

Manipulation des facteurs

Qu'est-ce que c'est ?

Les facteurs sont des vecteurs un peu particuliers, facilitant la manipulation de données qualitatives (qu'elles soient numériques ou caractères). En effet, en plus de stocker les différents éléments comme un vecteur classique, il stocke également l'ensemble des différentes modalités possibles dans un attribut accessible via la commande `levels`.

Ils forment une classe d'objets et bénéficient de traitements particuliers lors de leur manipulation et lors de l'utilisation de certaines fonctions. Les facteurs peuvent être non ordonnés (homme, femme) ou ordonnés (niveaux de ski).

Dans quel cas l'utiliser dans R ?

Pour encoder les réponses à une question ferme (c'est-à-dire qu'une question ne laissant à son destinataire que des choix prédéfinis), on utilisera les facteurs. En effet, en statistique, un facteur est typiquement utilisé pour stocker les valeurs observées d'une variable qualitative ou catégorique.

Pour illustrer le fonctionnement des facteurs, nous allons créer un facteur avec des attributs par défaut, puis des niveaux personnalisés, puis des niveaux et des étiquettes personnalisés.

Création des facteurs

3 fonctions pour créer les facteurs

- 1. La fonction `factor`

`factor` permet de créer un facteur en définissant directement les différents éléments du facteur.

```
sexe <- factor(c("H", "H", "F", "H", "H", "F", "F", "F"))
sexe
```

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

```
[1] H H F H H F F F Levels: F H
```

- 2. La fonction `as.factor`

```
salto <- c(1:5,5:1)
salto
```

```
## [1] 1 2 3 4 5 5 4 3 2 1
```

```
[1] 1 2 3 4 5 5 4 3 2 1
```

```
salto.f <- as.factor(salto)
salto.f
```

```
## [1] 1 2 3 4 5 5 4 3 2 1
## Levels: 1 2 3 4 5
```

```
[1] 1 2 3 4 5 5 4 3 2 1 Levels: 1 2 3 4 5
```

- 3. La fonction `ordered`

La fonction `ordered` va quant à elle nous permettre de créer des facteurs ordonnés

```
niveau <- ordered(c("débutant","débutant","champion",
                    "champion","moyen","moyen","moyen",
                    "champion"),
                 levels=c("débutant","moyen","champion"))
niveau
```

```
## [1] débutant débutant champion champion moyen      moyen      moyen      champion
## Levels: débutant < moyen < champion
```

Niveaux d'un facteur

Les facteurs prennent leurs valeurs dans un ensemble de modalités prédéfinies (niveaux), et ne peuvent en prendre d'autres. Pour connaître les niveaux d'un facteur, on utilise la fonction 'levels'.

```
levels(sexe)
```

```
## [1] "F" "H"
```

Remarque : R affiche les niveaux d'un facteur sous forme de caractère. Cependant, en interne, R les stocke sous forme d'entiers (dans notre exemple 2="H" et 1="F"). Par défaut, les niveaux d'un facteur nouvellement créés sont l'ensemble des valeurs uniques du vecteur.

L'option `levels` permet de prédéfinir les niveaux d'un facteur. Dans l'exemple suivant, `de` représente le résultat du lancement d'un dé huit fois.

```
de <- factor(c(3, 2, 2, 1, 3, 1, 3, 1), levels = c(1, 2, 3, 4, 5, 6))
de
```

```
## [1] 3 2 2 1 3 1 3 1
## Levels: 1 2 3 4 5 6
```

```
str(de)
```

```
## Factor w/ 6 levels "1","2","3","4",...: 3 2 2 1 3 1 3 1
```

Renommer les modalités d'un facteur

On peut renommer le facteur en utilisant la fonction `levels`

```
levels(sexe) <- c("Femme", "Homme")
sexe
```

```
## [1] Homme Homme Femme Homme Homme Femme Femme Femme
## Levels: Femme Homme
```

Modifier les valeurs d'un facteur

On peut modifier la valeur d'un facteur facilement par indexation comme pour un vecteur avec la fonction `levels`

```
levels(sexe)
```

```
## [1] "Femme" "Homme"
```

```
levels(sexe)[1] <- "Femme"
sexe
```

```
## [1] Homme Homme Femme Homme Homme Femme Femme Femme
## Levels: Femme Homme
```

Autre exemple:

```
de[c(2, 3)] <- c(6, 5)
de
```

```
## [1] 3 6 5 1 3 1 3 1
## Levels: 1 2 3 4 5 6
```

L'ordre des niveaux

Par défaut, les niveaux d'un facteur nouvellement créés sont classés par ordre alphanumérique croissant ou selon l'ordre qui figure dans l'option `levels`. Cet ordre est utilisé chaque fois que le facteur est employé.

```
sexe <- factor(c("H", "H", "F", "H", "H", "F", "F", "F"))
sexe
```

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

```
summary(sexe)
```

```
## F H  
## 4 4
```

On peut modifier l'ordre des niveaux d'un facteur existant en utilisant l'option `levels` :

```
sexe <- factor(sexe, levels = c("H", "F"))  
sexe
```

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

```
summary(sexe)
```

```
## H F  
## 4 4
```

Ce qu'il faut retenir :

- 3 moyens de créer des facteurs : `factor` , `as.factor` et `ordered`
- On peut modifier, renommer les modalités du facteur avec la fonction `levels`

En savoir plus : voici les sujets que vous pouvez approfondir par rapport à ce tutorat. - La documentation sur les facteurs : [LINK](https://sodocumentation.net/fr/r/topic/1104/facteurs)

(<https://sodocumentation.net/fr/r/topic/1104/facteurs>) - Les vecteurs : [LINK](https://sodocumentation.net/fr/r/topic/1088/creation-de-vecteurs) (<https://sodocumentation.net/fr/r/topic/1088/creation-de-vecteurs>)

Sources :

Source Openclassrooms (<https://openclassrooms.com/fr/courses/4525256-initiez-vous-au-langage-r-pour-analyser-vos-donnees/6250873-utilisez-les-facteurs>)

Source Bookdown (<https://bookdown.org/ael/rexplor/chap3-2.html>)