

# CHAPITRE 1

INTRODUCTION

# OBJECTIF DU COURS

- La statistique = l'ensemble d'outils mathématiques qui permettent de...



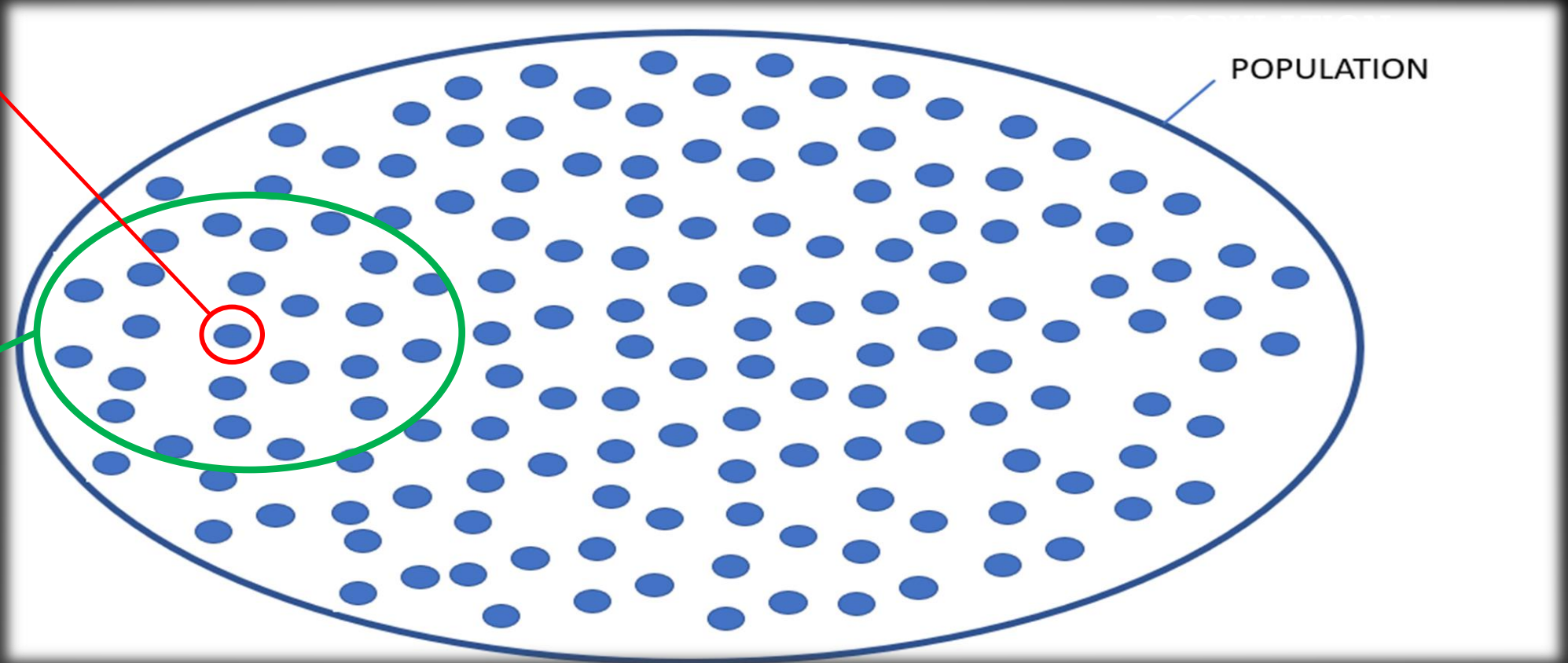
# RÉCOLTER LES DONNÉES

INDIVIDU  
(ou sujet)

POPULATION

ECHANTILLON

Collecte  
de **variables**



# RÉCOLTER LES DONNÉES

- Les variables seront encodées dans une **base de données**.
- La plupart du temps, on encodera

Une ligne par individu (ou sujet)

Une colonne par variable



## EXEMPLE

Données brutes

i	genre ( $W_i$ )	âge ( $X_i$ )	anciennete ( $Y_i$ )	langue ( $Z_i$ )
1	Masculine	19	1	Allemande
2	Masculine	25	5	Français
3	Féminin	21	2	Français
4	Masculine	27	8	Néerlandais
5	Masculine	30	8	Français
6	Masculine	32	9	Néerlandais
7	Féminin	30	9	Allemande
8	Féminin	19	1	Néerlandais
9	Féminin	21	1	Néerlandais
10	Masculine	21	1	Français
11	Masculine	24	8	français
12	Féminin	27	8	néerlandais
13	Féminin	26	7	néerlandais
14	Masculine	30	10	néerlandais
15	Féminin	26	8	néerlandais

Combien d'individus?  
Combien de variables?  
Combien vaut  $Y_3$ ?  $Z_4$ ?  
Comment note-t-on l'âge  
du 8<sup>ème</sup> sujet?

# TRAITER LES DONNÉES

= statistique **DESCRIPTIVE**

= simplifier l'information des données brutes

➡ Grâce à des graphiques (Chapitre 3)

➡ Ou des valeurs algébriques (Chapitre 4)

# TRAITER LES DONNÉES

Le choix du **graphique** ou des **valeurs algébriques**  
dépendra de la

**nature des variables** (chapitre 2)

# CHAPITRE 2

MESURE DES VARIABLES



# MESURE DES VARIABLES

- Qualitatives vs. Quantitative
  - Qualitatives = des catégories
  - Quantitative = des valeurs chiffrées

# VARIABLES QUALITATIVES

Nominale	Ordinale
Catégories Valeurs <b>non</b> intrinsèquement numériques	
<b>Sans</b> ordre logique	<b>Avec</b> ordre logique

# VARIABLES QUANTITATIVES

Discrète	Continue
Valeurs intrinsèquement numériques	
Valeurs <b>isolées</b> , généralement <b>entières</b>	N'importe quelle valeur est possible, dans un intervalle de valeurs parmi les réels

# EXERCICE

Pour chacune des variables suivantes, déterminez sur quel type d'échelle elles ont été mesurées

	Qualitative		Quantitative	
	Nominale	Ordinale	Discrète	Continue
Statut matrimonial				
Âge				
Genre				
Activité sportive (nulle, légère ou forte)				
Taille (en cm)				
Nombre d'enfants à charge				
Salaire				

# CHAPITRE 3

EXPLORATION GRAPHIQUE DES DONNÉES À UNE DIMENSION



# REPRÉSENTATION GRAPHIQUE

- Constitue ...
  - ... un accès visuel facile et immédiat sur les données
  - ... une représentation souvent **simplifiée** des données
- Étude des graphiques basés sur
  - Les tableaux de fréquences
  - Les quantiles
- Dépend de la **nature** des données

# GRAPHIQUE SUR BASE DES TABLEAUX DE FRÉQUENCES

- Étape 1 = établir un tableau de **distribution de fréquences (= simplification)**
- Étape 2 = Représentation graphique



# VARIABLES NOMINALES

## Tableau de fréquences

- 1) Déterminer les modalités de la variable (= les valeurs possibles)
- 2) Créer un tableau contenant autant de lignes que de modalités
- 3) Calculer les fréquences absolues et relatives
  - Fréquence absolue ( $n_j$ ) = nombre de valeur obtenant chaque valeur
  - Fréquence relative ( $f_j$ ) = proportion que ce nombre représente par rapport à l'échantillon total



# VARIABLES NOMINALES

**Tableau de fréquences**

<b>j</b>	<b><math>x_j</math></b>	<b><math>n_j</math></b>	<b><math>f_j</math></b>
1	Non	30	0.75
2	Oui	10	0.25

# VARIABLES NOMINALES

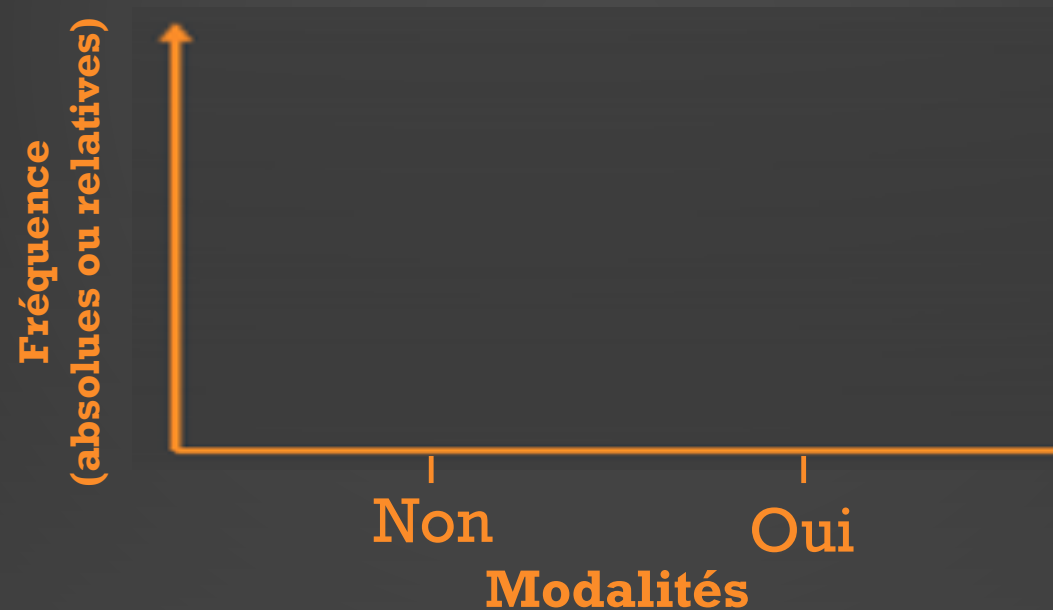
## Représentations graphiques

- 1) Diagramme en bâtons
- 2) Diagramme circulaire

# VARIABLES NOMINALES

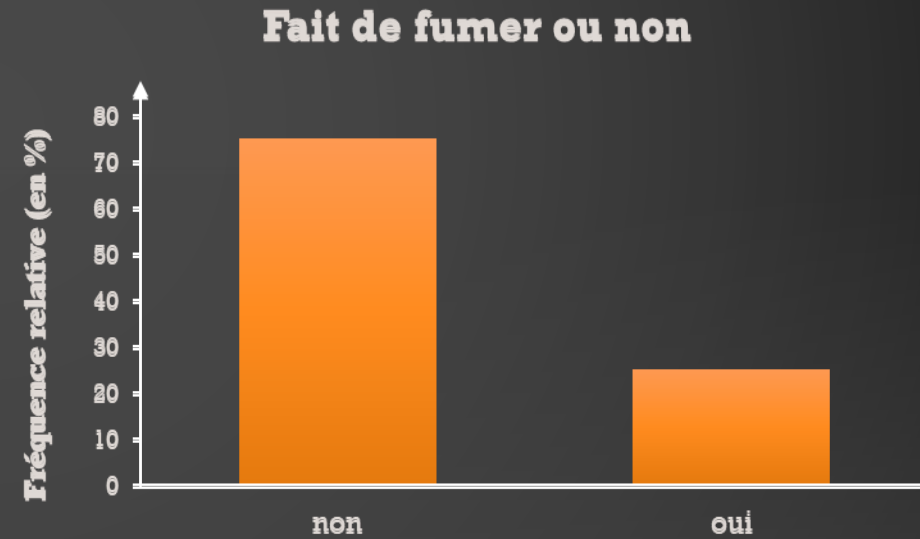
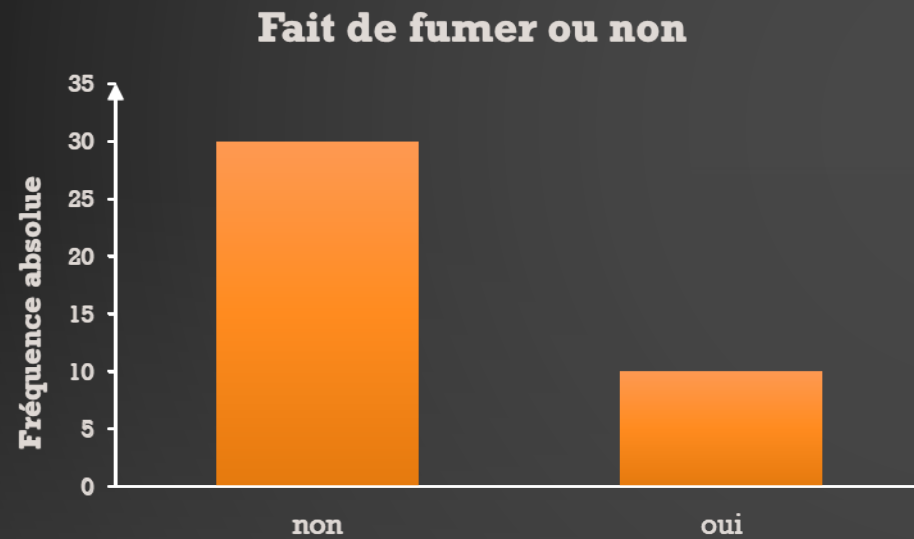
## Diagramme en bâtons

Fait de fumer ou non



# VARIABLES NOMINALES

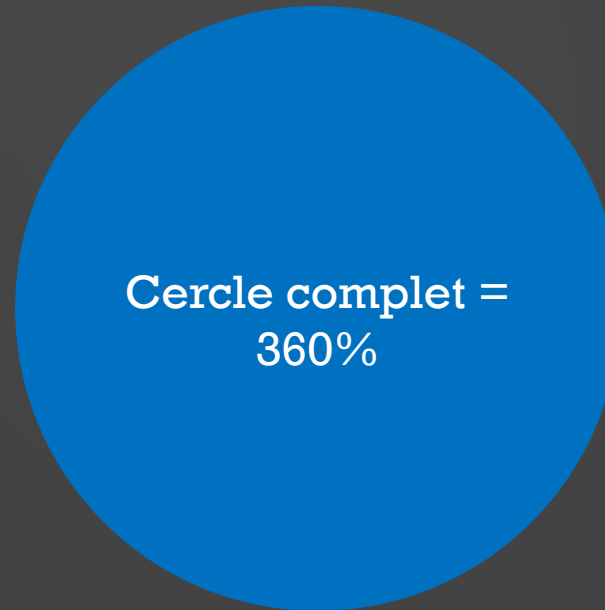
## Diagramme en bâtons



# VARIABLES NOMINALES

## Diagramme circulaire

Fait de fumer ou non



$$\begin{array}{l} \text{40 sujets} = 360^\circ \\ \text{1 sujet} = 9^\circ \end{array}$$

$$30 \text{ sujets} = 9^\circ \times 30 = 270^\circ$$

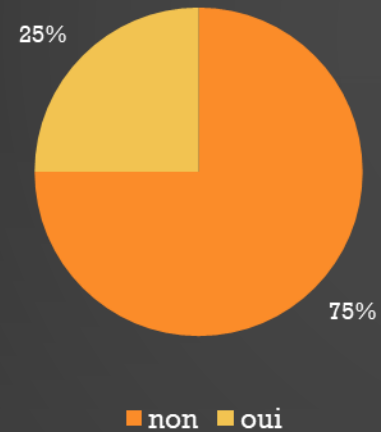
$$10 \text{ sujets} = 9^\circ \times 10 = 90^\circ$$



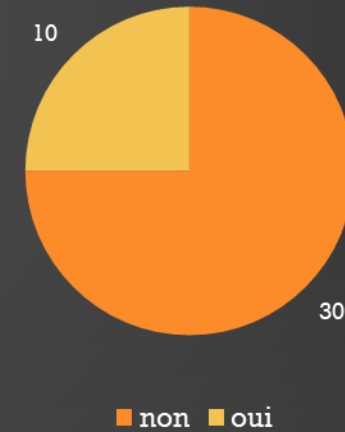
# VARIABLES NOMINALES

## Diagramme circulaire

Fait de fumer ou non



Fait de fumer ou non



# VARIABLES NOMINALES

## Diagramme en bâtons et diagramme circulaire

- La forme du graphe est la même!
- Ce qui change, c'est l'information donnée

	Distribution en fréquence absolue	Distribution en fréquence relative
<u>Information donnée:</u>	Dénombrer le nombre (non) fumeurs	Déterminer la proportion de (non) fumeurs parmi l'échantillon

# VARIABLES ORDINALE

- Imaginez que l'on connaisse la taille de T-shirt de 40 enfants de 4<sup>ème</sup> primaire

S	M	M	M	S	XS	XS	XS	S	S
M	S	S	M	S	S	M	M	M	M
M	M	M	M	S	S	M	M	M	L
L	L	L	L	S	S	S	S	S	S

# VARIABLES ORDINALES

**Tableau de fréquences**

j	$x_j$	$n_j$	$f_j$
1	XS	3	7,5%
2	S	16	40%
3	M	16	40%
4	L	5	12,5%

**Attention: contrairement au cas des variables nominales, l'ordre des catégories importe!**

# VARIABLES ORDINALES

## Représentations graphiques

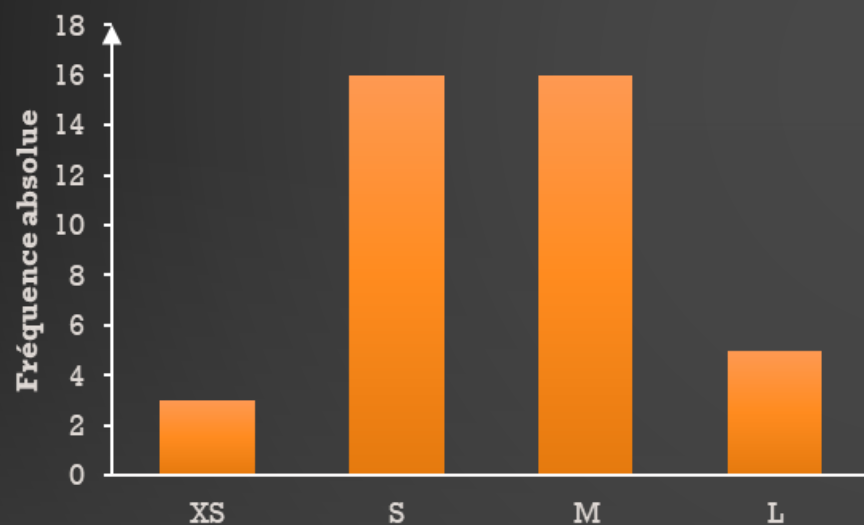
- 1) Diagramme en bâtons
- 2) Diagramme circulaire



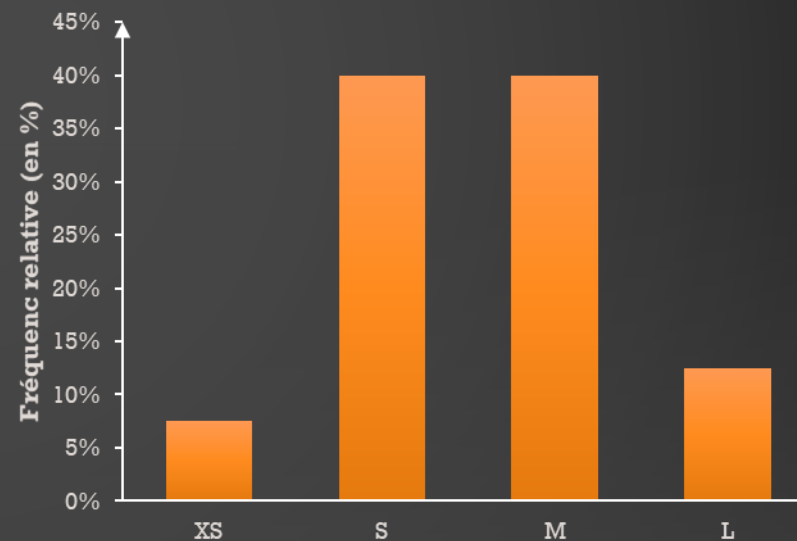
# VARIABLES ORDINALES

## Diagramme en bâtons

Taille de T-shirt d'enfants  
de 4ème primaire



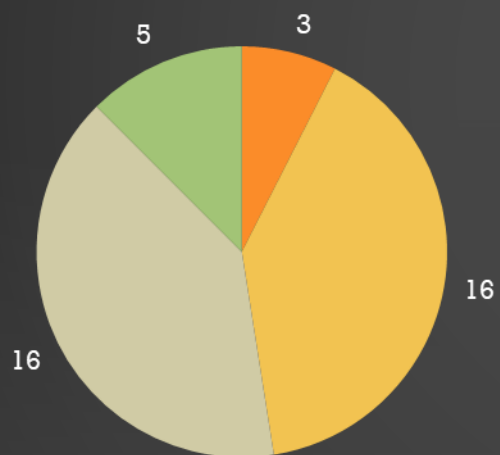
Taille de T-shirt d'enfants de  
4ème primaire



# VARIABLES ORDINALES

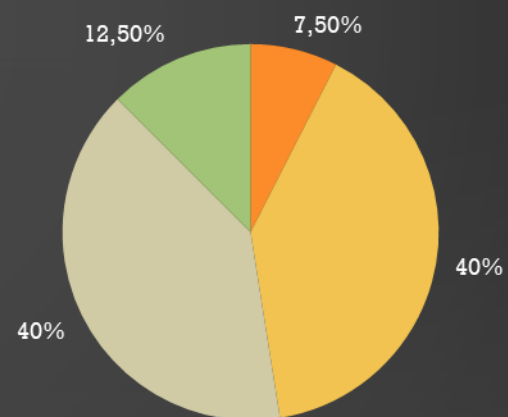
## Diagramme circulaire

Taille de T-shirt d'enfants de  
4ème primaire



■ XS ■ S ■ M ■ L

Taille de T-shirt d'enfants de  
4ème primaire



■ XS ■ S ■ M ■ L

# VARIABLES ORDINALES

## Diagramme en bâtons et diagramme circulaire

- La forme du graphe est la même!
- Ce qui change, c'est l'information donnée

	Distribution en fréquence absolue	Distribution en fréquence relative
<u>Information donnée:</u>	Dénombrer le nombre associé à chaque taille de T-shirt	Déterminer la proportion de (chaque taille de T-shirt parmi l'échantillon

# VARIABLES QUANTITATIVES DISCRÈTES

- Imaginez que l'on connaisse le nombre d'enfants au sein de 40 ménages

5	2	2	3	1	2	1	2	2	2
1	3	3	1	4	3	3	1	2	1
1	1	2	2	1	2	3	1	3	4
1	1	2	2	1	2	3	1	3	4

# VARIABLES QUANTITATIVES DISCRÈTES

**Tableau de fréquences**

j	$x_j$	$n_j$	$f_j$ (en %)	$N_j$	$F_j$ (en %)
1	1	14	35	14	35
2	2	13	32.5	27	67.5
3	3	9	22.5	36	90
4	4	3	7.5	39	97.5
5	5	1	2.5	40	100

# VARIABLES QUANTITATIVES DISCRÈTES

## Représentations graphiques

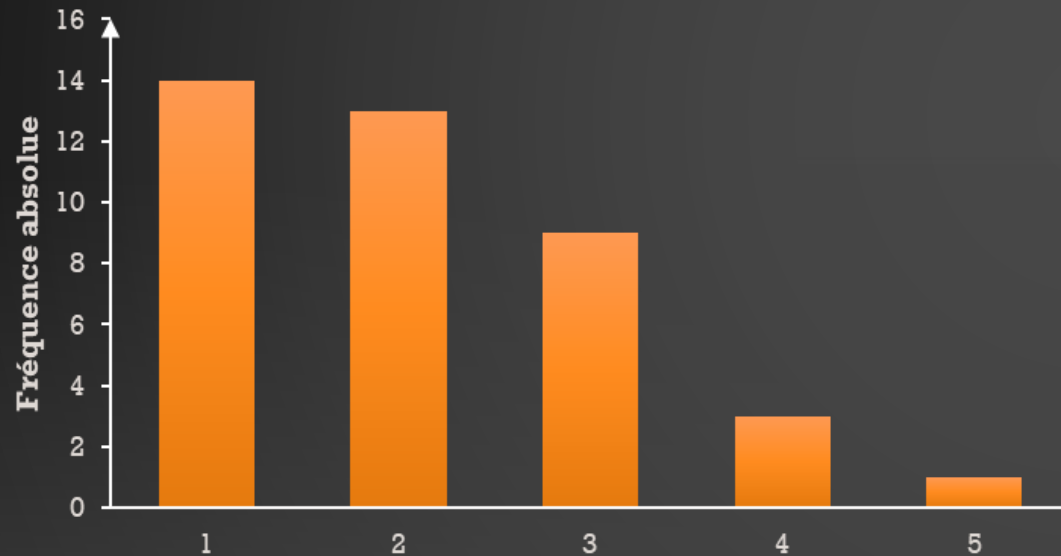
- 1) Diagramme en bâtons
- 2) Diagramme circulaire
- 3) Diagramme des fréquences cumulées
- 4) Diagramme en escalier



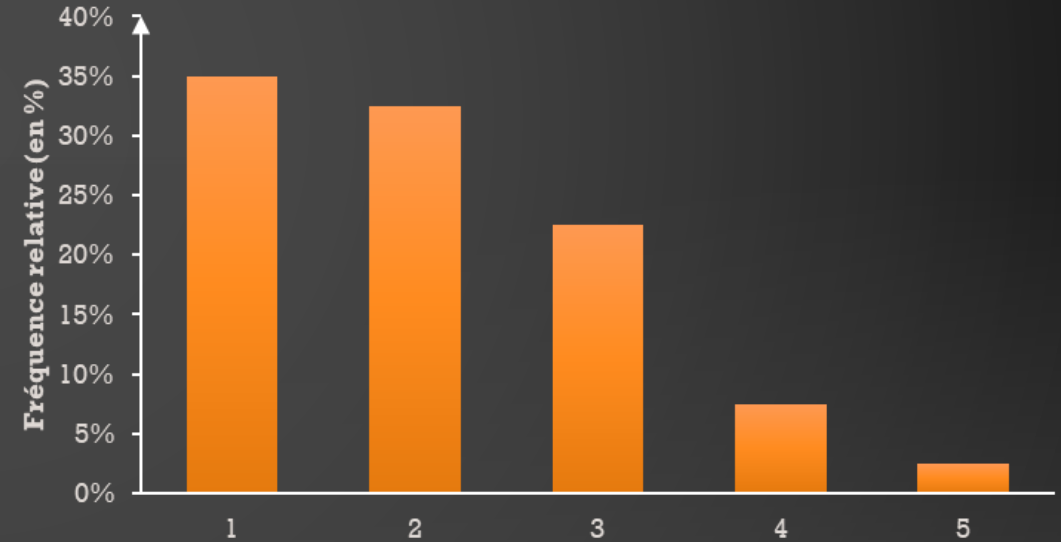
# VARIABLES QUANTITATIVES DISCRÈTES

## Diagramme en bâtons

Nombre d'enfants par ménage



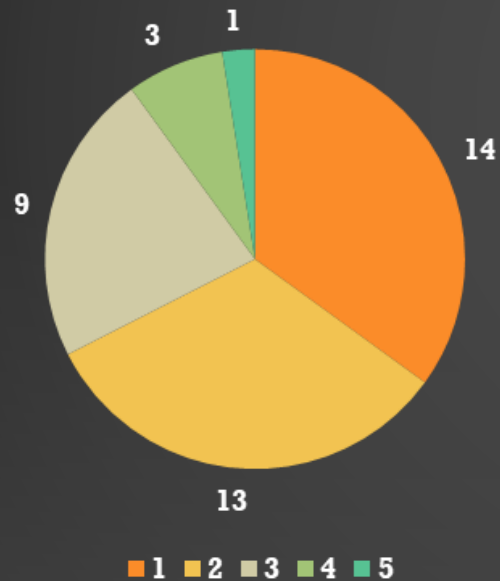
Nombre d'enfants par ménage



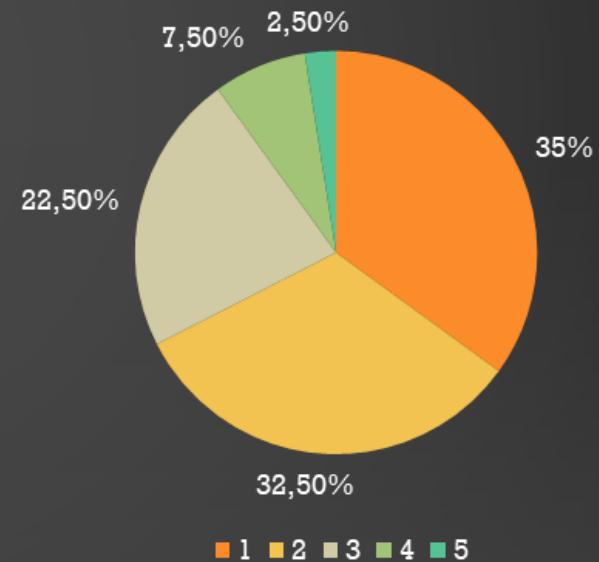
# VARIABLES QUANTITATIVES DISCRÈTES

## Diagramme circulaire

Nombre d'enfants par ménage



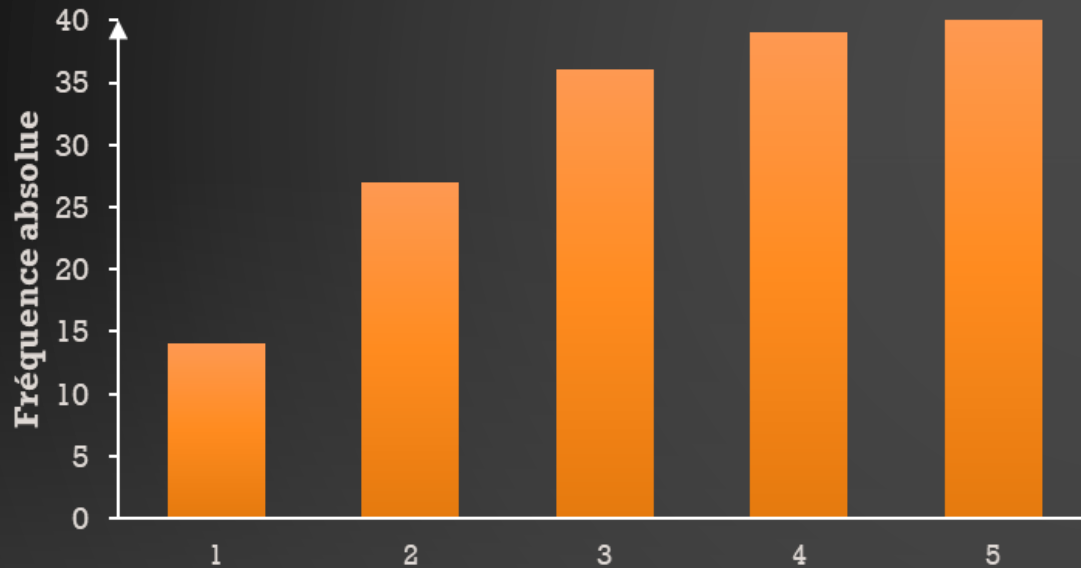
Nombre d'enfants par ménage



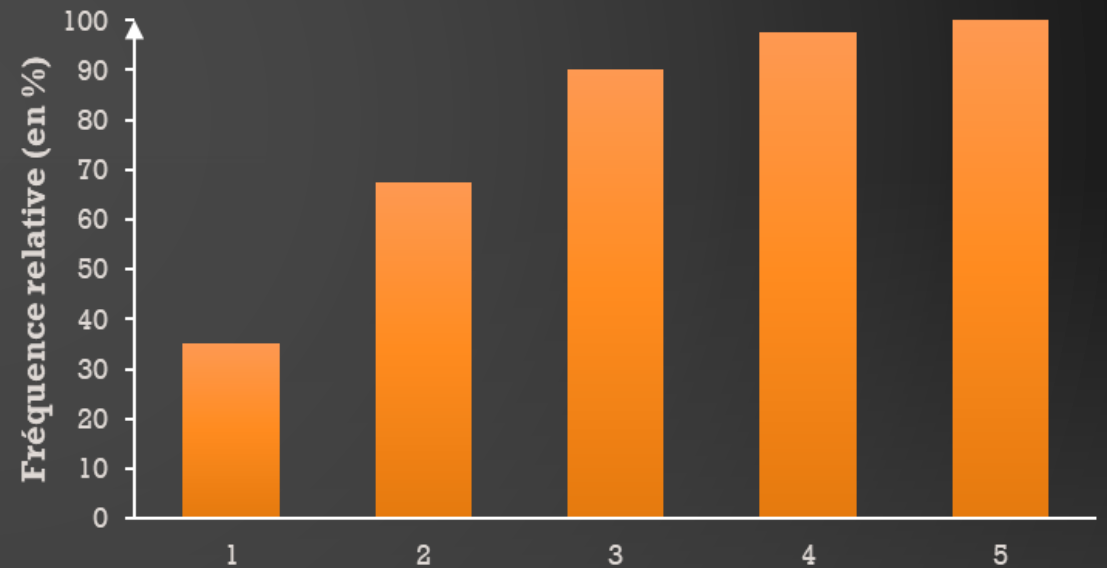
# VARIABLES QUANTITATIVES DISCRÈTES

## Diagramme des fréquences cumulées

Nombre d'enfants par ménage

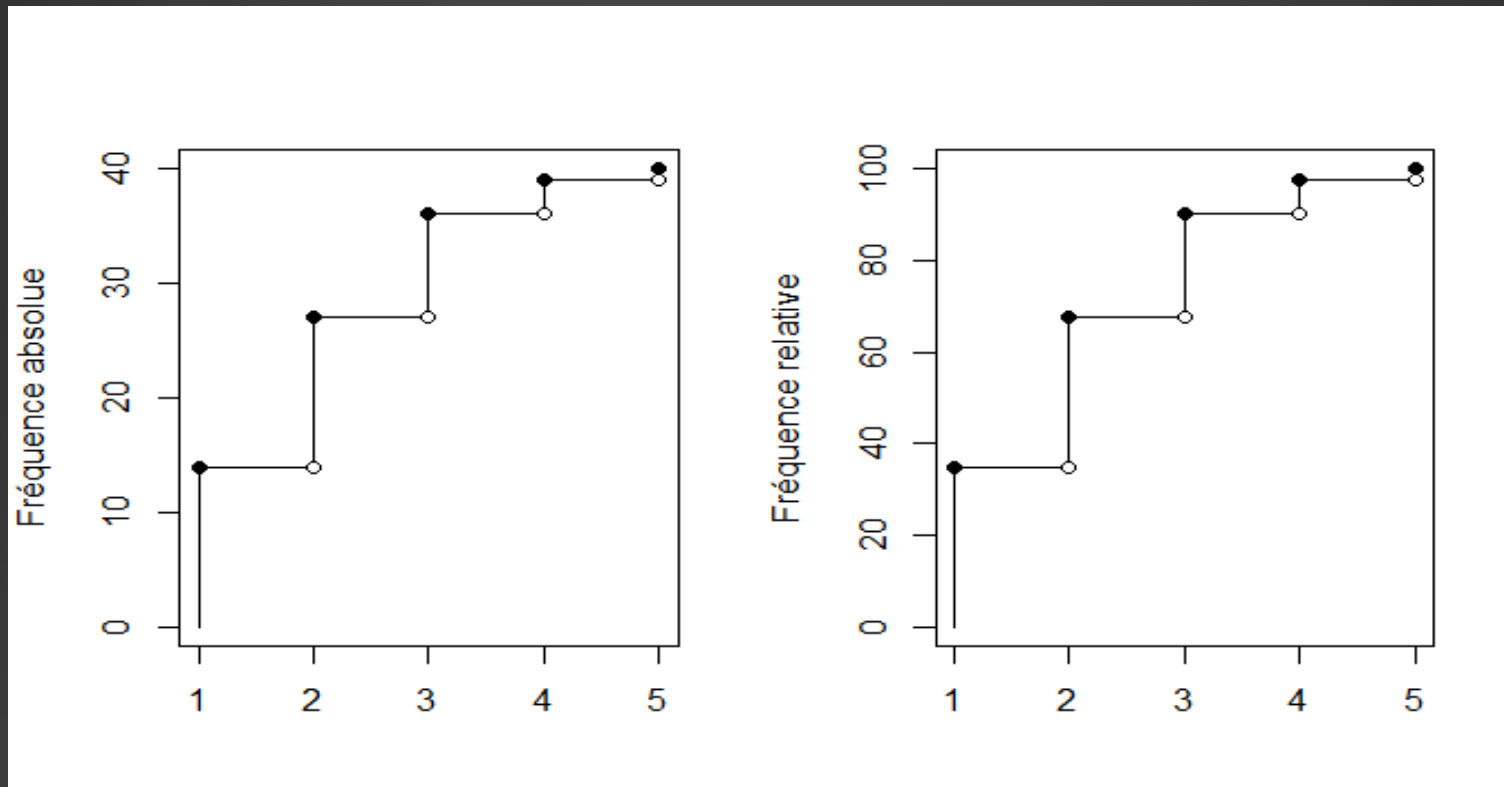


Nombre d'enfants par ménage



# VARIABLES QUANTITATIVES DISCRÈTES

## Diagramme en escalier



# VARIABLES QUANTITATIVES CONTINUES

**Tableau de fréquences**

j	$x_j$	Centre de classe	$n_j$	$f_j$ (en %)	$N_j$	$F_j$ (en %)
1	[0 – 20K]	10K	20	44.44	20	44.44
2	[20K-40K]	30K	9	20	29	64.44
3	[40K-60K]	50K	11	24.44	40	88.89
4	[60K-80K]	70K	3	6.67	43	95.56
5	[80K-100K]	90K	1	2.22	44	97.78
6	[100K-120K]	110K	1	2.22	45	100

# VARIABLES QUANTITATIVES CONTINUES

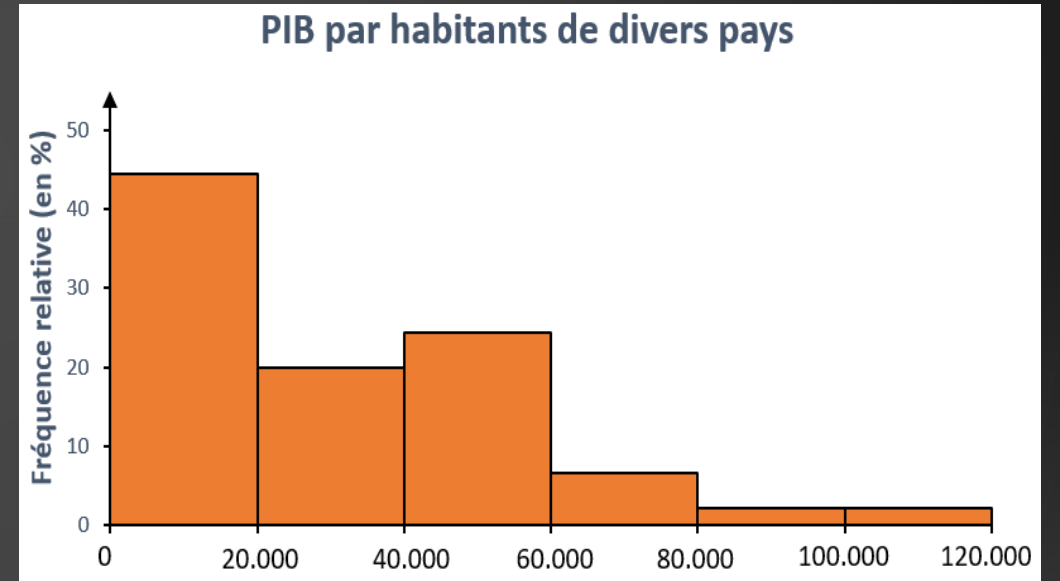
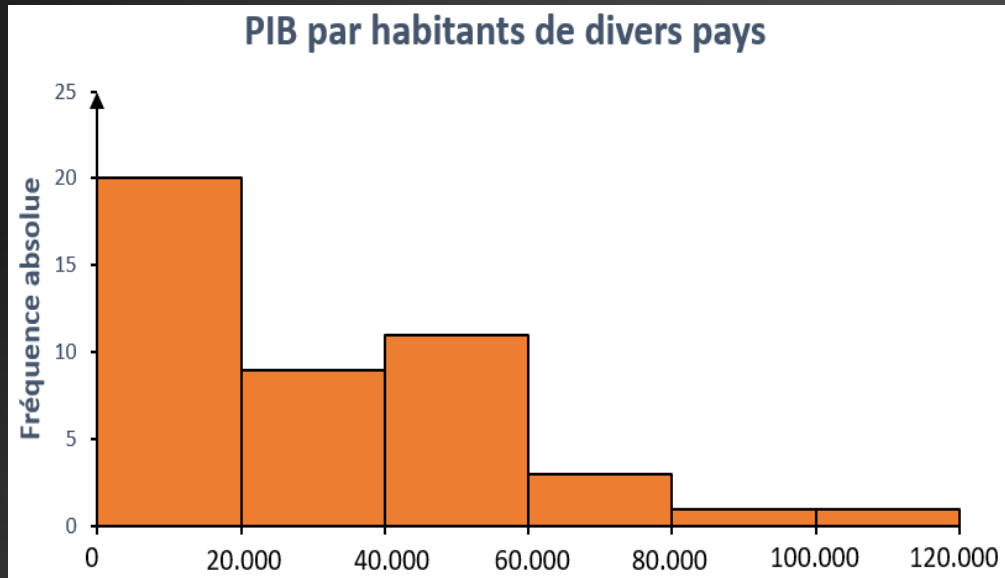
## Représentations graphiques

- 1) Histogramme (avec ou sans polygone des effectifs)
- 2) Histogramme des fréquences cumulées (avec ou sans polygone des effectifs cumulés)
- 3) Distributions



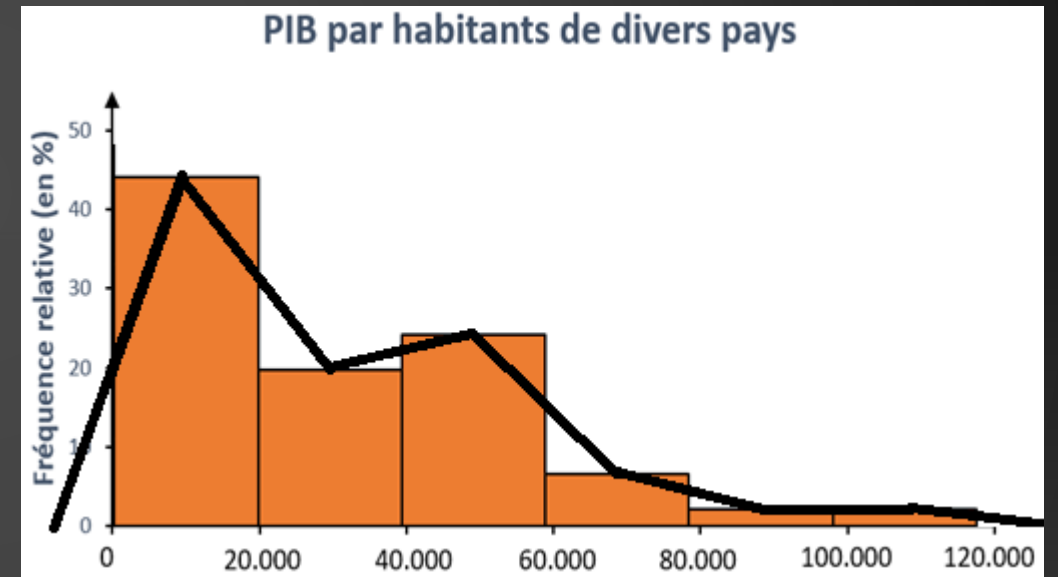
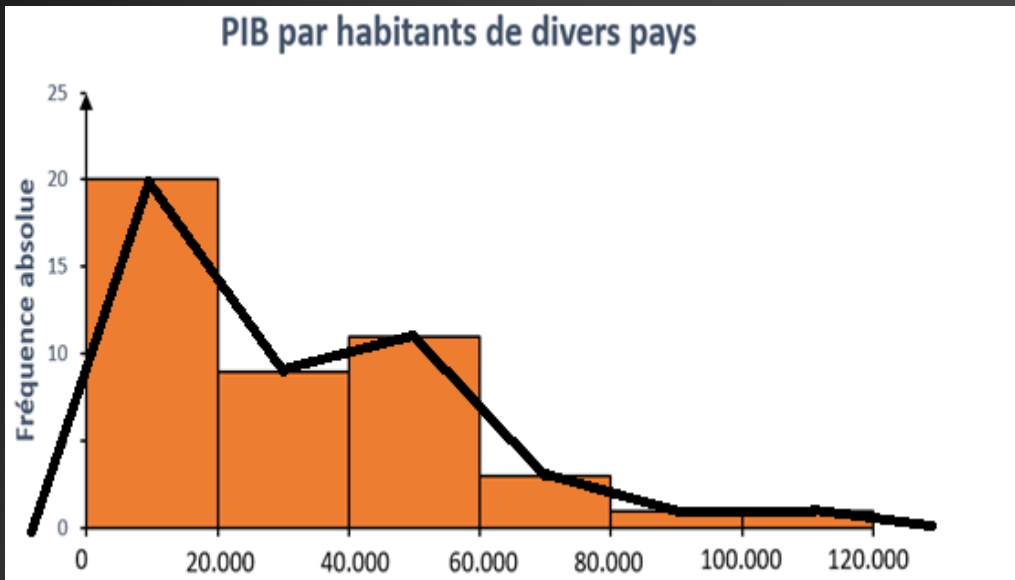
# VARIABLES QUANTITATIVES CONTINUES

## Histogrammes



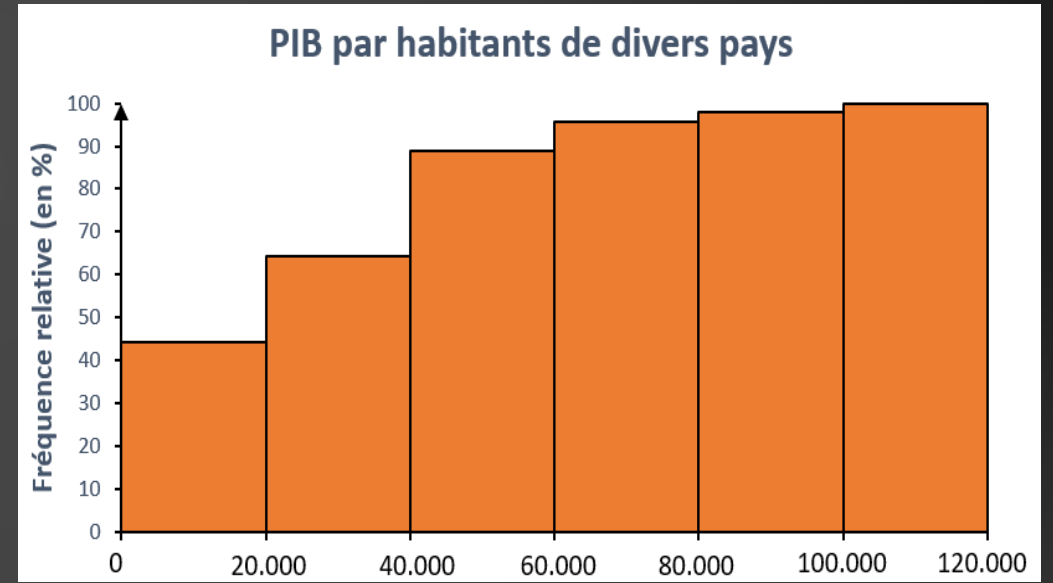
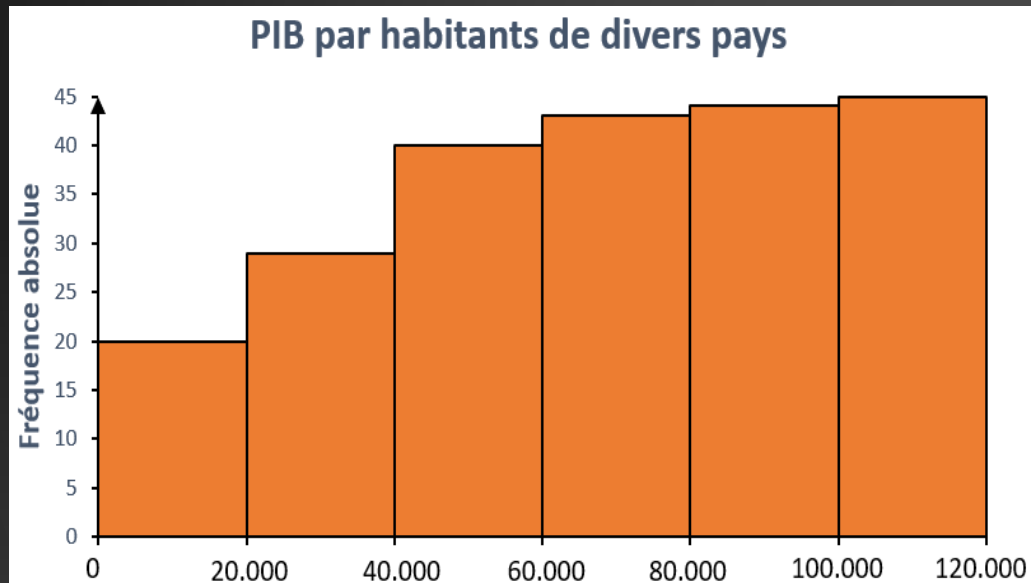
# VARIABLES QUANTITATIVES CONTINUES

## Histogrammes avec polygone des effectifs



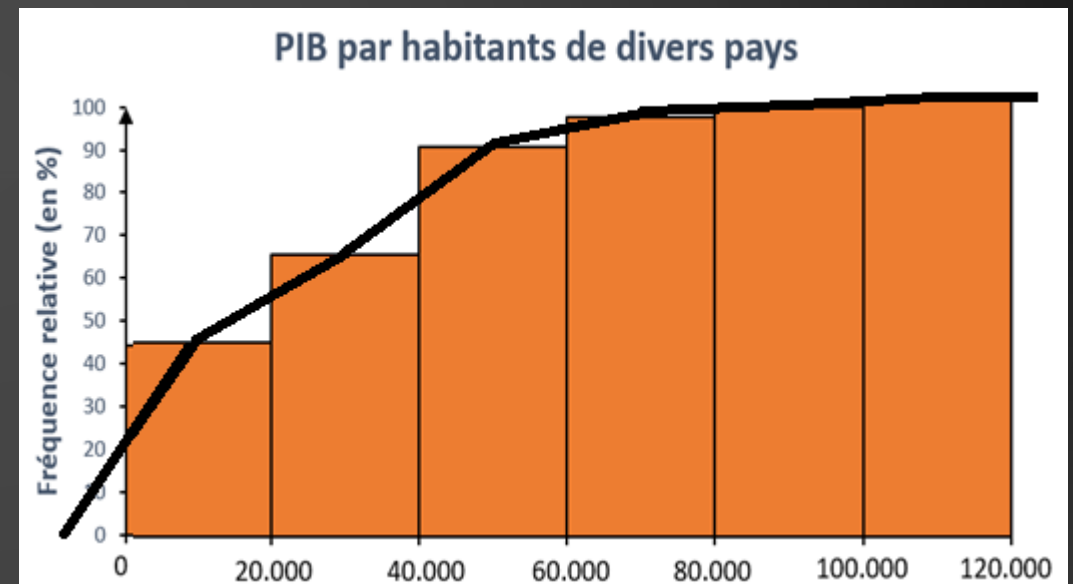
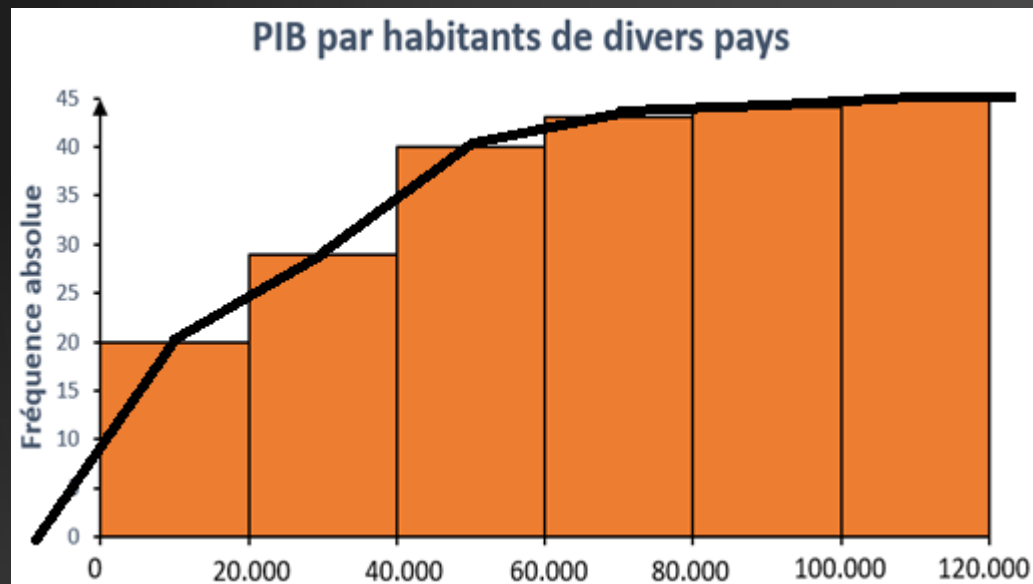
# VARIABLES QUANTITATIVES CONTINUES

## Histogrammes des fréquences cumulées



# VARIABLES QUANTITATIVES CONTINUES

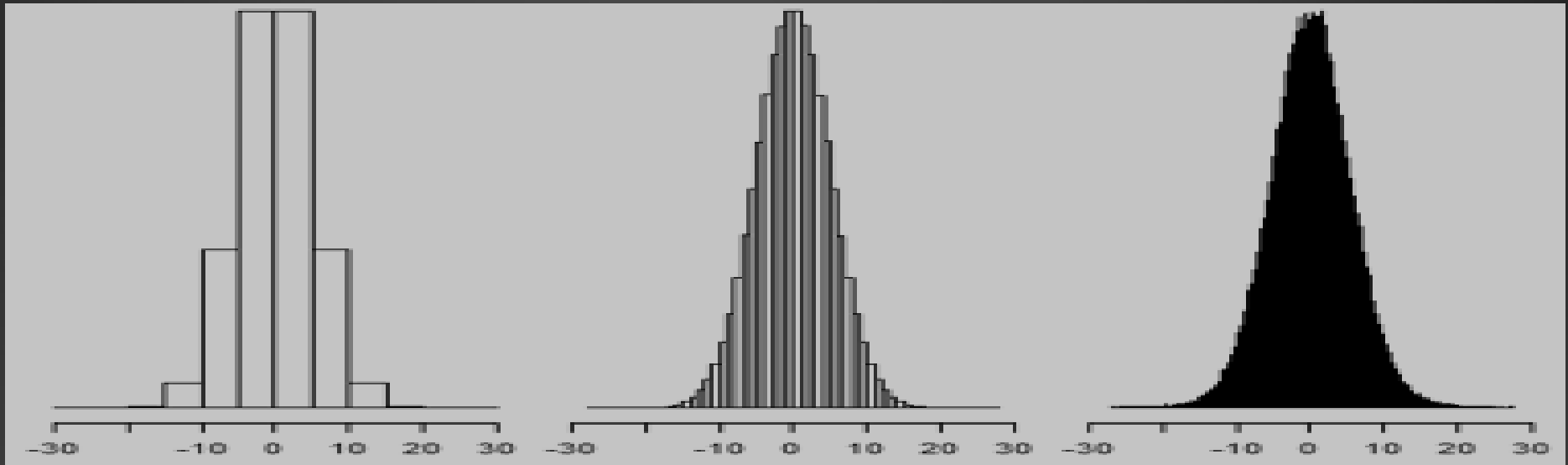
## Histogrammes des fréquences cumulées avec polygone des effectifs cumulés



# LES DISTRIBUTIONS

Imaginez qu'on rétrécisse de plus en plus la largeur d'un intervalle...

... à la fin, la largeur serait si petite qu'elle tendrait vers 0



# LES DISTRIBUTIONS

- Une infinité de distributions possibles
- Quatre types de formes principaux:

