

# Chapitre 4

## Statistique descriptive

# Echantillonnage statistique

On appelle **population**, un ensemble d'individus auquel on s'intéresse

La plupart du temps, la population d'intérêt est trop grande pour qu'on puisse prélever tout le monde. On se restreint à un certain nombre d'individus: un **échantillon**

**L'échantillonnage** est le fait de prélever un échantillon

**L'échantillonnage aléatoire simple** consiste à prélever au hasard et de façon indépendante  $n$  individus dans une population

# Echantillonnage statistique

Exemple: on souhaite connaître le poids à l'âge adulte des chats forestiers en France

**L'échantillonnage aléatoire simple** : on prend tous les chats forestiers de France, on leur donne un nom (ou un numéro), on met le tout dans une urne, on tire des noms au hasard et on fait en sorte de capturer les chats forestiers correspondants

Problèmes:

- 1) On ne connaît pas tous les chats
- 2) On ne peut pas les capturer sur commande
- 3) Certains chats sont plus difficiles à capturer que d'autres (les plus légers ?)

# Echantillonnage statistique

Solution 1: On pose des pièges dans différents endroits un peu partout en France

On couvre tous les types de milieu qui peuvent influencer sur le poids (type de végétation, climat,...)

On parle d'échantillon **représentatif**

Solution 2: On est moins exigeant, on **cible** la question sur une zone en particulier (poids des chats forestiers dans la forêt de bidule)

# La difficulté du traitement des données

| Nom du chat | Minou | Ronron | Patpat | Médor | Minou2 | Gratgrat | Toumoche | ... |
|-------------|-------|--------|--------|-------|--------|----------|----------|-----|
| Poids en Kg | 4,2   | 6,3    | 5,1    | 6,4   | 7,3    | 5,9      | 4,6      |     |

Problème: si on capture 230 chats le tableau n'est plus gérable

⇒ Besoin d'outils pour

- 1) Visualiser les données
- 2) Les résumer à l'aide d'indicateurs pertinents
- 3) Répondre à des questions théoriques (ex: les chats sauvages sont-ils plus gros que les chats domestiques ?)

STATISTIQUES  
DESCRIPTIVES  
(cours d'aujourd'hui)

STATISTIQUES  
INFÉRENTIELLES  
(cours suivants)

# Caractère statistique

## 1. Caractères qualitatifs

Caractère mesuré dans une échelle : nominale (type de globules blancs)  
ordinaire (stade d'une maladie)

## 2. Caractères quantitatifs discrets

Caractère qui ne peut prendre que certaines valeurs sur un intervalle  
(ex: taille d'une portée)

## 3. Caractères quantitatifs continus

Caractère pouvant prendre toutes les valeurs sur un intervalle donné  
(ex: taille, poids, âge,...)

# Caractère discret: quelques notations

Le caractère  $X$  peut valoir  $x_1, x_2, \dots, x_n$ . La  $i$ -ème valeur possible est appelée modalité  $x_i$ .

On appelle **effectif** de la modalité  $x_i$ , le nombre  $n_i$  d'individus qui ont la valeur  $x_i$  pour le caractère  $X$

On appelle **fréquence** de la modalité  $x_i$ , le nombre

$$f_i = \frac{n_i}{n}$$

On appelle **fréquence cumulée** en  $x_i$ , le nombre

$$f_i \text{ cum} = \sum_{p=1}^i f_p$$

# Tableau statistique

## 1. Caractère quantitatif discret

| Caractère $X$ :<br>$x_i$ : nombre de galles<br>par feuille | 0     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10 |
|--|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----|
| $n_i$ : nombre de feuilles<br>portant $x_i$ galles         | 182   | 98    | 46    | 28    | 12    | 5     | 2     | 1     | 0     | 1     | 0  |
| $f_i$ : fréq. relative                                     | 0,485 | 0,261 | 0,123 | 0,075 | 0,032 | 0,013 | 0,005 | 0,003 | 0     | 0,003 | 0  |
| <b><math>f_i</math> cum.</b> : fréq. relative<br>cumulée   | 0,485 | 0,746 | 0,869 | 0,944 | 0,976 | 0,989 | 0,994 | 0,997 | 0,997 | 1     | 1  |

$n = 375$  feuilles de hêtre examinées



# Tableau statistique

## 2. Caractère quantitatif continu

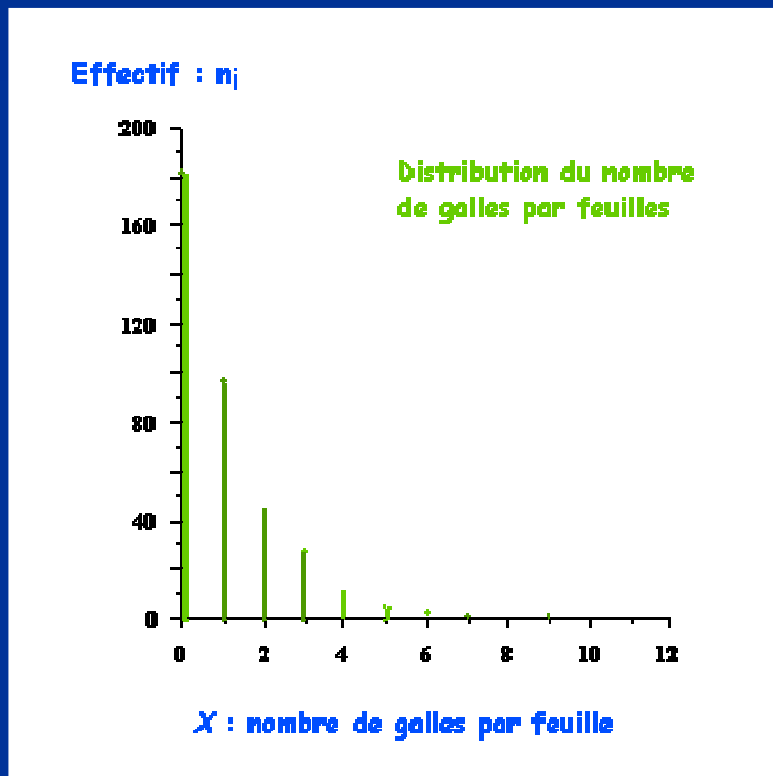
Une manière de représenter une variable continue sous forme de tableau consiste à regrouper les observations en classes

| Caractère $X$ :                                       | [140-145[ | [145-150[ | [150-155[ | [155-160[ | [160-165[ | [165-170[ | [170-175[ |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $x_i$ : longueur de la rectrice<br>bornes des classes |           |           |           |           |           |           |           |
| Valeur médiane des classes, $x'_i$                    | 142,5     | 147,5     | 152,5     | 157,5     | 162,5     | 167,5     | 172,5     |
| $n_i$ : nombre d'individu par classe<br>de taille $x$ | 1         | 1         | 9         | 17        | 16        | 3         | 3         |
| $f_i$ : fréquence relative                            | 0,02      | 0,02      | 0,18      | 0,34      | 0,32      | 0,06      | 0,06      |
| $f_{cum.}$ : fréquence relative<br>cumulée            | 0,02      | 0,04      | 0,22      | 0,56      | 0,88      | 0,94      | 1         |

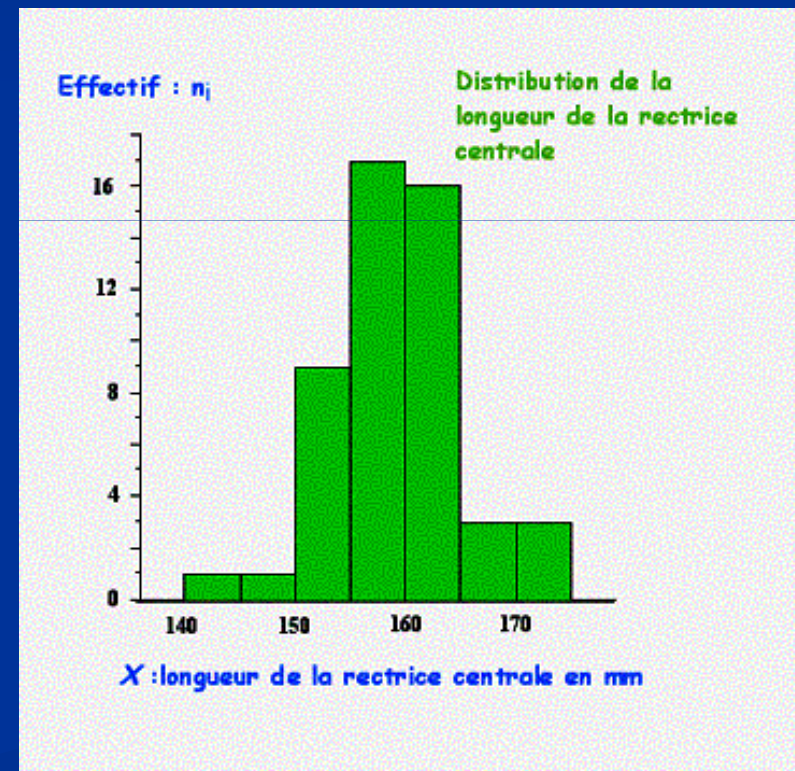
$n = 50$  rectrices examinées

# Représentation graphique

## 1. Caractère discret



## 2. Caractère continu

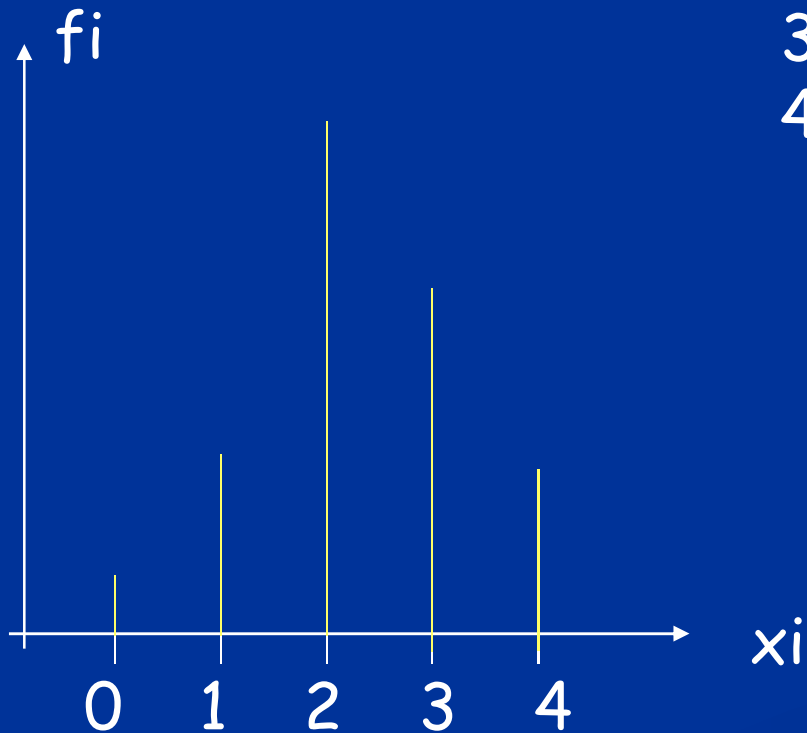


## Ex. nombre d'enfants dans une famille

3 2 2 3 2  
4 4 1 2 2  
4 3 2 0 2  
2 1 3 3 1



| $X$ | $n_i$ | $f_i$ | $\Sigma f_i$ |
|-----|-------|-------|--------------|
| 0   | 1     | 0.05  | 0.05         |
| 1   | 3     | 0.15  | 0.20         |
| 2   | 8     | 0.40  | 0.60         |
| 3   | 5     | 0.25  | 0.85         |
| 4   | 3     | 0.15  | 1            |
|     | <hr/> | <hr/> |              |
|     | 20    | 1     |              |



# Caractère quantitatifs: indicateurs statistiques

## 1. Indicateurs de position

- La moyenne arithmétique

Soit un échantillon de  $n$  valeurs observées  $x_1, x_2, \dots, x_i, \dots, x_n$  d'un caractère quantitatif  $X$

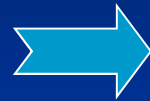
Données non groupées  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Données groupées pour caractère discret  $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i$

Données groupées pour caractère continu  $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i'$

## Ex. nombre d'enfants dans une famille

3 2 2 3 2  
4 4 1 2 2  
4 3 2 0 2  
2 1 3 3 1



| X | ni | fi   | Σfi  |
|---|----|------|------|
| 0 | 1  | 0.05 | 0.05 |
| 1 | 3  | 0.15 | 0.20 |
| 2 | 8  | 0.40 | 0.60 |
| 3 | 5  | 0.25 | 0.85 |
| 4 | 3  | 0.15 | 1    |

On peut faire le calcul à partir des données non groupées:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{20} (3+2+2+3+2+4+4+1+2+2+4+3+2+0+2+2+1+3+3+1) = 2.3$$

En fait on a additionné une fois 0, trois fois 1, huit fois 2, cinq fois 3 et trois fois 4, donc

$$\bar{x} = \frac{1}{20} (1 \times 0 + 3 \times 1 + 8 \times 2 + 5 \times 3 + 3 \times 4) = 2.3$$



Formule données  
groupées

## Attention:

Pour un caractère continu, la moyenne non groupée n'est pas égale à la moyenne après regroupement

# Caractère quantitatifs: indicateurs statistiques

## 1. Indicateurs de position

### ▪ La médiane

La médiane est la valeur du caractère pour laquelle la fréquence cumulée est de 0,5. Soit  $n$ , le nombre de mesures:

Si  $n$  est impair, la médiane est la valeur au rang  $(n+1)/2$

Si  $n$  est pair, la médiane est la **moyenne** des valeurs aux rangs  $n/2$  et  $(n/2) + 1$

### ▪ Le mode

Le mode est la valeur du caractère la plus fréquente

# Indicateurs statistiques

## 2. Indicateurs de dispersion

### ▪ La variance

Soit un échantillon de  $n$  valeurs observées  $x_1, x_2, \dots, x_i, \dots, x_n$  d'un caractère quantitatif  $X$

Données non groupées 
$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

Données groupées pour caractère discret 
$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$$

Données groupées pour caractère continu 
$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i (x'_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i'^2 - \bar{x}^2$$