
## P5        H   18       5
```

```
str(df5)
```

```
## 'data.frame': 5 obs. of 3 variables:
## $ genero : Factor w/ 2 levels "H","M": 1 2 2 2 1
## $ edad : num 23 45 20 30 18
## $ familia: num 2 3 4 2 5
```

```
dimnames(df5) = list(
  c("Antonio", "Ricardo", "JuanGabriel", "María", "Margarita"),
  c("Sexo", "Años", "MiembrosFamilia")
)
df5
```

```
##           Sexo Años MiembrosFamilia
## Antonio      H   23                2
## Ricardo      M   45                3
## JuanGabriel  M   20                4
## María        M   30                2
## Margarita    H   18                5
```

```
#Añadir una fila, mejor opción para adicionar (rbind)
df5 = rbind(df5, c("H", 30, 1))
df5
```

```
##           Sexo Años MiembrosFamilia
## Antonio      H   23                2
## Ricardo      M   45                3
## JuanGabriel  M   20                4
## María        M   30                2
## Margarita    H   18                5
## 6            H   30                1
```

```
#convertir en carácter un factor
df5$Sexo = as.character(df5$Sexo)
#Definir una nueva columna para el data frame
df5$Ingresos = c(10000, 12000, 15000, 12000, 20000, 10000)
df5
```

```
##           Sexo Años MiembrosFamilia Ingresos
## Antonio      H   23                2  10000
## Ricardo      M   45                3  12000
## JuanGabriel  M   20                4  15000
## María        M   30                2  12000
## Margarita    H   18                5  20000
## 6            H   30                1  10000
```