



---

# Structures de données en R

*(Premiers pas avec R)*

Christophe Pallier et Christophe Lalanne

`christophe.pallier@m4x.org / christophe.lalanne@gmx.net`

Mastère de Sciences Cognitives, EHESS – Paris 5 – ENS



## Facteurs

---

Les facteurs sont des variables qui permettent de définir des sous-groupes.

```
data=c(rnorm(50,mean=10),rnorm(50,mean=11))
```

```
group=rep(1:2,c(50,50))
```

```
plot(data~group)
```

```
groupf=factor(group,labels=c('g1','g2'))
```

```
levels(groupf)
```

```
data[groupf=='g1']
```

```
data[groupf=='g2']
```

```
plot(data~groupf)
```

```
by(data,groupf,summary)
```

```
tapply(data,groupf,mean)
```



## *Facteurs (suite)*

---

```
a=gl(4,10,labels=paste('a',1:4,sep=' '))
b=gl(2,1,40,labels=paste('b',1:2,sep=' '))
y=1:40+rnorm(40,sd=10)
a:b
plot(y~a)
plot(y~b)
plot(y~factor(a:b))
```



## Tableaux (*data.frames*)

---

Un `data.frame` est un ensemble de vecteurs qui ont tous la même longueur. C'est la structure idéale pour stocker un tableau de données à deux dimensions.

```
a=gl(10,10)
```

```
b=gl(2,50)
```

```
y=rnorm(100,mean=as.numeric(b))
```

```
dat=data.frame(a,b,y)
```

```
dat
```

```
names(dat)
```

```
dat$a
```

```
dat$y
```

Lecture/écriture de `data.frames`

Un tableau enregistré dans un fichier texte (par exemple un fichier excel sauvé au format 'csv') peut être lu avec la fonction '`{read.table()}`'

```
a<-read.table('http://www.pallier.org/ressources/tp.stats/donnees/sommeil1')
```

```
names(a)
```

```
a
```



## Exercices

---

- ▶ Générer un facteur déclarant 4 sous-groupes de 100 éléments chacun.
- ▶ Définir un vecteur de données aléatoires distribuées normalement dont les moyennes des sous-groupes précédents soient, respectivement, 100, 110, 90 et 100, et l'écart-type soit 10.
- ▶ Lire le fichier de données  
'<http://www.pallier.org/ressources/tp.stats/donnees/pedago.txt>' et calculer les effectifs de groupes définis par la colonne "pédago" et les moyennes des notes pour chacun de ces sous-groupes. (Faire un graphique en bâton (barplot) de ces moyennes)



## Structures de contrôle

---

- ▶ Execution conditionnelle:

```
if (a==1) { 2} else {3}
```

- ▶ Boucle

```
for (i in 1:10) { print(i) }
```

- ▶ Fonctions

```
plus1 <- function (x) { x+1 }  
plus1(5)
```

Exercices:

- ▶ Définir une fonction qui renvoie la différence entre le maximum et le minimum d'un vecteur
- ▶ Définir une fonction qui calcule la factorielle d'un entier (en utilisant une boucle).
- ▶ Ecrire une boucle qui va mettre dans un vecteur les données numériques stockées dans des fichiers 'a1.dat', 'a2.dat', ... 'a100.dat'



*fonctions*

---