

Manejo de dataframes

Un DataFrame en R es una estructura de datos bidimensional que se utiliza para almacenar conjuntos de datos tabulares, donde las filas representan observaciones y las columnas representan variables. Es una de las estructuras de datos más utilizadas en R y proporciona una forma eficiente de trabajar con datos heterogéneos.

Creación de DataFrames:

Puedes crear un DataFrame utilizando la función `data.frame()` o convirtiendo otras estructuras de datos, como matrices o listas, utilizando la función `as.data.frame()`.

```
# Crear un DataFrame desde cero
df <- data.frame(
  Nombre = c("Juan", "María", "Pedro"),
  Edad = c(25, 30, 22),
  Puntuacion = c(85, 92, 78)
)

# Convertir una matriz en un DataFrame
matriz <- matrix(c(1, 2, 3, 4, 5, 6), ncol = 2)
df_desde_matriz <- as.data.frame(matriz)
```

La función `data.frame()` en R se utiliza para crear DataFrames, que son estructuras de datos bidimensionales. Aquí están los parámetros principales de la función `data.frame()` junto con ejemplos para ilustrar su funcionamiento:

Parámetros Principales de `data.frame()`:

1. ... (Puntos Suspensivos):

- **Descripción:** Permite especificar columnas del DataFrame mediante argumentos adicionales. Cada argumento adicional representa una columna del DataFrame.
- **Ejemplo:**

```
# Crear un DataFrame con columnas específicas
df <- data.frame(
  Nombre = c("Juan", "María", "Pedro"),
  Edad = c(25, 30, 22),
```

```
Puntuacion = c(85, 92, 78)
)
```

2. row.names:

- **Descripción:** Permite especificar nombres para las filas del DataFrame. Puede ser un vector de nombres o `NULL` para no asignar nombres a las filas.
- **Ejemplo:**

```
# Crear un DataFrame con nombres de fila
df_con_nombres <- data.frame(
  Nombre = c("Juan", "María", "Pedro"),
  Edad = c(25, 30, 22),
  Puntuacion = c(85, 92, 78),
  row.names = c("Obs1", "Obs2", "Obs3")
)
```

3. check.rows:

- **Descripción:** Controla si las filas deben tener la misma longitud. Si `check.rows = TRUE` (por defecto), se verifica que todas las filas tengan la misma longitud.
- **Ejemplo:**

```
# Crear un DataFrame con verificación de longitud de fila
df_verificado <- data.frame(
  Nombre = c("Juan", "María", "Pedro"),
  Edad = c(25, 30, 22),
  Puntuacion = c(85, 92) # Error: Longitud desigual
)
```

En resumen, `data.frame()` permite crear DataFrames en R y ofrece flexibilidad en la especificación de columnas y nombres de fila. El uso de los parámetros adicionales permite personalizar la creación del DataFrame según las necesidades específicas de los datos.

Acceso y Manipulación de DataFrames:

Puedes acceder a las columnas de un DataFrame utilizando el operador `$` o mediante el uso de corchetes `[]`. Además, puedes realizar operaciones de subconjunto y manipulación de datos utilizando funciones específicas.

```
# Acceder a columnas por nombre
edades <- df$Edad

# Filtrar filas según una condición
jovenes <- df[df$Edad < 30, ]

# Agregar una nueva columna
df$Categoria <- ifelse(df$Puntuacion > 80, "A", "B")
```

Funciones Útiles para DataFrames:

- `summary()` : Proporciona un resumen estadístico de las columnas numéricas.
- `head()` y `tail()` : Muestran las primeras o últimas filas del DataFrame.
- `str()` : Muestra la estructura de un DataFrame.
- `subset()` : Permite seleccionar un subconjunto de filas y columnas basado en condiciones.

```
# Resumen estadístico
summary(df)

# Mostrar las primeras 3 filas
head(df, 3)

# Estructura del DataFrame
str(df)

# Seleccionar subconjunto
subset_df <- subset(df, Edad > 25)
```

Ventajas de DataFrames sobre Matrices:

1. **Heterogeneidad:** Los DataFrames pueden contener columnas de diferentes tipos de datos, mientras que las matrices deben contener elementos del mismo tipo.
2. **Nombres de Columna:** Los DataFrames permiten asignar nombres a las columnas, lo que facilita el acceso y la manipulación de datos.
3. **Facilidad de Manipulación:** Los DataFrames ofrecen funciones específicas para manipular y analizar datos, como `subset()`, `summary()`, y otras, que facilitan la exploración y transformación de datos.

4. **Compatibilidad con Data Frames en R:** Muchas funciones en R devuelven resultados como DataFrames, lo que facilita la integración y el análisis de datos de diferentes fuentes.

En resumen, los DataFrames son una estructura de datos versátil y poderosa en R que facilita el manejo y análisis de datos tabulares. Son especialmente útiles en análisis de datos, estadísticas y tareas relacionadas con la ciencia de datos.