



Introducción a R

Mathieu Kessler

Departamento de Matemática Aplicada y Estadística
Universidad Politécnica de Cartagena

Cartagena, Enero 2010



Guión

- 1 Qué es R.
- 2 Estructura de R
- 3 El entorno de trabajo en R:
 - Línea de comandos
 - Ficheros asociados a R
 - Directorio de trabajo
- 4 Los principales objetos en R
 - Vectores
 - Los conjuntos de datos (data frames)



Guión

- 1 Qué es R.
- 2 Estructura de R
- 3 El entorno de trabajo en R:
 - Línea de comandos
 - Ficheros asociados a R
 - Directorio de trabajo
- 4 Los principales objetos en R
 - Vectores
 - Los conjuntos de datos (data frames)



Un entorno para llevar a cabo de manera eficiente cálculos estadísticos y realizar gráficos



Un entorno para llevar a cabo de manera eficiente cálculos estadísticos y realizar gráficos

- R es libre (Licencia GPL) <http://www.r-project.org/>



Un entorno para llevar a cabo de manera eficiente cálculos estadísticos y realizar gráficos

- R es libre (Licencia GPL) <http://www.r-project.org/>
- Es extensible: tiene un núcleo (componente base) y un gran número de paquetes (librerías) adicionales.



Un entorno para llevar a cabo de manera eficiente cálculos estadísticos y realizar gráficos

- R es libre (Licencia GPL) <http://www.r-project.org/>
- Es extensible: tiene un núcleo (componente base) y un gran número de paquetes (librerías) adicionales.
- Muchos recursos y documentación libres.



Un poco de historia

- El programa S: Finales de los 80, en los laboratorios Bell.
- Una versión comercial: S plus.
- R empieza en 1995, por R. Gentleman & R. Ihaka, University of Auckland, como un dialecto de S.
- Casi total compatibilidad con S-plus.



Un poco de historia

- El programa S: Finales de los 80, en los laboratorios Bell.
- Una versión comercial: S plus.
- R empieza en 1995, por R. Gentleman & R. Ihaka, University of Auckland, como un dialecto de S.
- Casi total compatibilidad con S-plus.

Ahora

- El entorno predilecto de los estadísticos, y cada vez más presente en aplicaciones.
- Un equipo de desarrolladores voluntarios
- Una enorme comunidad internacional.



Guión

- 1 Qué es R.
- 2 Estructura de R
- 3 El entorno de trabajo en R:
 - Línea de comandos
 - Ficheros asociados a R
 - Directorio de trabajo
- 4 Los principales objetos en R
 - Vectores
 - Los conjuntos de datos (data frames)



La principal fuente de información:

<http://www.cran.r-project.org/>

R consiste de dos tipos de componentes:

- El núcleo base de R. (base)
- Los paquetes contribuidos. (contrib)



La principal fuente de información:

<http://www.cran.r-project.org/>

R consiste de dos tipos de componentes:

- El núcleo base de R. (base)
 - Los paquetes contribuidos. (contrib)
- 1 Es necesario instalar el núcleo base.
 - 2 Es opcional instalar paquetes contribuidos.
 - 3 El núcleo base se instala descargando e ejecutando el archivo R-2.10.0-win32.exe desde la página de CRAN.
 - 4 Los paquetes contribuidos se pueden descargar e instalar desde el programa R directamente (si se dispone de conexión internet).



Guión

- 1 Qué es R.
- 2 Estructura de R
- 3 El entorno de trabajo en R:
 - Línea de comandos
 - Ficheros asociados a R
 - Directorio de trabajo
- 4 Los principales objetos en R
 - Vectores
 - Los conjuntos de datos (data frames)



Línea de comandos

Al ejecutar R, tenemos directamente acceso a la línea de comandos. Podemos usarla como una simple calculadora:

```
> 1 + 3
```

```
[1] 4
```

Definir constantes y hacer referencia a ellas:

```
> x = 1
```

```
> x + 3
```

```
[1] 4
```

```
> exp(x)
```

```
[1] 2.718282
```



Un truco muy útil:

En la línea de comandos, podemos recorrer los comandos ya introducimos usando las flechas \uparrow y \downarrow



Ficheros asociados a R

- 1 **Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.



Ficheros asociados a R

- 1 **Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones



Ficheros asociados a R

- 1 **Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es .R



Ficheros asociados a R

- 1 **Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es `.R`
 - Hay un sencillo editor con R para ello (ver también Tinn-R)



Ficheros asociados a R

- 1 Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es .R
 - Hay un sencillo editor con R para ello (ver también Tinn-R)
- 2 Ficheros Workspaces** (área de trabajo). Contiene distintos objetos de R, (conjuntos de datos, variables, constantes, etc...)



Ficheros asociados a R

- 1 Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es `.R`
 - Hay un sencillo editor con R para ello (ver también Tinn-R)
- 2 Ficheros Workspaces** (área de trabajo). Contiene distintos objetos de R, (conjuntos de datos, variables, constantes, etc...)
 - Extensión `.Rdata`, `.Rda`



Ficheros asociados a R

- 1 Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es .R
 - Hay un sencillo editor con R para ello (ver también Tinn-R)
- 2 Ficheros Workspaces** (área de trabajo). Contiene distintos objetos de R, (conjuntos de datos, variables, constantes, etc...)
 - Extensión .Rdata, .Rda
 - Se cargan desde R con “Cargar área de trabajo”, menú Archivo.



Ficheros asociados a R

- 1 Ficheros scripts:** ficheros sin formato que contienen instrucciones y comandos de R.
 - Permiten repetir análisis, gráficas y facilitan modificaciones
 - Su extensión es .R
 - Hay un sencillo editor con R para ello (ver también Tinn-R)
- 2 Ficheros Workspaces** (área de trabajo). Contiene distintos objetos de R, (conjuntos de datos, variables, constantes, etc...)
 - Extensión .Rdata, .Rda
 - Se cargan desde R con “Cargar área de trabajo”, menú Archivo.
- 3 Ficheros de resultados** Contiene las salidas de R a las distintas instrucciones que hemos introducido.



Directorio de trabajo

- El directorio de trabajo es el directorio donde busca o guarda R por defecto los ficheros asociados.
- Se cambia con “Cambiar dir. de trabajo” desde el menú “Archivo”
- Es buena práctica cambiar directamente el directorio de trabajo al inicio de nuestro trabajo.



Guión

- 1 Qué es R.
- 2 Estructura de R
- 3 El entorno de trabajo en R:
 - Línea de comandos
 - Ficheros asociados a R
 - Directorio de trabajo
- 4 Los principales objetos en R
 - Vectores
 - Los conjuntos de datos (data frames)



Vector

- Un vector es una concatenación de valores.



Vector

- Un vector es una concatenación de valores.
- Para concatenar valores se usa la función “c”:

```
> c(1.34, 3.53, 2.3, 5.98, 8.3)
```

```
[1] 1.34 3.53 2.30 5.98 8.30
```



Vector

- Un vector es una concatenación de valores.
- Para concatenar valores se usa la función “c”:

```
> c(1.34, 3.53, 2.3, 5.98, 8.3)
```

```
[1] 1.34 3.53 2.30 5.98 8.30
```

- Si definimos un vector llamado *peso*:

```
> peso = c(54, 70, 65, 78, 68, 85)
```

Podemos acceder a los elementos de *peso*, usando corchetes:

```
> peso[1]
```

```
[1] 54
```

```
> peso[4]
```

```
[1] 78
```



Los conjuntos de datos

- **EL** objeto más usado en R.



Los conjuntos de datos

- **EL** objeto más usado en R.
- Es una colección de varios vectores de misma longitud.



Los conjuntos de datos

- **EL** objeto más usado en R.
- Es una colección de varios vectores de misma longitud.
- Una fila \leftrightarrow Un individuo
- Cada columna \leftrightarrow una variable



Los conjuntos de datos

- **EL** objeto más usado en R.
- Es una colección de varios vectores de misma longitud.
- Una fila \leftrightarrow Un individuo
- Cada columna \leftrightarrow una variable
- Se construyen juntando vectores usando la instrucción

data.frame:

También hemos medido alturas:

```
> altura = c(160, 170, 172, 185, 160, 175)
```

Ahora juntamos altura y peso en un dataframe:

```
> constitución = data.frame(altura, peso)
```



Los conjuntos de datos (data frames)

De manera que

```
> constitución
```

	altura	peso
1	160	54
2	170	70
3	172	65
4	185	78
5	160	68
6	175	85



¿Cómo acceder a variables o elementos de un dataframe?

Los elementos de un dataframe se pueden acceder de dos maneras distintas:

- 1 Como los elementos de una matriz, usando corchetes:
 - `constitución[i, j]` devuelve el elemento de la fila i and columna j .
 - `constitución[i,]` devuelve toda la fila i (individuo)
 - `constitución[, j]` devuelve toda la columna j (variable)



¿Cómo acceder a variables o elementos de un dataframe?

Los elementos de un dataframe se pueden acceder de dos maneras distintas:

- 1 Como los elementos de una matriz, usando corchetes:
 - `constitución[i, j]` devuelve el elemento de la fila i and columna j .
 - `constitución[i,]` devuelve toda la fila i (individuo)
 - `constitución[, j]` devuelve toda la columna j (variable)
- 2 Columnas enteras se pueden acceder por su nombre, usando `$`:

```
> constitución$peso
```

```
[1] 54 70 65 78 68 85
```



Cálculo de Índice de Masa Corporal

El Índice de Masa Corporal:

$$IMC = \frac{\textit{peso}(kg)}{\textit{altura}^2(m^2)}.$$

Sirve para detectar sobrepeso.



Cálculo de Índice de Masa Corporal

El Índice de Masa Corporal:

$$IMC = \frac{\text{peso}(kg)}{\text{altura}^2(m^2)}$$

Sirve para detectar sobrepeso.

Añadimos una nueva variable (columna) al dataframe constitución:

```
> constitución$IMC = constitución$peso/(constitución$altura/100)^2
> constitución
```

	altura	peso	IMC
1	160	54	21.09375
2	170	70	24.22145
3	172	65	21.97134
4	185	78	22.79036
5	160	68	26.56250
6	175	85	27.75510