

# 4M007 PROJET

Comportement Bancaire de 500 clients

Makhatch Abdulvagabov

UPMC

17/12/2017

Titre de partie

Traitement des données manquantes

Compléments

Titre de partie

## Un slide avec en-tête

Ceci est un fichier RMarkdown pour une présentation au format Beamer (création de slides en pdf). Il se compile sous RStudio avec la commande Knit to PDF (Beamer).

L'objectif de cette présentation est triple :

- donner un modèle de présentation Beamer créée avec RMarkdown;
- donner des outils pour traiter les données manquantes;
- utiliser des blocs de code à options (echo=FALSE,include=FALSE).

Un slide sans en-tête

# Traitement des données manquantes

# Création d'un faux jeu de données

*Là un bout de code est évalué mais vous ne voyez ni le code ni la sortie.*

Voici le faux jeu de données créé et son résumé numérique (*là vous voyez la sortie, mais pas le code*).

*(là on n'a plus beaucoup de place donc on va écrire tout petit)*

```
##   Salaire Age Ville
## A    2500  40 Lille
## B    1570  25 <NA>
## C    3027  30 Lyon
## D         NA  50 <NA>
```

```
##      Salaire      Age      Ville
## Min.   :1570   Min.   :25.00   Lille:1
## 1st Qu.:2035   1st Qu.:28.75   Lyon :1
## Median :2500   Median :35.00   NA's :2
## Mean   :2366   Mean   :36.25
## 3rd Qu.:2764   3rd Qu.:42.50
## Max.   :3027   Max.   :50.00
## NA's    :1
```

Le tableau contient 4 individus et 3 variables. L'âge moyen des individus est de `mean(fakedata$Age)` ans. *(là je viens d'intégrer du code R au texte)*.

Quand on a beaucoup de données, `summary` permet de voir facilement s'il y a des données manquantes (codées en général `NA`).

## Localisation des données manquantes

Lorsque le tableau de données n'est pas trop grand, on peut visualiser où sont les données manquantes.

```
##   Salaire   Age Ville
## A   FALSE FALSE FALSE
## B   FALSE FALSE  TRUE
## C   FALSE FALSE FALSE
## D    TRUE FALSE  TRUE
```

Sinon, on peut au moins faire le bilan du nombre de valeurs manquantes par individu

```
NAind<-apply(localiseNA,1,sum);NAind
```

```
## A B C D
## 0 1 0 2
```



ou par variables

```
NAvar<-apply(localiseNA,2,sum);NAvar
```

```
## Salaire      Age      Ville  
##          1         0         2
```

# Suppression des individus avec des données manquantes

Ici les individus comportant des données manquantes se trouvent dans les lignes

```
## B D  
## 2 4
```

la liste des individus comportant des données manquantes est

```
## [1] "B" "D"
```

et la liste des variables comportant des données manquantes

```
## [1] "Salaire" "Ville"
```

Lorsque peu d'individus ont des données manquantes (disons moins de 5%), on peut les supprimer du fichiers.

```
newfakedata<-na.omit(fakedata);newfakedata
```

```
##      Salaire Age Ville  
## A      2500  40 Lille  
## C      3027  30  Lyon
```

# Compléments

## Insérer une figure avec un titre

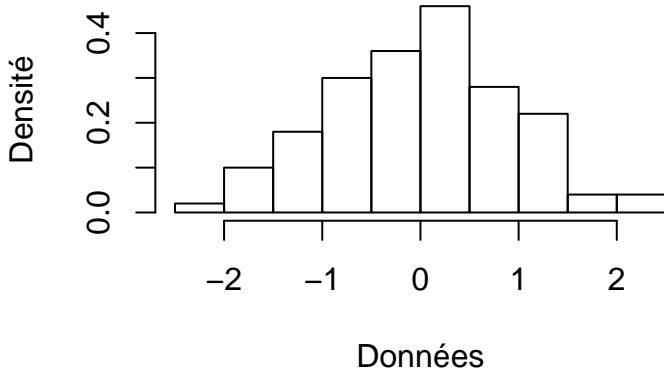


Figure 1: Histogramme des données

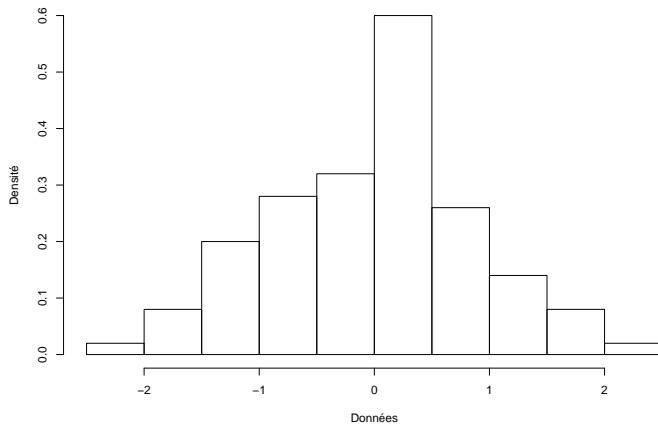


Figure 2: Histogramme des données