

UNIVERSIDAD
AUSTRAL



Valores que inspiran

Maestría en Ciencia de Datos



UNIVERSIDAD
AUSTRAL

| INGENIERÍA

Algoritmos y estructura de datos

Programación en R



Algoritmos y estructura de datos

Programación en R

¿Qué es R?

R es un lenguaje y entorno para computación estadística y gráficos.

R puede considerarse como una implementación diferente de S.

R proporciona una amplia variedad de técnicas estadísticas (modelos lineales y no lineales, pruebas estadísticas clásicas, análisis de series temporales, clasificación, agrupamiento, ...) y gráficas, y es altamente extensible.

Algoritmos y estructura de datos

Programación en R

Operadores

Operadores aritméticos

+ : Suma
- : Resta
***** : Multiplicación
/ : División
^ : Potenciación
%/% : División entera
%% : Modulo

Operadores racionales

! : Negación
& : Conjunción elemento
a elemento
&& : Conjunción lógica
| : Disyunción elemento
a elemento
|| : Disyunción lógica

Operadores lógicos

== : iguales
!= : Distintos
> : Mayor
>= : Mayor o igual
< : Menor
<= : Menor



Algoritmos y estructura de datos

Programación en R

Vectores

En R utiliza diferentes estructuras de datos.

La estructura más simple es el vector, que es una colección ordenada de elementos.

El vector, como todos los elementos en R, es un objeto.

En un vector, todos sus elementos son del mismo tipo de dato.



Algoritmos y estructura de datos

Programación en R

Vectores

La función **c()** nos permite construir un vector. En este contexto, la función, puede tener un número arbitrario de vectores como argumento y cuyo resultado es el vector obtenido mediante la concatenación de todos ellos.

La operador de asignación, ('<-').

La asignación puede realizarse también mediante la función **assign()**, o apuntando la flecha hacia el otro lado ('->').



Algoritmos y estructura de datos

Programación en R

Vectores

Vectores numéricos : son aquellos vectores que contienen valores de tipo numéricos (Numeric, Integer, Complex).

Ej. `c(1, 2, 3, 4)`

Vectores lógicos : son aquellos vectores que contienen valores de tipo lógicos (TRUE, FALSE).

Ej. `c(TRUE, FALSE, FALSE, TRUE)`

Vectores de carácter : son aquellos vectores que contiene valores de tipo carácter (character, string).

Ej. `c('a', 'b', 'cd')`



Algoritmos y estructura de datos

Programación en R

Aritmética vectorial

Los vectores pueden usarse en expresiones aritméticas, en cuyo caso las operaciones se realizan elemento a elemento.

Si opero con más de un vector, y no lo son, el resultado será un vector de la longitud del más largo, y el más corto será *reciclado*, repitiéndolo tantas veces como sea necesario.



Algoritmos y estructura de datos

Programación en R

Sucesiones

: este operador permite construir una sucesiones utilizando el incremento natural (Ej. $1:4 = [1, 2, 3, 4]$). Este operador tiene prioridad sobre cualquier otro.

seq() : la función permite generar sucesiones complejas (Ej. $\text{seq}(\text{from}=1, \text{to}=2, \text{by}=.2) = [1, 1.2, 1.4, 1.6, 1.8, 2]$).

rep() : la función permite duplicar un objeto de formas diversas (Ej. $\text{rep}(4, \text{times}=3) = [4, 4, 4]$).



Algoritmos y estructura de datos

Programación en R

Valores faltantes

En ocasiones puede que no todas las componentes de un vector sean conocidas.

Cuando falta un elemento, lo que se denomina 'valor faltante', se le asigna un valor especial, NA.



Algoritmos y estructura de datos

Programación en R

Vectores indices

Puede seleccionar un subvector de un vector añadiendo al nombre del mismo un vector de índices entre corchetes, `[y]`.

En general podrá obtener un subvector de cualquier expresión cuyo resultado sea un vector, sin mas que añadirle un vector de índices entre corchetes.



Algoritmos y estructura de datos

Programación en R

Vectores índices

Un vector lógico : en este caso el vector de índices debe tener la misma longitud que el vector al que refiere. Sólo se seleccionarán los elementos correspondientes a valores T del vector de índices y se omitirá el resto.

Un vector de números naturales positivos : en este caso los elementos del vector de índices deben pertenecer al conjunto $[1, 2, \dots, \text{length}(x)]$. El resultado es un vector formado por los elementos del vector referido que corresponden a estos índices y en el orden en que aparecen en el vector de índices.



Algoritmos y estructura de datos

Programación en R

Vectores indices

Un vector de números naturales negativos : en este caso, los índices indican los elementos del vector referido que deben excluirse.

Un vector de caracteres : esta opción solo puede realizarse si el vector posee el atributo *names* (nombres) para identificar sus componentes, en cuyo caso se comportará de modo similar a un vector de números naturales positivos.



Algoritmos y estructura de datos

Programación en R

Clases de objetos

Vector : este es el tipo de objeto más básico de R.

Matrices : las matrices o, más generalmente, variables indexadas (Arrays) son generalizaciones multidimensionales de los vectores. De hecho, son vectores indexados por dos o más índices y que se imprimen de modo especial.

Factores : Los factores sirven para representar datos categóricos.

Listas : las listas son una forma generalizada de vector en las cuales los elementos no tienen por qué ser del mismo tipo y a menudo son a su vez vectores o listas.



Algoritmos y estructura de datos

Programación en R

Clases de objetos

Data Frames : las hojas de datos (data frames) son estructuras similares a una matriz, en que cada columna puede ser de un tipo distinto a las otras. Las hojas de datos son apropiadas para describir 'matrices de datos' donde cada fila representa a un individuo y cada columna una variable, cuyas variables pueden ser numéricas o categóricas.

Funciones : las funciones son también objetos de R que pueden almacenarse en el espacio de trabajo, lo que permite extender las capacidades de R fácilmente.



Algoritmos y estructura de datos

Programación en R

Atributos de un objeto

Las entidades que manipula R se conocen con el nombre de *objetos*.

Los objetos que poseen todos sus elementos del mismo tipo o *modo* se denominan estructuras 'atómicas'.

Los tipos o *modo* 'atómicos' en R son:

- **character** (caracteres/cadena de caracteres)
- **numeric** (Números reales)
- **complex** (Números complejos)
- **logical** (Booleanos True/False)



Algoritmos y estructura de datos

Programación en R

Atributos de un objeto

R también maneja objetos modos, no atómicos y de estructura recursiva.

Los tipos o *modo* 'no atómicos' en R son:

- **list** (listas)
- **function** (Funciones)
- **expression** (Expresiones)

Otro atributo de un objeto es su longitud (*length*)



Algoritmos y estructura de datos

Programación en R

Atributos de un objeto

La función *attributes(objeto)* proporciona una lista de todos los atributos no intrínsecos definidos para el objeto en ese momento

Cada objeto pertenece a una clase (*class*), y ello permite utilizar en R programación dirigida a objetos.



Algoritmos y estructura de datos

Programación en R

Factores nominales y ordinales

Un *factor* es un vector utilizado para especificar una clasificación discreta de los elementos de otro vector de igual longitud.

En R existen factores **nominales** y factores **ordinales**.



Algoritmos y estructura de datos

Programación en R

Factores nominales y ordinales

Un *factor* es un vector utilizado para especificar una clasificación discreta de los elementos de otro vector de igual longitud.

En R existen factores **nominales** y factores **ordinales**.



Algoritmos y estructura de datos

Programación en R

Matrices

Una variable indexada (*array*) es una colección de datos, por ejemplo numéricos, indexada por varios índices.

Un vector de dimensiones es un vector de números enteros positivos.

Si su longitud es k entonces la variable indexada correspondiente es k -dimensional.



Algoritmos y estructura de datos

Programación en R

Listas

Una lista es un objeto consistente en una colección ordenada de objetos, conocidos como componentes.

A diferencia de los vectores, las listas pueden tener elementos de distinto tipo.

Una lista puede tener listas como elementos.



Algoritmos y estructura de datos

Programación en R

Data frames

Una hoja de datos (Data frame) es una lista que pertenece a la clase "data.frame".

Los componentes deben ser vectores (numéricos, cadenas de caracteres, o lógicos), factores, matrices numéricas, listas u otras hojas de datos.

Las matrices, listas, y hojas de datos contribuyen a la nueva hoja de datos con tantas variables como columnas, elementos o variables posean, respectivamente.



Algoritmos y estructura de datos

Programación en R

Data frames

Los vectores numéricos y los factores se incluyen sin modificar, los vectores no numéricos se fuerzan a factores cuyos niveles son los únicos valores que aparecen en el vector.

Los vectores que constituyen la hoja de datos deben tener todos la misma longitud, y las matrices deben tener el mismo tamaño de filas.



Algoritmos y estructura de datos

Programación en R

Funciones

R permite crear objetos del modo *function*, que constituyen nuevas funciones de R, que se pueden utilizar a su vez en expresiones posteriores.

```
NombreDeFuncion <-function(arg 1, arg 2, ...) { expresión }
```



Algoritmos y estructura de datos

Programación en R

Instrucciones de control

If (*expresión condicional*)
 expresión

else
 expresión

ifelse (*expresión condicional*,
 expresión verdadera,
 expresión falsa)



Algoritmos y estructura de datos

Programación en R

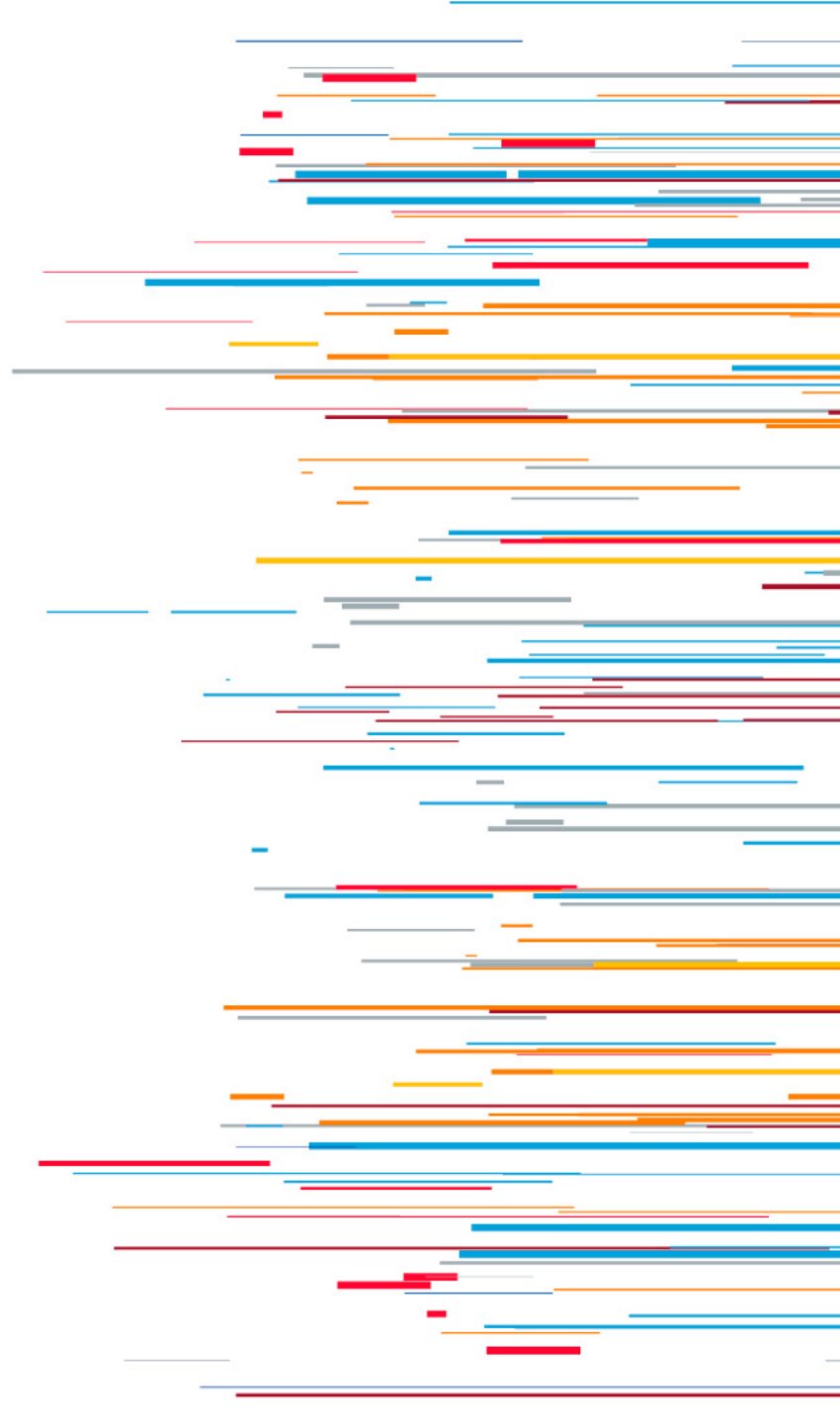
Instrucciones de repetición

for (*i in expresión*) *expresión*

while (*expresión condicional*) *expresión*

repeat *expresión*

break *corta la repetición*



Muchas gracias.

www.austral.edu.ar



UNIVERSIDAD
AUSTRAL

| INGENIERÍA