

Material Imprimible

Curso de Análisis de datos con R

Módulo I

Contenidos:

- R Base y R Studio
- Objetos y Atributos
- Vectores y listas
- Dataframes

R Base y R Studio

R es un entorno y lenguaje de programación estadística que proporciona una amplia gama de herramientas para análisis de datos y visualización. El lenguaje R permite a los usuarios realizar análisis estadísticos avanzados, modelado predictivo y generación de gráficos de alta calidad. Típicamente, el servidor de R es conocido como R Base y es donde se ejecutan los comandos programados.

R Studio es un entorno de desarrollo integrado (IDE) diseñado específicamente para facilitar la programación en R. Ofrece una interfaz de usuario amigable que incluye herramientas para escribir y depurar código, así como paneles para gestionar gráficos, historiales de comandos y entornos de trabajo. R Studio mejora significativamente la experiencia de programar en R al proporcionar una organización eficiente de proyectos, herramientas de colaboración y una integración perfecta con la consola de R Base.

La combinación de R y R Studio ofrece a los analistas de datos y científicos una poderosa plataforma para llevar a cabo investigaciones estadísticas y análisis de datos. La comunidad de usuarios activa y comprometida contribuye al crecimiento continuo del universo R mediante la creación y compartición de paquetes, lo que enriquece la funcionalidad y las capacidades del lenguaje. Estas tecnologías son ampliamente utilizadas en las organizaciones para abordar desafíos complejos relacionados con el análisis de datos, la minería de datos y la generación de informes. La capacidad de extender funcionalidades mediante la instalación de paquetes específicos hace que R sea una herramienta versátil y adaptable a diversas necesidades analíticas, reduciendo costos y trabajando en entornos escalables de código abierto.

Objetos y Atributos

En R, los objetos son estructuras que almacenan datos y pueden ser manipulados mediante funciones y operadores. Algunos tipos de objetos comunes incluyen vectores, matrices, listas, data frames y factores. Cada tipo de objeto tiene atributos específicos que proporcionan información adicional sobre los datos almacenados.

Los atributos en R son metadatos asociados a un objeto que describen sus características. Por ejemplo, si tenemos un DataFrame, sus dimensiones (filas y columnas) es uno de sus atributos.

Vectores

Un vector es una colección de elementos del mismo tipo. Se puede crear un vector usando la función `c()`, donde cada elemento está separado por comas. Es regla que el vector tenga SIEMPRE el mismo tipo de datos

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

Listas

Una lista es una colección ordenada de elementos que pueden ser de diferentes tipos.

Se debe utilizar la función `list()` para crearlas.

```
lista_ejemplo <- list(nombre = "Curso Capacitarte", alumnos= 50, calificaciones = c(90, 85, 95))
```

DataFrames

Un data frame es una estructura de datos bidimensional similar a una tabla de una base de datos. Se compone de filas y columnas, combinando características de listas y matrices, donde cada columna puede tener un tipo de datos diferente.

Se debe usar la función `data.frame()`

```
dataframe_ejemplo <- data.frame(  
  Nombre = c("Ana", "Nestor", "Carlos"),  
  Edad = c(28, 32, 25),  
  Salario= c(125, 548, 320)  
)
```

También, se puede usar la función nativa `read.csv` para cargar desde archivos plano (csv o txt) que tengamos en nuestra computadora. A lo largo del curso, usaremos esta funcionalidad.

```
mi_data_frame <- read.csv("curso_capacitarte.csv")
```

Es importante definir entre las comillas (" ") todo el directorio del archivo con su ubicación exacta.