

Cheat sheet Práctica 2: Generación de nuevas variables con R. Ordenación de casos.

Juan Antonio Villatoro García – Grupo BioestadísticaR



**UNIVERSIDAD
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo BioestadísticaR formado por:

Antonio Martín Andrés

Juan de Dios Luna del Castillo,

Pedro Femia Marzo,

Miguel Ángel Montero Alonso,

Christian José Acal González,

Pedro María Carmona Sáez,

Juan Manuel Melchor Rodríguez,

José Luis Romero Béjar,

Manuela Expósito Ruíz,

Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de BioestadísticaR.

Todos los materiales están protegidos por la Licencia Creative Commons CC BY-NC-ND que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Práctica 2: : Generación de nuevas variables con R. Ordenación de casos



Recodificación de variables

- Se usa en casos en que los valores de las variables sean difíciles de interpretar estadísticamente
- Consiste en asignar números o nuevas palabras a categorías en función de la necesidad
- 3 posibles casos:
 - Recodificación entre variable tipo factor a tipo factor
 - Alterar orden de los niveles
 - Agrupar valores numéricos

Ejemplo: asignamos No cuando no bebe y Sí cuando bebe algo:

```
> library(foreign)
> osteo=read.spss("osteo.sav",to.data.frame = TRUE)
> alcohol <- osteo$alcohol
> levels(alcohol)
[1] "No"      "Moderado" "Excesivo"
> alcohol.sn <- vector()
> alcohol.sn[alcohol=="No"]=0
> alcohol.sn[alcohol=="Moderado" | alcohol=="Excesivo"]=1
> alcohol.sn=factor(alcohol.sn,labels=c("No","Sí"))
> alcohol.sn
 [1] Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí No
 [18] Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí
 [35] Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí No Sí Sí Sí No No No Sí
 [52] Sí Sí Sí No Sí Sí Sí Sí Sí Sí Sí Sí No No No Sí No No
 [69] Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí Sí No Sí Sí Sí Sí
 [86] Sí Sí No Sí Sí Sí Sí Sí No
Levels: No Sí
```

Manejo de datos faltantes

- Información desconocida dentro de una variable
- Denotados como NA

Orden	SIGNIFICADO
which(x)	Indica las posiciones de un objeto que cumple una condición
is.na(x)	Señala como TRUE aquellos valores faltantes dentro de un objeto

Ejemplo: posiciones de los valores faltantes en la variable pthi del conjunto de datos osteo

```
> head(osteo$pthi)
[1] 49.0 NA 54.3 NA 49.8 59.9
> which(is.na(osteo$pthi))
 [1]  2  4  8  9 11 14 15 17 23 25 26 27 30 31 32 33 35
 [18] 57 62 63 84 86 88 89 91 92 93
```

Ordenar casos

- La función sort() en los data.frame solo ordena una variable permaneciendo las otras variables (información) inamovible
- La función order() permite ordenar la información de un data.frame en función de la una variable o varias variables.

Ejemplo: ordenamos este data.frame en función de la variable edad

```
> edad=c(26,32,29,26,27,25,27)
> peso=c(64.75,55.25,61.20,58.75,63.30,59.80,51.30)
> altura=c(1.80,1.65,1.71,1.62,1.71,1.83,1.63)
> datos_original=data.frame(edad,peso,altura)
> datos_original
  edad peso altura
1  26 64.75  1.80
2  32 55.25  1.65
3  29 61.20  1.71
4  26 58.75  1.62
5  27 63.30  1.71
6  25 59.80  1.83
7  27 51.30  1.63
> datos.ordenados_e=datos_original[order(edad),]
> datos.ordenados_e
  edad peso altura
6  25 59.80  1.83
1  26 64.75  1.80
4  26 58.75  1.62
5  27 63.30  1.71
7  27 51.30  1.63
3  29 61.20  1.71
2  32 55.25  1.65
```

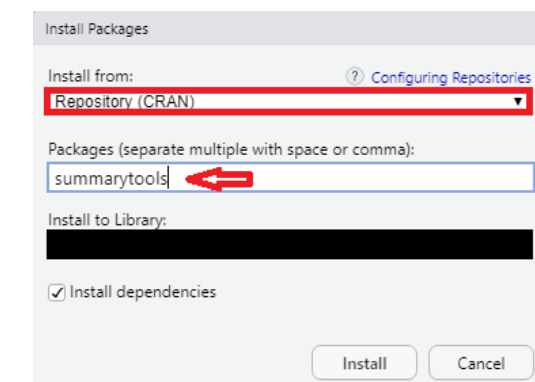
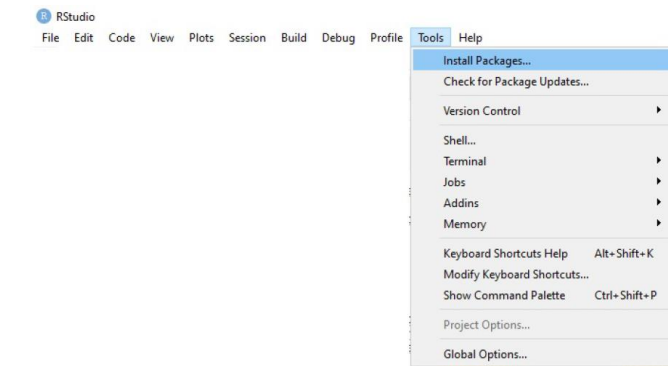
Ejemplo: ordenamos este data.frame en función de la variable edad y peso.

```
> datos.ordenados_e_p=datos_original[order(edad,peso),]
> datos.ordenados_e_p
  edad peso altura
6  25 59.80  1.83
4  26 58.75  1.62
1  26 64.75  1.80
7  27 51.30  1.63
5  27 63.30  1.71
3  29 61.20  1.71
2  32 55.25  1.65
```

Instalación de librerías

Librería externa

- Función **install.packages()** e introducir el nombre del paquete
- En el menú **Packages** (panel 2 de Rstudio) → botón **Install packages** → Escribimos el nombre del paquete



Librería interna

- En el menú **Packages** (panel 2 de Rstudio) → botón **Install package** → **Packages Archive File (zip, tar.gz)** → **Seleccionamos la librería interna (Browse)**

