



# Statistique exploratoire

Thierry Dhorne

3 mars 2016



# Méthode d'analyse

## ❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

- on travaille d'abord variable par variable
  - statistique univariée
- puis on s'intéresse aux variables deux à deux
  - pour étudier les liens
  - statistique bivariée
- enfin on analyse le tableau globalement
  - pour faire une synthèse
  - statistique multivariée
- ★ pas étudiée dans ce cours



❖ Méthode  
d'analyse

Exploration  
univariée d'une  
variable qualitative

- ❖ Présentation des données
- ❖ Étude exploratoire
- ❖ Analyse critique
- ❖ Restitution graphique  
Diagramme en barres ou en colonnes
- ❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration  
univariée d'une  
variable quantitative  
entière

# *Exploration univariée d'une variable qualitative*



# Présentation des données

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

❖ Présentation des données

❖ Étude exploratoire

❖ Analyse critique

❖ Restitution graphique  
Diagramme en barres ou en colonnes

❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration univariée d'une variable quantitative entière

- population ou un échantillon d'effectif  $n$  sur laquelle est définie une variable qualitative  $Y$  à  $I$  modalités  $y_i$

➤ Exemple :

Population : les étudiants d'une classe d'effectif 45.

Sexe :

*F H F H F H F F F F H F H H H H H H H H H F*

*H H F H H H F F H H H H H F H H H H F H H H*

$Y = \text{sexe}$  ; variable qualitative de modalité  $y_i$  : masculin / féminin où  $i = 1, 2$  (car 2 modalités)



# Étude exploratoire

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

❖ Présentation des données

❖ Étude exploratoire

❖ Analyse critique

❖ Restitution graphique  
Diagramme en barres ou en colonnes

❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration univariée d'une variable quantitative entière

- information synthétique des données
- tableau des effectifs ou des fréquences (absolues ou relatives)

$y_i$	$n_i$	$f_i$
$y_1 = \text{Homme}$	31	68.9 %
$y_2 = \text{Femme}$	14	31.1 %
total	45	100 %



## Analyse critique

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

❖ Présentation des données

❖ Étude exploratoire

❖ Analyse critique

❖ Restitution graphique  
Diagramme en barres ou en colonnes

❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration univariée d'une variable quantitative entière

➤ comparer par rapport à la fréquence dans la population globale des jeunes Français pour les mêmes classes d'âge

$y_i$	$n_i$	$f_i$
$y_1 = \text{Homme}$	2404787	50.5 %
$y_2 = \text{Femme}$	2355020	49.5 %
total	4759807	100 %



# Restitution graphique

## Diagramme en barres ou en colonnes

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

❖ Présentation des données

❖ Étude exploratoire

❖ Analyse critique

❖ Restitution graphique  
Diagramme en barres ou en colonnes

❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration univariée d'une variable quantitative entière

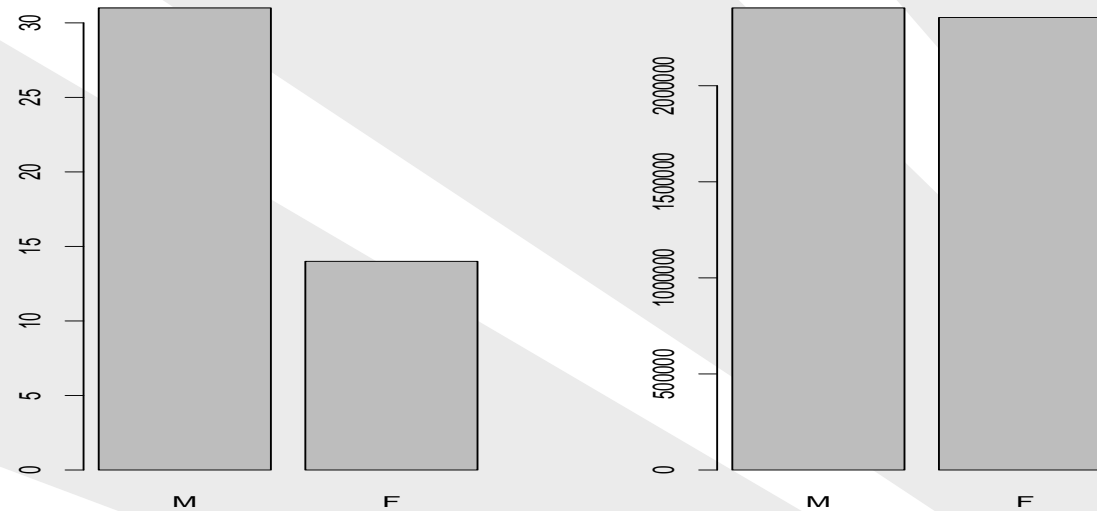


Diagramme en barres de la variable sexe :

- à gauche population étudiante
- à droite population de référence



# Restitution graphique

## Diagramme en secteurs ou camembert

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

❖ Présentation des données

❖ Étude exploratoire

❖ Analyse critique

❖ Restitution graphique  
Diagramme en barres ou en colonnes

❖ Restitution graphique  
Diagramme en secteurs ou camembert

Exploration univariée d'une variable quantitative entière

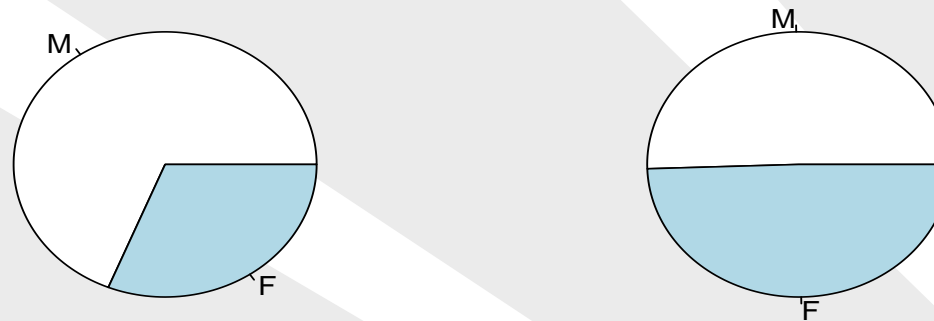


Diagramme en secteurs de la variable sexe :

- à gauche population étudiante
- à droite population de référence





❖ Méthode  
d'analyse

Exploration  
univariée d'une  
variable qualitative

Exploration  
univariée d'une  
variable quantitative  
entière

- ❖ Présentation des données
- ❖ Étude exploratoire
- ❖ Diagramme en bâtons
- ❖ Fonction de répartition empirique
- ❖ Résumé numérique
- ❖ Tendance centrale
- ❖ Mode
- ❖ Médiane
- ❖ Moyenne
- ❖ Autres indicateurs de position
- ❖ Valeurs extrêmes
- ❖ Quartiles
- ❖ Déciles

# *Exploration univariée d'une variable quantitative entière*



# Présentation des données

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

● population d'effectif  $N$  sur laquelle est définie une variable quantitative discrète  $Z$  à  $c$  modalités  $z_i$ .

➤ Exemple :

Population : les étudiants d'une classe d'effectif  $N=45$ .

Nombre de bonnes réponses au QCM de statistique :

7 6 5 5 6 4 8 3 4 5 4 4 9 5 3 8 7 6 5 8 8 8 6 9 7 7 8 7 6 10 2 8  
9 5 6 6 10 8 7 7 5 6 4 8 7

$Z$  = nombre de bonnes réponses ; variable quantitative discrète de modalités  $z_i : 2/3/4/5/6/7/8/9/10$ . Où  $i = 1, 2, \dots, 9$



# Étude exploratoire

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- tableau de distribution statistique ou tableau de données groupées ou tri à plat
- la distribution statistique associée à chaque valeur de la variable  $Z$ , l'effectif et/ou la fréquence d'individus présentant cette valeur

valeur	effectif	fréquence	effectif cumulé	fréquence cumulée
2	1	2.22	1	2.22
3	2	4.44	3	6.67
4	5	11.11	8	17.78
5	7	15.56	15	33.33
6	8	17.78	23	51.11
7	8	17.78	31	68.89
8	9	20.00	40	88.89
9	3	6.67	43	95.56
10	2	4.44	45	100.00



# Étude exploratoire

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

★ les effectifs sont parfois appelés fréquences absolues, dans ce cas on appelle les fréquences « fréquences relatives » pour éviter toute confusion



# Étude graphique

## Diagramme en bâtons

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

➤ outil le plus adapté pour visualiser une série statistique entière

● pour chaque valeur de la variable on représente un bâton dont la hauteur est la fréquence absolue ou relative (ou moins souvent la fréquence cumulée).

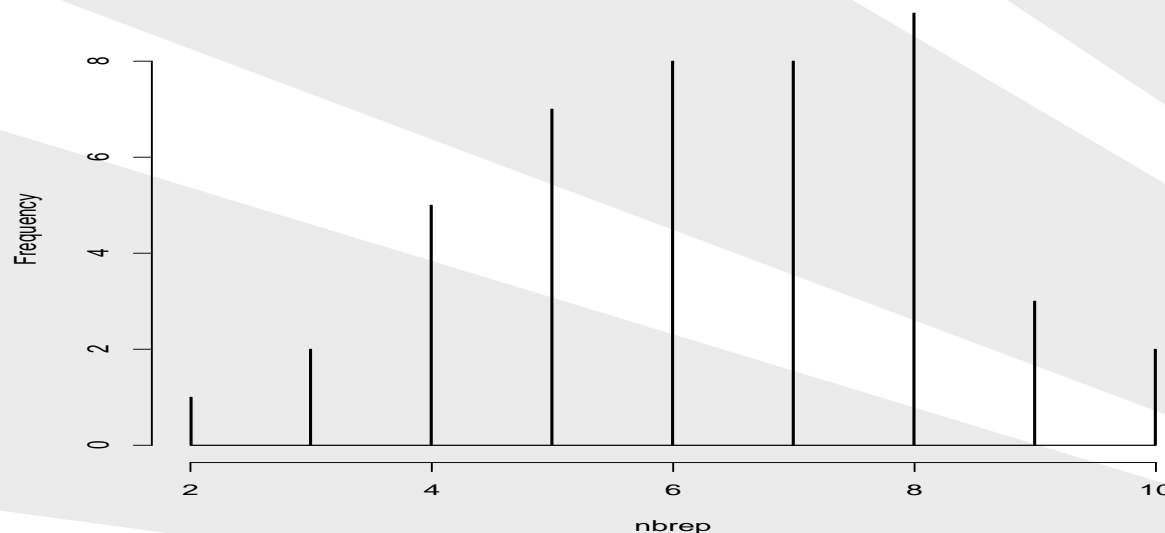


Diagramme en bâtons du nombre de bonnes réponses - Effectifs



# Étude graphique

## Fonction de répartition empirique

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- On peut aussi visualiser la fonction de répartition empirique encore appelée courbe cumulative des fréquences

### Fonction de répartition empirique

$$F_n(x) = \frac{\text{nombre d'éléments dans l'échantillon} \leq x}{n}$$



# Étude graphique

## Fonction de répartition empirique

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

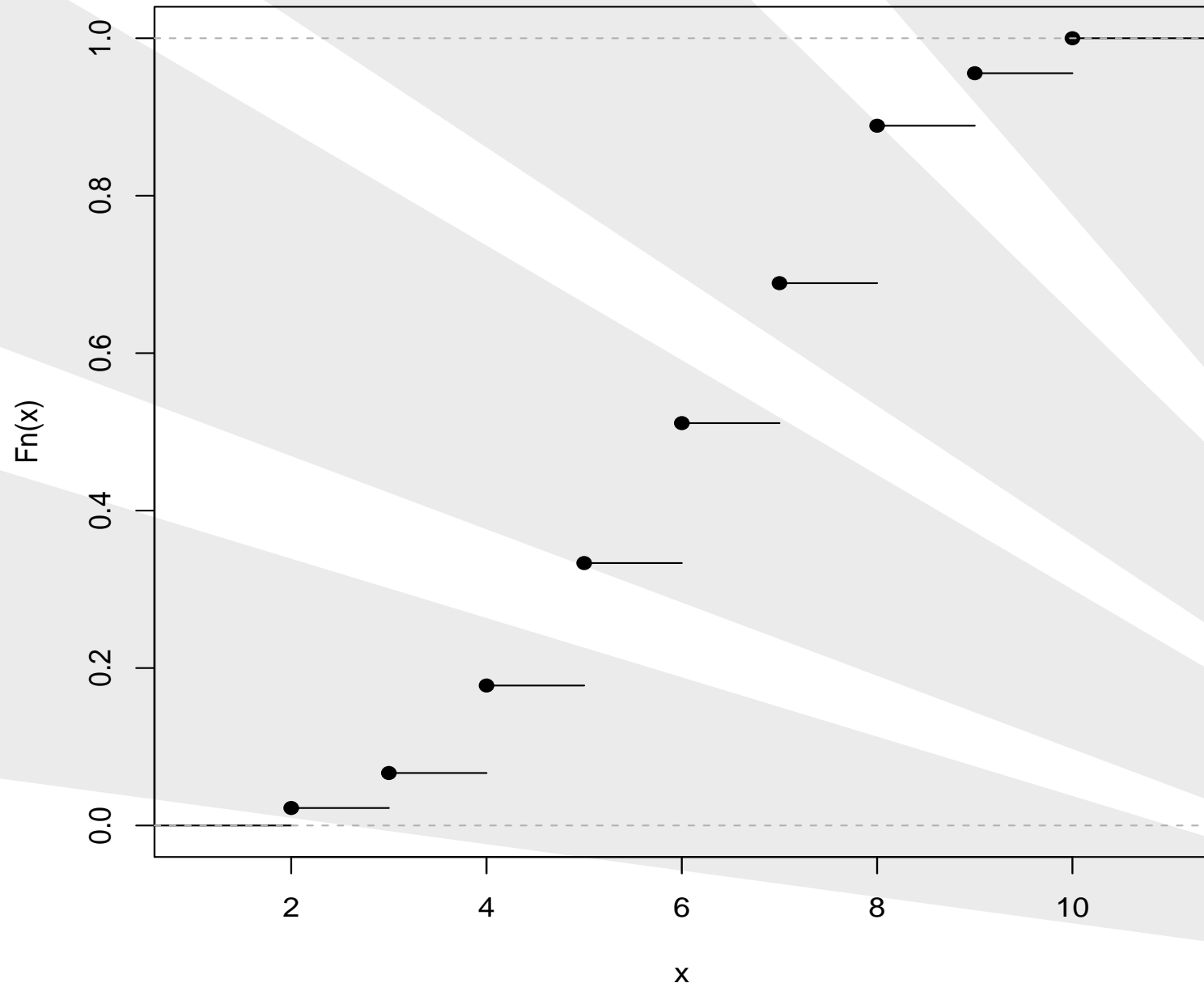
❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles





# Résumé numérique

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendance centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- caractéristiques de tendance centrale
- autres indicateurs de position
- caractéristiques de dispersion
- caractéristiques de forme





# Caractéristiques de tendance centrale

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendance centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- pour caractériser le cœur (centre) de distribution
- la notion de cœur est intuitive mais non rigoureuse
- il existe donc plusieurs indicateurs
  - le mode
  - la médiane
  - la moyenne



## Mode

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendance centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- c'est la valeur la plus fréquente (s'il n'y en a qu'une)
- le mode n'est donc pas toujours défini
- mais il est facile repérer visuellement
- et parfois difficile à calculer
- ★ ne pas confondre mode et maximum (le mode correspond au maximum en ordonnée)
- il peut exister des modes locaux (valeurs plus fréquentes localement)
- dans ce cas on parle de série multimodale (bimodale, trimodale,...)



# Médiane

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- c'est la valeur qui coupe l'échantillon ordonné en deux parties égales
  - c'est-à-dire qui correspond à l'indice moyen pour la série ordonnée
  - c'est un indicateur assez difficile à calculer (il faut trier la série)

## Médiane

Soit la série statistique :  $\{x_1, x_2, \dots, x_r, \dots, x_n\}$

On considère la série ordonnée :

$\{x_{(1)}, x_{(2)}, \dots, x_{(s)}, \dots, x_{(n)}\}$

La médiane vaut :

- si  $n = 2p + 1$ ,  $x_{(p+1)}$
- si  $n = 2p$ ,  $\frac{x_{(p)} + x_{(p+1)}}{2}$



# Fonction de répartition empirique et médiane

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

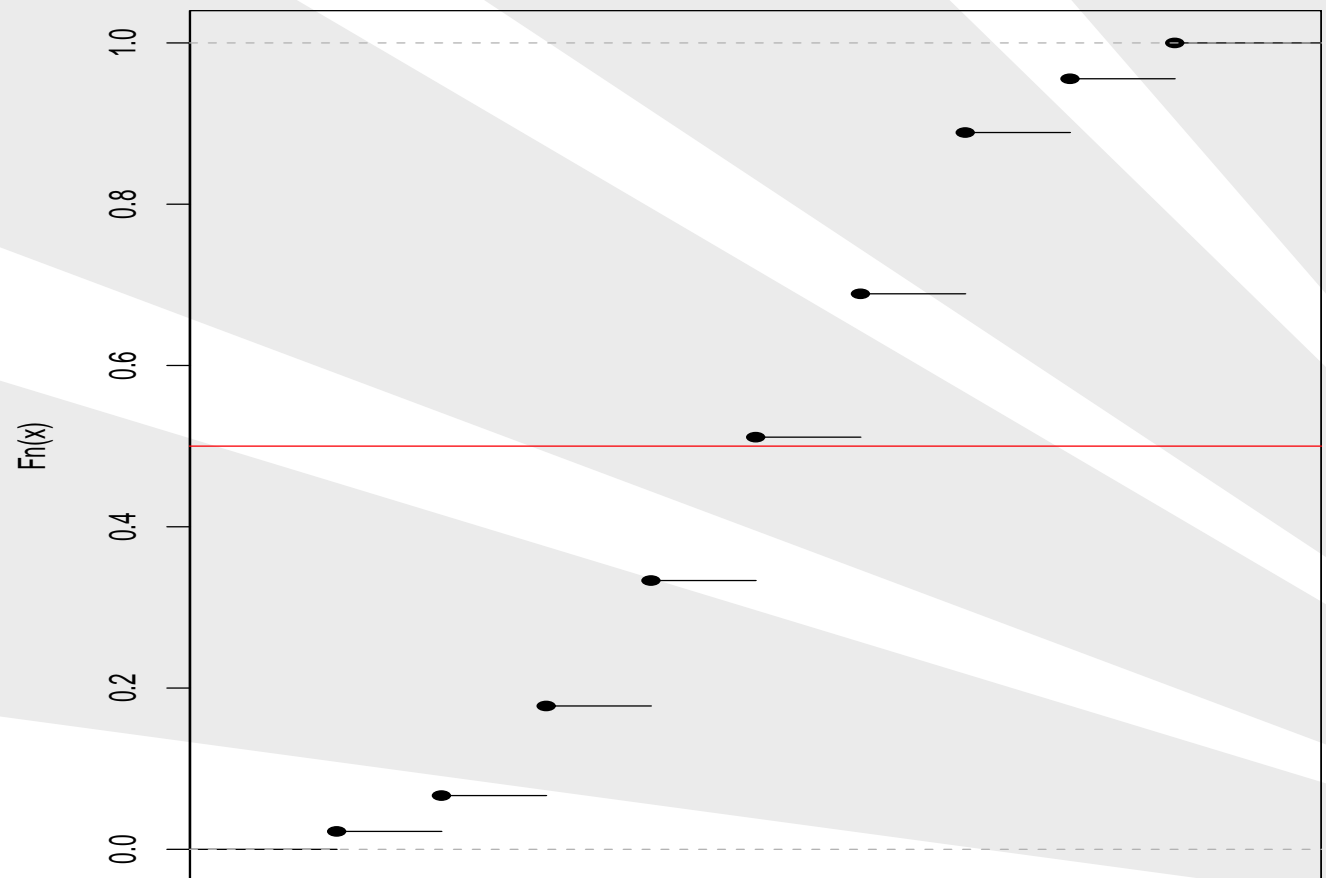
❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- On peut visualiser la médiane sur la fonction de répartition empirique





# Moyenne

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendance centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- c'est l'indicateur le plus classiquement utilisé
- très facile à calculer
- il correspond au barycentre (point d'équilibre de la série)

## Moyenne

$$\bar{x} = \frac{1}{n} \sum_{r=1}^n x_r = \frac{\sum_{i=1}^I n_i x_i}{\sum_{i=1}^I n_i}$$



## Autres indicateurs de position

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- on utilise d'autres indicateurs de position pour identifier des positions non centrales de la série

- proportions particulières

- extrémités

- ....



# Autres indicateurs de position

## Valeurs extrêmes

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendance centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

● il s'agit tout simplement

- du minimum
- du maximum

de la série



# Autres indicateurs de position

## Quartiles

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- on généralise la médiane qui coupe en deux la série
- aux quartiles qui coupent en quatre
- il y a donc 3 quartiles
- ★ dont le deuxième est la médiane
- on considère donc le premier et le dernier (troisième quartile)
- le calcul exact des quartiles se fait par interpolation linéaire





# Autres indicateurs de position

## Déciles

❖ Méthode d'analyse

Exploration univariée d'une variable qualitative

Exploration univariée d'une variable quantitative entière

❖ Présentation des données

❖ Étude exploratoire

❖ Diagramme en bâtons

❖ Fonction de répartition empirique

❖ Résumé numérique

❖ Tendances centrale

❖ Mode

❖ Médiane

❖ Moyenne

❖ Autres indicateurs de position

❖ Valeurs extrêmes

❖ Quartiles

❖ Déciles

- on « coupe » la série en 10
- on s'intéresse essentiellement aux déciles extrêmes
  - premier et dernier décile
  - pour avoir une idée de l'extension de la série