

Principios de Estadística

Leonardo Collado Torres y María Gutiérrez Arcelus

Licenciatura en Ciencias Genómicas, UNAM

www.lcg.unam.mx/~lcollado/index.php

www.lcg.unam.mx/~mgutierr/index.php

Cuernavaca, México
Febrero - Junio, 2009

Introducción y R básico

Principios de
Estadística

Orígenes

Uso básico de
R

1 Orígenes

2 Uso básico de R

De donde viene R

Principios de
Estadística

Orígenes

Uso básico de
R

- Para muchos R es un dialecto porque es un derivado del lenguaje S creado por John Chambers y co en los Bell Labs. En si, R fue escrito a mitad de los 90s por Ross Ihaka y Robert Gentleman.
- Desde 1997, R ha sido manejado por el *R Development Core Team* y se ha mantenido como open-source.
- Una ventaja de R es que se puede usar en varias plataformas: UNIX, Windows, Mac.
- R en si es un lenguaje de computación creado para facilitar la manipulación de datos, hacer cálculos y gráficas de alto nivel. Es por esto que R es fuerte en estadística.

Propiedades de R

Principios de
Estadística

Orígenes

Uso básico de
R

R es un ambiente para trabajar en estadística computacional y al mismo tiempo es un lenguaje de programación. Hay usuarios que solo van a usar las funciones básicas de R (como una calculadora) mientras otros incluso harán paquetes que ligen R con C. En fin, R:

- es efectivo en el manejo de datos y su almacenamiento.
- tiene muchos operadores para hacer cálculos en arreglos (vectores) y matrices.
- tiene una gama de herramientas para el análisis de datos. Hay muchos paquetes disponibles, como la familia de Bioconductor.
- tiene un sistema de gráficas muy útil para el análisis de datos. Excel es cosa del pasado ;)

Propiedades de R

Principios de
Estadística

Orígenes

Uso básico de
R

- ya viene con modelos estadísticos.
- hay muchos manuales y un sistema de ayuda bastante bueno. Además hay una comunidad internacional que te puede ayudar :).

Abrir y cerrar R

Principios de
Estadística

Orígenes

Uso básico de
R

- Para abrir R simplemente tienen que escribir el comando `R` en UNIX. Lo primero que verán es una pequeña descripción de R incluyendo la versión que tienen instalada.
- Al abrir R, este busca en el directorio donde están información de alguna sesión previa. Esto luego sera útil con los *workspace*.
- Para cerrar R simplemente escriban `q()`. Les va a pedir si quieren guardar una imagen del workspace – por ahora digan que no.

Workspace

Principios de
Estadística

Orígenes

Uso básico de
R

- Muchas veces tienes que interrumpir tu trabajo. R tiene toda una funcionalidad llamada workspace que te ayuda a retomar tu trabajo de sesiones previas.
- Cuando guardas el workspace se crean dos archivos: `.RData` y `.Rhistory` en el directorio donde estes trabajando. Estos almacenan todos los objetos que haya definido el usuario (vectores, matrices, listas, funciones). La próxima vez que abras R en ese directorio, carga todo lo que creaste antes automáticamente.
- Hay una serie de funciones que les pueden ayudar para organizar su trabajo en R. `getwd` te da tu directorio de trabajo actual, `setwd` lo cambia, `history` te muestra los últimos 25 comandos que usaste y `history(max.show=Inf)` te muestra todos.

Workspace

Principios de
Estadística

Orígenes

Uso básico de
R

- Otras funciones útiles son `savehistory` y `loadhistory`. Además si quieres ver que objetos tienes en tu sesión puedes usar `objects` o `ls`; puedes quitar objetos con `rm`.
- Si quieren guardar el workspace manualmente o con un nombre diferente a `.Rdata` usen `save.image()`. De igual forma, pueden cargar un workspace manualmente con `load`.
- Para checar las opciones del ambiente de R usen `options`. Por ejemplo, pueden cambiar cuantos dígitos se imprimen en el output.

R como Calculadora

Principios de
Estadística

Orígenes

Uso básico de
R

- R es un *expression language*. Aka¹, una R no es igual a una r. Los nombres de variables tienen que empezar por un punto² o caracteres alfanuméricos.

```
> 2 + 2
> 2^2
> r <- c(1:3, 4.5, 109)
> pi * r^2
> sqrt(36)
> sin(2 * pi)
> exp(1)
> log(10)
> log(10, base = 10)
```

¹also known as

²Una letra le tiene que seguir al punto para que sea un nombre válido

Asignación de valores

Principios de
Estadística

Orígenes

Uso básico de
R

- En R, hay 3 formas de asignar valores, aunque en general se usan solo dos: `=` y `<-`
- Preferencialmente usen `<-` simplemente para evitar confusiones. Es que el signo `=` se usa para el paso de valores en las funciones.

```
> A <- c(a = 1, b = 2) ["b"]
```

```
> A = c(a = 1, b = 2) ["b"]
```

```
> A
```

```
b
```

```
2
```

- Aquí queda más clara la asignación en la primera línea, aunque las dos hacen lo mismo.