

Factores

Téllez Gerardo Rubén

2/4/2021

Factores

Similar a un vector, con estructura más rica que permite usarlo para clasificar las observaciones.

- **levels:** atributo del factor. Cada elemento del factor es igual a un nivel. Los niveles clasifican las entradas del factor. Se ordenan por orden alfabético
- Para definir un factor, primero se define un vector y se transforma por las funciones **factor()** o **as.factor()**.

Normalmente los niveles son palabras.

```
nombres <- c("Juan", "Antonio", "Ricardo", "María", "Rosa", "Yuko", "Ernesto")
nombres.factor = factor(nombres)

# as.factor convierte le vector en factores
# toma como niveles los que aparecen dentro del factor

# factor() define un factor a partir de lvector
# Un par de parámetros permite modificar algunas cosas
```

Todo lo que se quiera convertir de texto a factor para clasificar los datos, se usan los factores en lugar del string.

Cuando se hace un vector con la intención de clasificar datos, es más conveniente convertirlo a factor.

La Función factor()

- **factor(vector, levels=...)**: define un factor a partir del vector y dispone de algunos parámetros que permiten modificar el factor que se crea.
 - **levels:** permite especificar los niveles e incluso añadir niveles que no aparecen en el vector.
 - **labels:** permite cambiar los nombres de los. niveles
- **levels(factor)**: regresa los niveles del factor.

Cuando usemos factor(), podemos especificar los niveles de factor

```
sexo = c("M", "H", "H", "M", "H", "H", "M", "H", "H", "M", "H", "H", "M")
sexo.factor = factor(sexo)
sexo.factor
```

```
## [1] M H H M H H M H H M H H M
## Levels: H M
```

```
sexo.factor_2 = as.factor(sexo)
sexo.factor_2
```

```
## [1] M H H M H H M H H M H H M
## Levels: H M
```

```
#La diferencia se nota al requerir más niveles.
sexo.factor_3 = factor(sexo, levels = c("M", "H", "Q"))
#Se especificó un nivel adicional
sexo.factor_3
```

```
## [1] M H H M H H M H H M H H M
## Levels: M H Q
```

```
#Para aportar etiquetas
sexo.factor_4 = factor(sexo, levels = c("M", "H", "Q"), labels = c("Mujer", "Hombre", "Quimera"))
sexo.factor_4
```

```
## [1] Mujer Hombre Hombre Mujer Hombre Hombre Mujer Hombre Hombre Mujer
## [11] Hombre Hombre Mujer
## Levels: Mujer Hombre Quimera
```

Es muy importante dar etiquetas significativas.

Función levels()

La función levels() permite cambiar el nombre directamente a los nombres del factor

```
levels(sexo.factor_4)
```

```
## [1] "Mujer" "Hombre" "Quimera"
```

```
levels(sexo.factor_4) = c("Femenino", "Masculino", "Hibrido")
levels(sexo.factor_4)
```

```
## [1] "Femenino" "Masculino" "Hibrido"
```

Bastante útil cuando en particular tenemos por ejemplo notas

```

notas <- c(0, 6, 10, 10, 6, 0, 10, 6, 6, 10, 0, 0)
notas.factor <- factor(notas)
levels(notas.factor)

```

```
## [1] "0" "6" "10"
```

```

levels(notas.factor) <- c("Reprobado", "Suficiente", "Sobresaliente")
levels(notas.factor)

```

```
## [1] "Reprobado" "Suficiente" "Sobresaliente"
```

```
notas.factor
```

```

## [1] Reprobado Suficiente Sobresaliente Sobresaliente Suficiente
## [6] Reprobado Sobresaliente Suficiente Suficiente Sobresaliente
## [11] Reprobado Reprobado
## Levels: Reprobado Suficiente Sobresaliente

```

```

#Para agrupar niveles distintos al mismo nivel
levels(notas.factor) <- c("Reprobado", "Aprobado", "Aprobado")
notas.factor

```

```

## [1] Repobado Aprobado Aprobado Aprobado Aprobado Repobado Aprobado Aprobado
## [9] Aprobado Aprobado Repobado Repobado
## Levels: Repobado Aprobado

```

Factor ordenado

Es un factor donde los niveles siguen un orden

- **ordered(vector, levels=...)**: función que define un factor ordenado y tiene los mismos parámetros que factor.

```
notas
```

```
## [1] 0 6 10 10 6 0 10 6 6 10 0 0
```

```
ordered(notas, labels = c("Reprobado", "Suficiente", "Sobresaliente"))
```

```

## [1] Reprobado Suficiente Sobresaliente Sobresaliente Suficiente
## [6] Reprobado Sobresaliente Suficiente Suficiente Sobresaliente
## [11] Reprobado Reprobado
## Levels: Reprobado < Suficiente < Sobresaliente

```

Los niveles ahora estará ordenados de manera que **Reprobado < Suficiente < Sobresaliente**