

TALLER 2 DE R ESTUDIO

- **Matrices**

En R hay varias formas de crear una matriz:

Mediante la función **matrix()**, cuyos parámetros son:

- **data** Vector que contiene los valores que formarán la matriz. Debe tener en cuenta que si no es suficientemente grande, se repetirá las veces que sea necesario
- **nrow** Número de filas. Si no especifica, se toma **nrow =1**
- **ncol** Número de columnas
- **byrow** Variable lógica que indica si la matriz debe construirse por filas o por columnas. El valor predeterminado es F
- **dimnames** Lista de longitud 2 con los nombres de las filas y las columnas.
- Mediante los operadores **rbind()** (para pegar vectores por filas) y **cbind()** (para pegar vectores por columnas)

matriz = rbind (x1, x2,...)

matriz = cbind (x1, x2...)

Instrucciones

M = matrix (1: 9, nrow = 3, byrow = TRUE) # la matriz se rellena por filas

M

N = matrix(1: 9, nrow = 3, byrow = FALSE) # la matriz se rellena por columnas

> N

También podemos utilizar las funciones cbind y rbind para construir matrices:

- v1 = c(1,2,3,4)
> v2 = c(5,6,7,8)
> v3= c(9,10,11,12)
> FA = rbind(v1,v2,v3)
> FA

Construcción de una base de datos

A continuación se construirá la primera base de datos a partir de tres variables. Para esto, como se observa en los siguientes comandos, se parte por la construcción de tres variables de 9 casos cada una:

- Género. Variable nominal con valores 1 y 2, que representan las categorías de respuesta “hombre” y “mujer”.
- Ingreso. Variable de razón con valores arbitrarios de ingreso monetario.
- Acuerdo en torno al aborto libre. Variable ordinal tipo escala Likert con valores 1, 2, 3, 4 o 5, que representan las categorías “nada de acuerdo”, “un poco de acuerdo”, “ni de acuerdo ni en desacuerdo”, “bastante de acuerdo”, “muy de acuerdo”.

#Creación de las variables: todas tienen la misma cantidad de casos

```
genero <- c(1,1,2,1,2,2,2,1,2)
```

```
ingreso <- c(100000,300000,500000,340000,300000,500000,650000,410000,750000)
```

```
acuerdo <- c(1,1,3,2,4,1,5,3,2)
```

A partir de las variables ya creadas se puede construir una base de datos. Para esto se utiliza el comando `data.frame` asignando su resultado al objeto `aborto` que contendrá la base de datos construida.

Si la ejecución del comando es exitosa se verá un nuevo objeto de tipo `data` en el entorno de trabajo, donde además se indicará la dimensión de la base de datos (cantidad de casos y variables).

```
aborto <- data.frame(genero, ingreso, acuerdo)
```

Luego de construida esta base de datos interesa guardarla como un archivo reutilizable para posteriores análisis. Para ello es útil el comando `save`: indicando el nombre del objeto a guardar como archivo y definiendo también el nombre del archivo. Como se ve en la siguiente línea de comando, el nombre del archivo resultante se indica con el argumento `file = "nombre_archivo.extensión"` (el nombre del archivo va entre comillas).

```
save(aborto, file = "aborto.RData") #Se indica primero el objeto a guardar  
                                     #y luego el nombre del archivo, entre  
                                     comillas.
```

Taller 1 datos

Elaborar el data frame

Elaborar la regresión lineal de:

Peso vs altura

Edad vs altura

Edad vs peso

Peso Altura edad

77 1.63 23

58 1.63 23

89 1.85 26

55 1.62 23

47 1.60 26

60 1.63 26

54 1.70 22

58 1.65 23

75 1.78 26

65 1.70 24

82 1.77 28

85 1.83 42

75 1.74 25

65 1.65 26

LIMPIAR DATOS

```
id <- c(1,4,3,NA,7,6,9,4,0,8)
```

```
Texto <- c("a", "b", NA, "NA","a","b","b","b","c","d")
```

```
valor <- c(2,8,7,5,1,9,4,3,7,2)
```

```
Datos <- data.frame (id=id, texto=texto, valor=valor)
```

```
Head (datos, n=10)
```

Taller limpieza de datos

Estadísticos descriptivos con R-Studio

Tabla de Frecuencias

Para crear tablas de frecuencia en *R* y *R Studio* se emplea la función **table** o la función **prop.table**, dependiendo de si la tabla muestra las frecuencias absolutas o las frecuencias relativas. La sintaxis de estas órdenes es la siguiente:

```
> table(x) # para frecuencias absolutas
```

```
> prop.table(tab) # para las frecuencias relativas
```

La principal diferencia entre las dos funciones reside en el tipo de los argumentos que necesita cada una.

- **table** construye la tabla de frecuencias absolutas a partir de la variable que recibe como argumento
- **prop.table** recibe como argumento una tabla o una matriz que representa una tabla de frecuencias absolutas, y a partir de ella construye la tabla de frecuencias relativas asociada. Es decir, **prop.table** recibe como argumento el resultado que devuelve la función **table**.

Taller base de datos se de datos en Excel