

Séance 2

**Représentation de l'information :
tableaux et graphiques**

Objectifs de la séance

Les données statistiques sont souvent consignées dans des **tableaux** et représentées sous forme de **graphiques**. **Selon la nature des caractères** (quantitatifs ou qualitatifs), **les représentations seront différentes**. Le but de ce chapitre est :

- De présenter les modes de **représentations numériques** des données (tableaux) ;
- De présenter les modes de **représentations graphiques** des données (histogrammes, diagrammes en bâtons, etc.) ;
- D'introduire deux notions fondamentales pour la construction de ces représentations : les **effectifs** et les **fréquences**.

LE TABLEAU STATISTIQUE COMPLET

Individus	Caract. X	Caract. Y	Caract. Z
Ind. 1	1	4	Fort
Ind. 2	2	6	Faible
Ind. 3	2	8	Faible
Ind. 4	3	8	Moyen
Ind. 5	5	10	Fort

Relation population → caractère

Dans un tableau complet, les modalités de la variable statistique X sont notées x_i , i variant de 1 à n (soit $x_1; x_2; \dots; x_i; \dots; x_n$), les x_i peuvent parfois être semblables

LE TABLEAU STATISTIQUE CONDENSE

Modalités Z	Effectif (n_j)
Faible	2
Moyen	1
Fort	2

Relation caractère → population

Dans un tableau condensé, une modalité quelconque du caractère X est notée x_j , j variant de 1 à k (soit $x_1; x_2; \dots; x_j; \dots; x_k$). A chaque modalité x_j correspond un effectif partiel n_j .

Exemple: étude sur les employés d'une entreprise

Salariés	Locomotion	Niveau d'études	Âge	Salaires mensuels nets (€)
Dupond	Voiture	Master	36	2400
Claude	Voiture	Bac	20	1350
Garisson	Voiture	Master	24	1800
Toto	Transport en commun	Doctorat	28	4500
Martin	Voiture	Doctorat	40	4900
Steen	2 roues	Bac	25	1350
Jefferson	Voiture	Licence	25	1600
Douglas	A pied	Licence	20	1500
Bryan	Transport en commun	Master	40	2400
Marteau	Voiture	Licence	20	1500
Pertus	Voiture	Bac	40	2000
Carrière	A pied	Bac	20	1300
Bistouri	Transport en commun	Master	28	1700
Birhut	Voiture	Master	36	1900
Vasquez	A pied	Master	24	1500
Urena	Transport en commun	Master	60	5000
Ndione	Voiture	Licence	40	1820
Pauli	Voiture	Licence	25	1350
Sanchez	2 roues	Master	51	5000
Muller	Voiture	Bac	20	2000
Norma	Voiture	Bac	60	4900

Quelle est la population observée, sa taille, les individus qui la composent, les caractères analysés et leur type ?

Exemple: étude sur les employés d'une entreprise

Salariés	Locomotion	Niveau d'études	Âge	Salaires mensuels nets (€)
Dupond	Voiture	Master	36	2400
Claude	Voiture	Bac	20	1350
Garisson	Voiture	Master	24	1800
Toto	Transport en commun	Doctorat	28	4500
Martin	Voiture	Doctorat	40	4900
Steen	2 roues	Bac	25	1350
Jefferson	Voiture	Licence	25	1600
Douglas	A pied	Licence	20	1500
Bryan	Transport en commun	Master	40	2400
Marteau	Voiture	Licence	20	1500
Pertus	Voiture	Bac	40	2000
Carrière	A pied	Bac	20	1300
Bistouri	Transport en commun	Master	28	1700
Birhut	Voiture	Master	36	1900
Vasquez	A pied	Master	24	1500
Urena	Transport en commun	Master	60	5000
Ndione	Voiture	Licence	40	1820
Pauli	Voiture	Licence	25	1350
Sanchez	2 roues	Master	51	5000
Muller	Voiture	Bac	20	2000
Norma	Voiture	Bac	60	4900

- La population observée est l'ensemble des employés d'une entreprise (N=21)
- Les individus sont des employés
- 4 caractères sont observés : le mode de transport (qualitatif nominal), le niveau d'études (qualitatif ordinal), l'âge (quantitatif discret), le salaire (quantitatif continu)

➤ EFFECTIF de la MODALITÉ

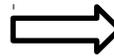
Définition: pour chaque modalité de la variable on compte le nombre d'individus ayant cette modalité

Notation: n_j

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Locomotion
Dupond	Voiture
Claude	Voiture
Garisson	Voiture
Toto	Transport en commun
Martin	Voiture
Steen	2 roues
Jefferson	Voiture
Douglas	A pied
Bryan	Transport en commun
Marteau	Voiture
Pertus	Voiture
Carrière	A pied
Bistouri	Transport en commun
Birhut	Voiture
Vasquez	A pied
Urena	Transport en commun
Ndione	Voiture
Pauli	Voiture
Sanchez	2 roues
Muller	Voiture
Norma	Voiture

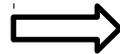


Locomotion	Effectifs (n_j)
Voiture	
Transport en commun	
2 roues	
A pied	
Total	

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Locomotion
Dupond	Voiture
Claude	Voiture
Garisson	Voiture
Toto	Transport en commun
Martin	Voiture
Steen	2 roues
Jefferson	Voiture
Douglas	A pied
Bryan	Transport en commun
Marteau	Voiture
Pertus	Voiture
Carrière	A pied
Bistouri	Transport en commun
Birhut	Voiture
Vasquez	A pied
Urena	Transport en commun
Ndione	Voiture
Pauli	Voiture
Sanchez	2 roues
Muller	Voiture
Norma	Voiture



Locomotion	Effectifs (n_j)
Voiture	12
Transport en commun	4
2 roues	2
A pied	3
Total	21

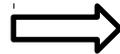


Variable qualitative → modalités sans ordre particulier

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Niveau d'études
Dupond	Master
Claude	Bac
Garisson	Master
Toto	Doctorat
Martin	Doctorat
Steen	Bac
Jefferson	Licence
Douglas	Licence
Bryan	Master
Marteau	Licence
Pertus	Bac
Carrière	Bac
Bistouri	Master
Birhut	Master
Vasquez	Master
Urena	Master
Ndione	Licence
Pauli	Licence
Sanchez	Master
Muller	Bac
Norma	Bac

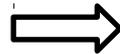


Niveau d'étude	Effectifs (n_j)
Bac	
Licence	
Master	
Doctorat	
Total	

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Niveau d'études
Dupond	Master
Claude	Bac
Garisson	Master
Toto	Doctorat
Martin	Doctorat
Steen	Bac
Jefferson	Licence
Douglas	Licence
Bryan	Master
Marteau	Licence
Pertus	Bac
Carrière	Bac
Bistouri	Master
Birhut	Master
Vasquez	Master
Urena	Master
Ndione	Licence
Pauli	Licence
Sanchez	Master
Muller	Bac
Norma	Bac



Niveau d'étude	Effectifs (n_i)
Bac	6
Licence	5
Master	8
Doctorat	2
Total	21

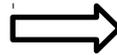


Variable qualitative ordinaire → modalités dans l'ordre croissant ou décroissant

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Âge
Dupond	36
Claude	20
Garisson	24
Toto	28
Martin	40
Steen	25
Jefferson	25
Douglas	20
Bryan	40
Marteau	20
Pertus	40
Carrière	20
Bistouri	28
Birhut	36
Vasquez	24
Urena	60
Ndione	40
Pauli	25
Sanchez	51
Muller	20
Norma	60

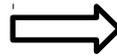


Âge	Effectifs (n_j)
20	
24	
25	
28	
36	
40	
51	
60	
Total	

I. Les représentations numériques

2. Effectifs et fréquences

Salariés	Âge
Dupond	36
Claude	20
Garisson	24
Toto	28
Martin	40
Steen	25
Jefferson	25
Douglas	20
Bryan	40
Marteau	20
Pertus	40
Carrière	20
Bistouri	28
Birhut	36
Vasquez	24
Urena	60
Ndione	40
Pauli	25
Sanchez	51
Muller	20



Âge	Effectifs (n_j)
20	5
24	2
25	3
28	2
36	2
40	4
51	1
60	2
Total	21

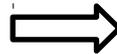


Variable quantitative discrète → modalités dans l'ordre croissant ou décroissant

I. Les représentations numériques

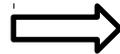
2. Effectifs et fréquences

Salariés	Salaires mensuels nets (€)
Dupond	2400
Claude	1350
Garisson	1800
Toto	4500
Martin	4900
Steen	1350
Jefferson	1600
Douglas	1500
Bryan	2400
Marteau	1500
Pertus	2000
Carrière	1300
Bistouri	1700
Birhut	1900
Vasquez	1500
Urena	5000
Ndione	1820
Pauli	1350
Sanchez	5000
Muller	2000
Norma	4900



Salaire	Effectifs (n_j)
1300	
1350	
1500	
1600	
1700	
1800	
1820	
1900	
2000	
2400	
4500	
4900	
5000	
Total	

Salariés	Salaires mensuels nets (€)
Dupond	2400
Claude	1350
Garisson	1800
Toto	4500
Martin	4900
Steen	1350
Jefferson	1600
Douglas	1500
Bryan	2400
Marteau	1500
Pertus	2000
Carrière	1300
Bistouri	1700
Birhut	1900
Vasquez	1500
Urena	5000
Ndione	1820
Pauli	1350
Sanchez	5000
Muller	2000
Norma	4900

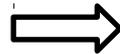


Salaire	Effectifs (n_j)
1300	1
1350	3
1500	3
1600	1
1700	1
1800	1
1820	1
1900	1
2000	2
2400	2
4500	1
4900	2
5000	2
Total	21



Variable quantitative continue → modalités dans l'ordre croissant ou décroissant

Salariés	Salaires mensuels nets (€)
Dupond	2400
Claude	1350
Garisson	1800
Toto	4500
Martin	4900
Steen	1350
Jefferson	1600
Douglas	1500
Bryan	2400
Marteau	1500
Pertus	2000
Carrière	1300
Bistouri	1700
Birhut	1900
Vasquez	1500
Urena	5000
Ndione	1820
Pauli	1350
Sanchez	5000
Muller	2000
Norma	4900



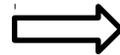
Salaire	Effectifs (n_j)
1300	1
1350	3
1500	3
1600	1
1700	1
1800	1
1820	1
1900	1
2000	2
2400	2
4500	1
4900	2
5000	2
Total	21



Variable quantitative continue → modalités dans l'ordre croissant ou décroissant

Quand on peut ordonner les modalités, il est possible de calculer des **EFFECTIFS CUMULÉS** (N_j)

Salaire	Effectifs (n_j)
1300	1
1350	3
1500	3
1600	1
1700	1
1800	1
1820	1
1900	1
2000	2
2400	2
4500	1
4900	2
5000	2
Total	21



Salaire	Effectifs (n_j)	Effectifs cumulés (N_j)
1300	1	1
1350	3	4
1500	3	7
1600	1	8
1700	1	9
1800	1	10
1820	1	11
1900	1	12
2000	2	14
2400	2	16
4500	1	17
4900	2	19
5000	2	21
Total	21	

Lecture : 7 employés gagnent au plus 1 500 euros

➤ **FRÉQUENCE SIMPLE**

Définition: proportion d'individus pour lesquels une variable statistique a pris une valeur donnée

Notation: f_j

Calcul:

$$f_j = \frac{n_j}{n}$$

➤ FRÉQUENCE SIMPLE

Définition: proportion d'individus pour lesquels une variable statistique a pris une valeur donnée

Notation: f

Calcul: $f_j = \frac{n_j}{n}$ $\Rightarrow \sum_j^k f_j = 1$

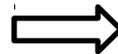


[0;1] ou [0%;100%]

Se lit : la somme des f_j , j variant de 1 à k est égale à 1

➤ FRÉQUENCE SIMPLE

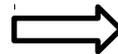
Niveau d'étude	Effectifs n_j
Bac	6
Licence	5
Master	8
Doctorat	2
N	21



Niveau d'étude	n_j	f_j
Bac	6	
Licence	5	
Master	8	
Doctorat	2	
N	21	

➤ FRÉQUENCE SIMPLE

Niveau d'étude	Effectifs n_j
Bac	6
Licence	5
Master	8
Doctorat	2
Total	21



Niveau d'étude	n_j	f_j
Bac	6	0.29 ou 29 %
Licence	5	0.24 ou 24 %
Master	8	0.38 ou 38 %
Doctorat	2	0.09 ou 9 %
Total	21	1 ou 100 %

➤ FRÉQUENCE CUMULÉE

Définition: Résultat de l'addition, de proche en proche, des fréquences d'une distribution observée

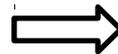
Notation: F



IL FAUT D'ABORD AVOIR CLASSÉ LES MODALITÉS

➤ FRÉQUENCE CUMULÉE

Niveau d'étude	n_j	f_j
Bac	6	0.29
Licence	5	0.24
Master	8	0.38
Doctorat	2	0.09
Total	21	1

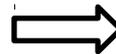


Niveau d'étude	n_j	f_j	F_j croissante
Bac	6	0.29	0.29
Licence	5	0.24	0.53
Master	8	0.38	0.91
Doctorat	2	0.09	1
Total	21	1	

Lecture : 53 % des employés ont au plus une Licence

➤ FRÉQUENCE CUMULÉE

Niveau d'étude	n_j	f_j
Bac	6	0.29
Licence	5	0.24
Master	8	0.38
Doctorat	2	0.09
Total	21	1



Niveau d'étude	n_j	f_j	F_j décroissante
Bac	6	0.29	1
Licence	5	0.24	0.71
Master	8	0.38	0.47
Doctorat	2	0.09	0.09
Total	21	1	

Lecture : 71% des employés ont au moins une Licence

Si le caractère auquel on s'intéresse est **quantitatif continu** ou **quantitatif discret avec de nombreuses valeurs**, il est nécessaire pour passer du tableau complet au tableau condensé, de regrouper les valeurs en classes, de **discrétiser la variable**.

➤ CLASSE

Définition : intervalle de valeurs d'une variable continue, l'ensemble des classes formant une partition de l'ensemble des valeurs possibles de la variable.

Notation : [borne inférieure – borne supérieure[

➤ CLASSE

Propriétés:

- ✓ **Bornes** d'une classe = valeurs minimale et maximale de la classe
- ✓ **Amplitude** d'une classe = **longueur de l'intervalle** :
Elle n'est pas toujours la même selon les classes $a = b_{\text{sup}} - b_{\text{inf}}$
- ✓ **Centre** d'une classe $c = \frac{b_{\text{inf}} + b_{\text{sup}}}{2}$

➤ CLASSE

Construction : la **discrétisation** d'une variable passe par deux opérations qui supposent deux choix qu'ils convient d'objectiver :

- ✓ nombre de classes ;
- ✓ méthode pour fixer les bornes des classes.

➤ CLASSE

Le choix du nombre de classes :

- ✓ Il faut trouver un compromis entre un découpage trop fin (beaucoup de classes) qui impliquerait la présence de classes avec de trop faibles effectifs et un découpage grossier (peu de classes) qui entraînerait une perte importante d'information
- ✓ Ce choix est conditionné par les **possibilités de représentation cartographique** (pas plus de 9 classes), **l'effectif total** de la population et **l'allure de la variable**.

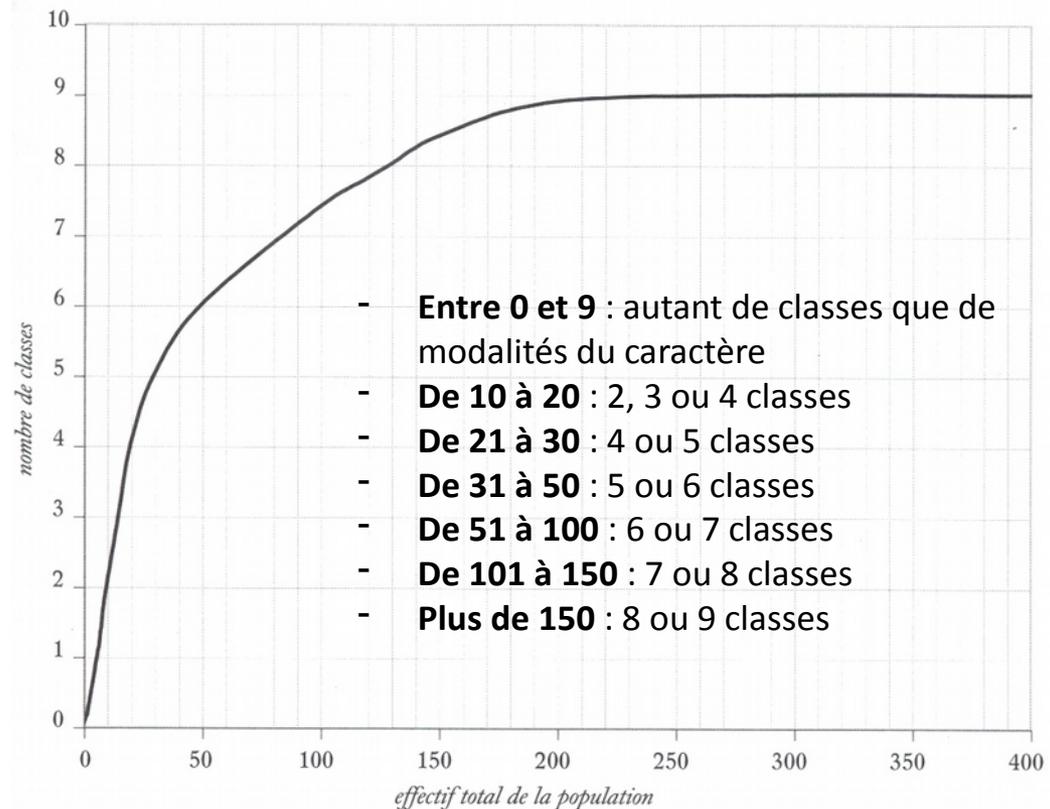
➤ CLASSE

Le choix du nombre de classe :

Formule de Brooks-Carruthers : $k = 5 \times \log(N)$

Formule de Sturges-Huntsberger : $k = 1 + (10/3) \times \log(N)$

Nombre de classes
en fonction
de l'effectif
de la population



D'après Dumolard et al.
à partir de Huntsberger et Brooks

➤ CLASSE

Le choix des bornes de classes:

- Les limites des classes doivent couvrir l'ensemble du domaine de variation de la variable : des valeurs ne doivent pas être laissées en dehors du champ couvert par les classes
- Les bornes des classes doivent être des valeurs simples (éviter 0,555890093 comme borne de classe, prendre plutôt 0,5)
- Si certaines valeurs ont une signification dans le domaine d'étude alors il faut les envisager comme borne de classes (ex : si l'on étudie la variation de la population des communes entre deux dates, il faut que la borne 0 apparaisse)
- Les bornes doivent respecter l'allure de la variable que l'on peut par exemple visualiser par **un scalogramme**

Bornes de classes

Valeurs du caractère



➤ CLASSE

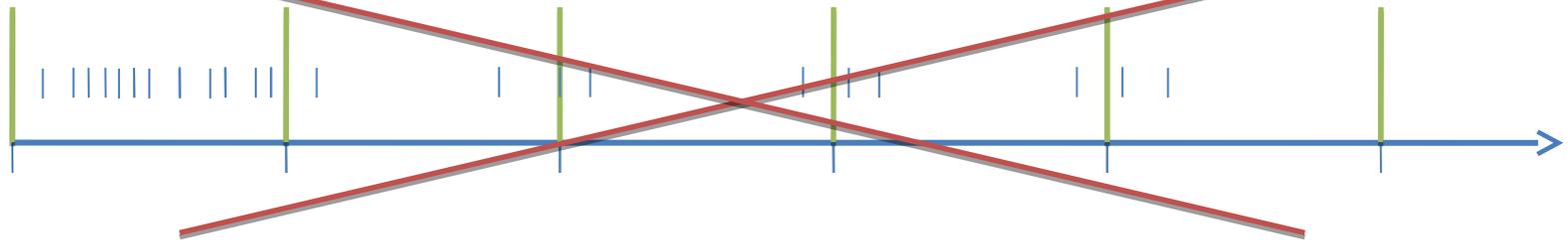
Le choix des bornes de classes:

Quand la distribution n'offre pas de seuils nets, et se répartit de façon homogène, il est commode de réaliser **des classes d'égale amplitude**. Il faut éviter cette méthode quand la distribution est dissymétrique (beaucoup de faibles ou des grandes valeurs)

Répartition homogène des individus : classes d'égale amplitude



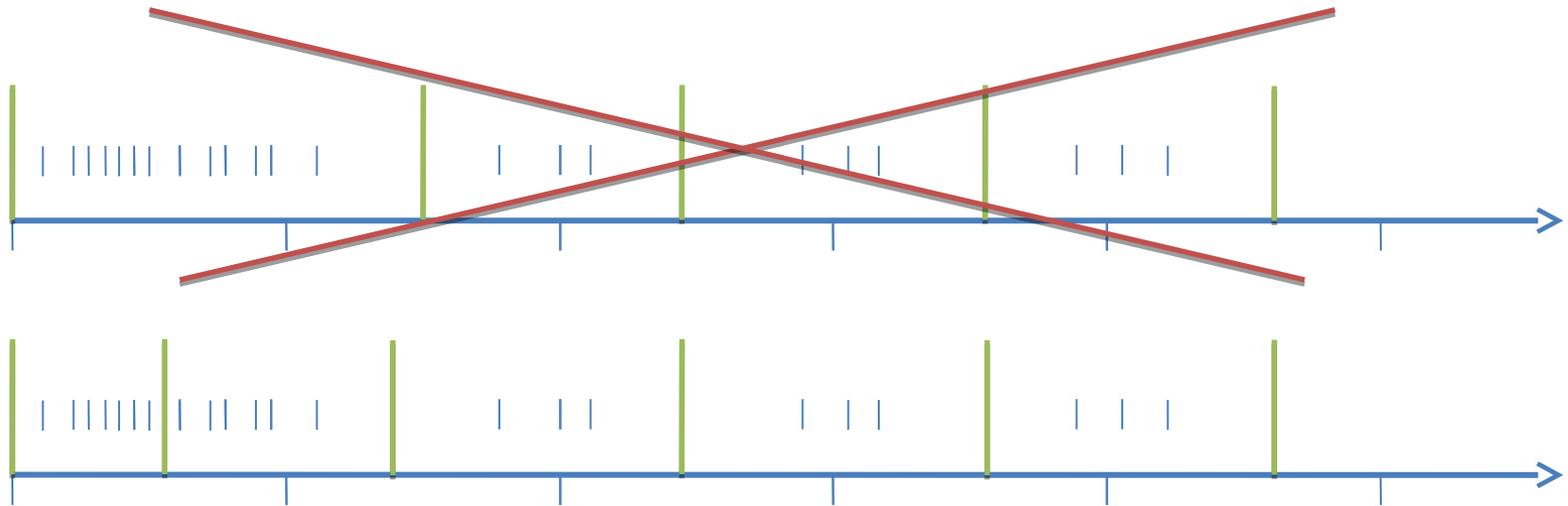
Variables dissymétriques : chercher des discontinuités



➤ CLASSE

Le choix des bornes de classes:

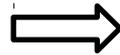
- Quand une classe a une large amplitude et une forte fréquence par rapport aux autres, il faut la fractionner



- Si l'on veut comparer un même caractère sur plusieurs populations, il faut procéder à un découpage commun de toutes les distributions (ex : densités de population des communes françaises à différentes dates)

I. Les représentations numériques 3. Découpage en classes (discrétisation)

Âge	Effectifs
20	5
24	2
25	3
28	2
36	2
40	4
51	1
60	2
Total	21

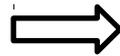


Âge	Effectifs
[20; 30[12
[30; 40[2
[40; 50[4
[50;60]	3
Total	21

Classes d'égale amplitude (a = 10 ans)

I. Les représentations numériques 3. Découpage en classes (discrétisation)

Salaire	Effectifs
1300	1
1350	3
1500	3
1600	1
1700	1
1800	1
1820	1
1900	1
2000	2
2400	2
4500	1
4900	2
5000	2
Total	21



Salaire	Effectifs
[1300 ; 1600[7
[1600 ; 1900[4
[1900; 2400[5
[2400; 5000]	5
Total	21

Classes d'inégale amplitude

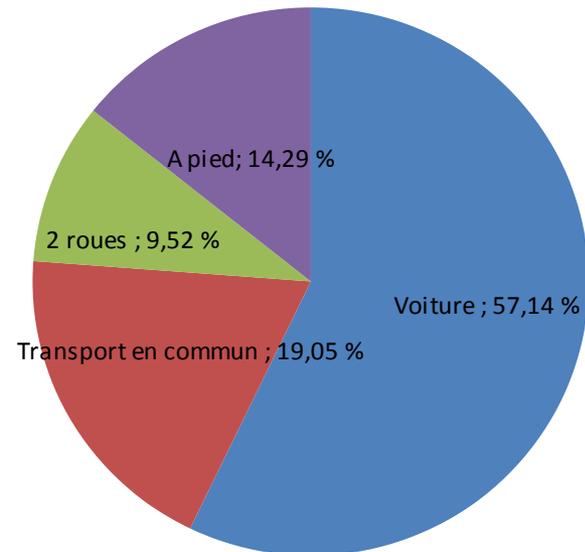
➤ DIAGRAMME CIRCULAIRE

Modalités représentées par des portions de disque proportionnelles à leur effectif/fréquence.

Locomotion	n_j	f_j
Voiture	12	0.57
Transport en commun	4	0.19
2 roues	2	0.09
A pied	3	0.14
Total	21	1

$$a_{(\text{voiture})} = 0.57 * 360 = 205^\circ$$

Répartition des salariés selon le type de transport



➤ DIAGRAMME EN BARRES ou EN BÂTONS

Modalités placées en abscisses, formant des bases de rectangles égales, et les effectifs (ou fréquences) en ordonnée

Locomotion	n_j	f_j
Voiture	12	0.57
Transport en commun	4	0.19
2 roues	2	0.09
A pied	3	0.14
Total	21	1

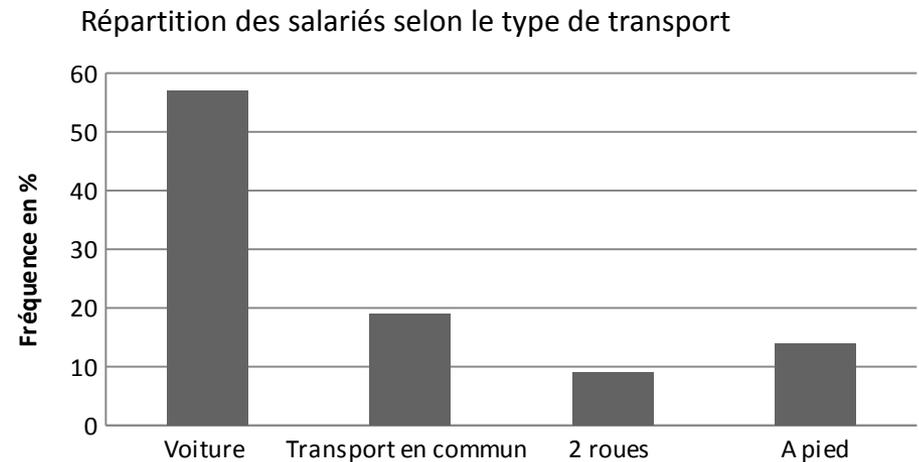


DIAGRAMME EN BARRES ou EN BÂTONS DES EFFECTIFS CUMULÉS

Si la variable peut être ordonnée, on peut représenter en plus des graphiques présentés précédemment les effectifs cumulés sur un diagramme en barres

Niveau d'étude	n_j	N_j
Bac	6	6
Licence	5	11
Master	8	19
Doctorat	2	21



Variable qualitative ordinale : toujours placer les modalités dans l'ordre

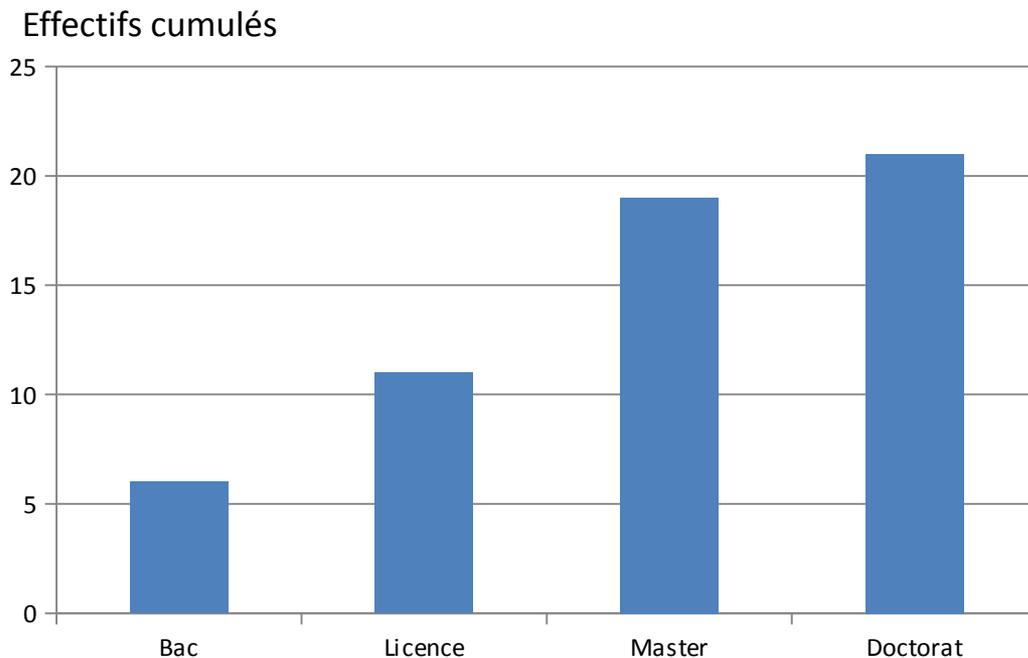
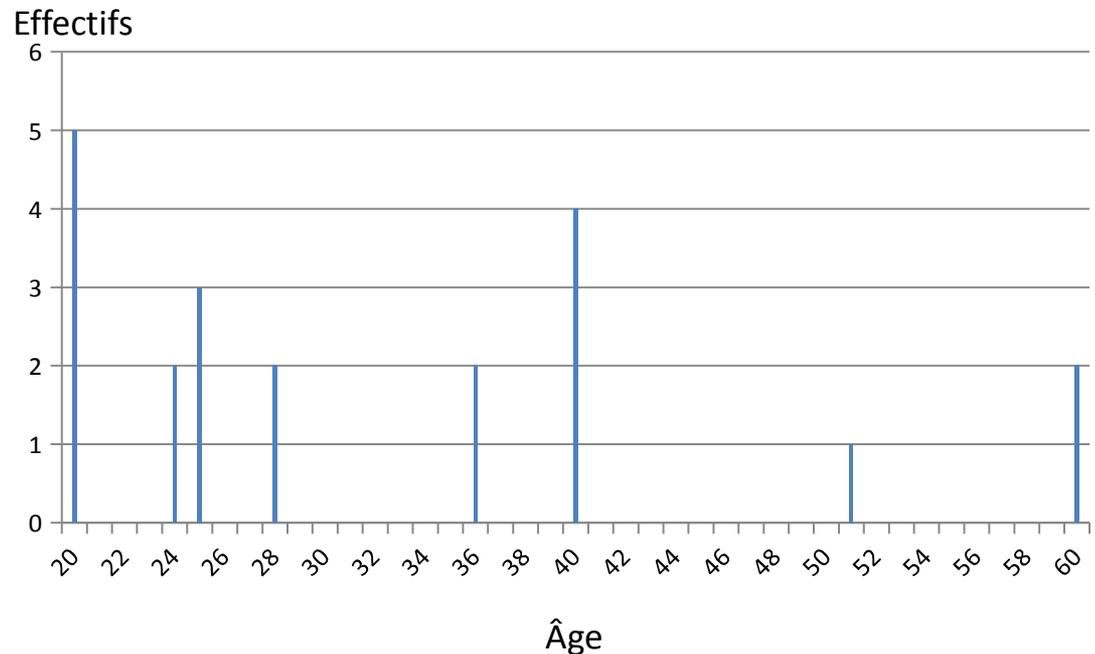


DIAGRAMME EN BARRES ou EN BÂTONS

Modalités placées en abscisses, les effectifs (ou fréquences) en ordonnée, au moyen de segments verticaux.

Répartition des salariés en fonction de l'âge

Âge	n_j	f_j
20	5	0,24
24	2	0,09
25	3	0,14
28	2	0,09
36	2	0,09
40	4	0,19
51	1	0,05
60	2	0,09
Total	21	1

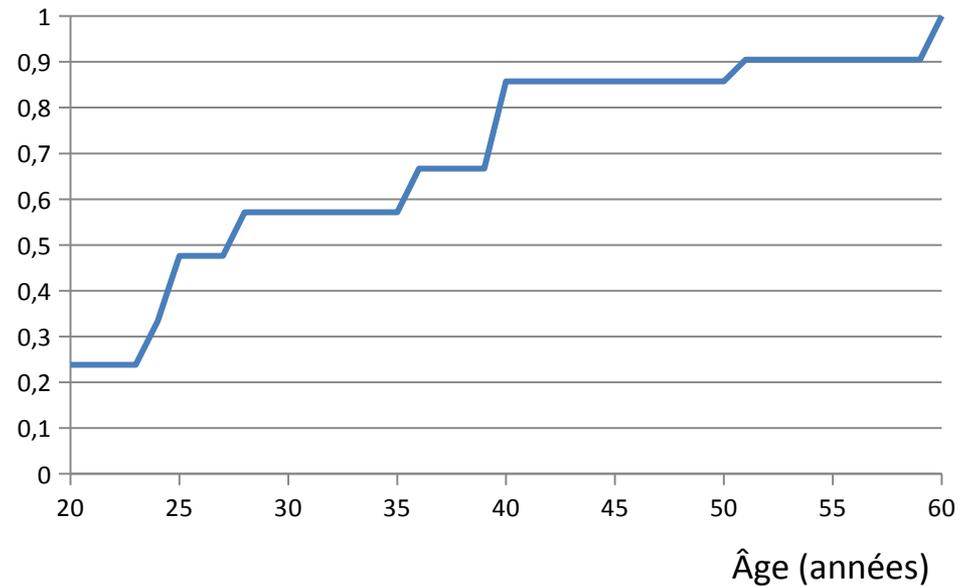


➤ COURBE CUMULATIVE (COURBE EN ESCALIER)

S'obtient au moyen des fréquences cumulées croissantes ou décroissantes

Âge	n_j	f_j	F_j croissantes
20	5	0,24	0.24
24	2	0,09	0.33
25	3	0,14	0.47
28	2	0,09	0.56
36	2	0,09	0.65
40	4	0,19	0.84
51	1	0,05	0.89
60	2	0,09	0.98 = 1
Total	21	1	

F_j croissantes



➤ HISTOGRAMME

Rectangles juxtaposés dont les bases sont les classes, et les surfaces sont proportionnelles aux effectifs (ou fréquences) associés.

 **2 CAS** 

- ✓ Classes d'égale amplitude: rectangles dont la surface est proportionnelle à l'effectif/fréquence associé
- ✓ Classes d'inégale amplitude : calcul de densité d'effectifs ou de fréquences


$$d_j = \frac{n_j}{a_j}$$


$$df_j = \frac{f_j}{a_j}$$

➤ HISTOGRAMME

Rectangles juxtaposés dont les bases sont les classes, et les surfaces sont proportionnelles aux effectifs (ou fréquences) associés.

⚠ 2 CAS ⚠

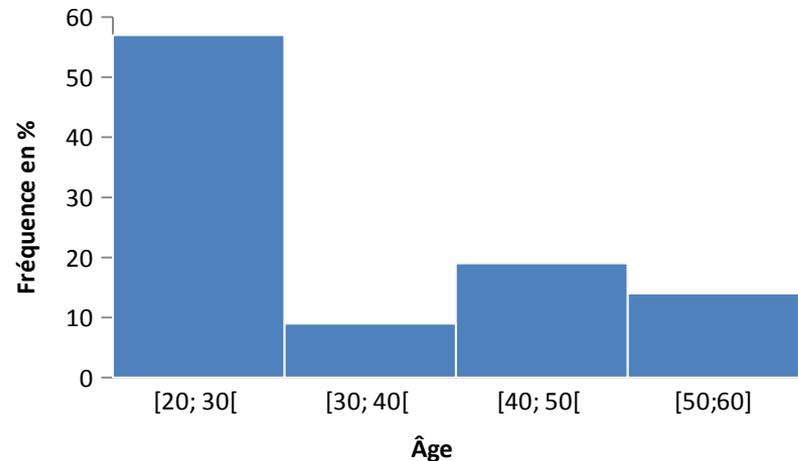
- ✓ Classes d'égale amplitude: rectangles dont la surface est proportionnelle à l'effectif/fréquence associé
- ✓ Classes d'inégale amplitude : calcul de densité d'effectifs ou de fréquences


$$d_j = \frac{n_j}{a_j}$$

$$df_j = \frac{f_j}{a_j}$$

➤ HISTOGRAMME : CLASSES D'ÉGALE AMPLITUDE

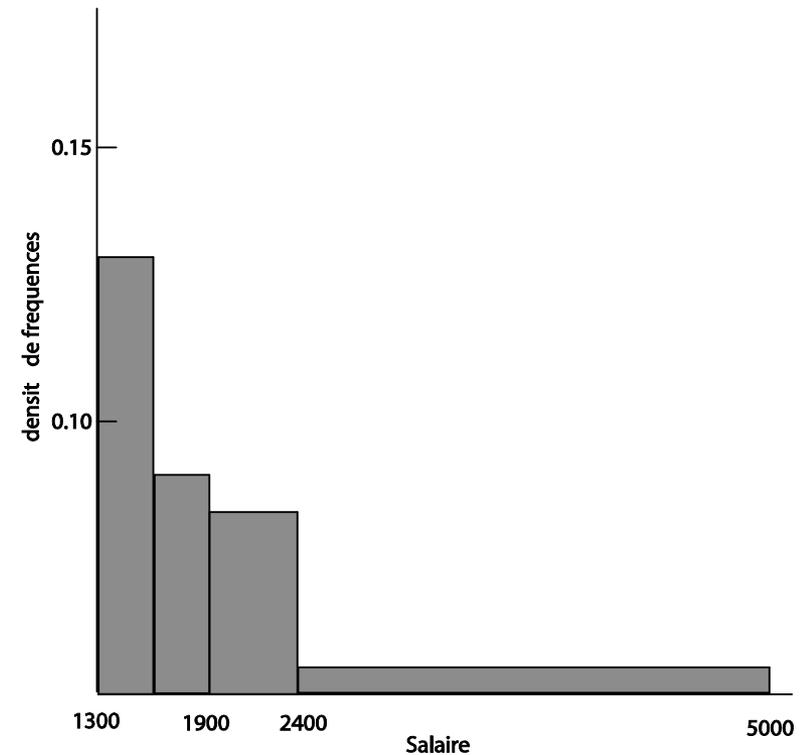
Âge	n_j	f_j
[20 - 30[12	0.57
[30 - 40[2	0.09
[40 - 50[4	0.19
[50 - 60]	3	0.14
Total	21	1



On se contente de reporter sur l'axe des ordonnées la fréquence des modalités. La largeur des « barres » sur l'axe des abscisses est identique car les classes sont de même amplitude.

➤ HISTOGRAMME : CLASSES D'INÉGALE AMPLITUDE

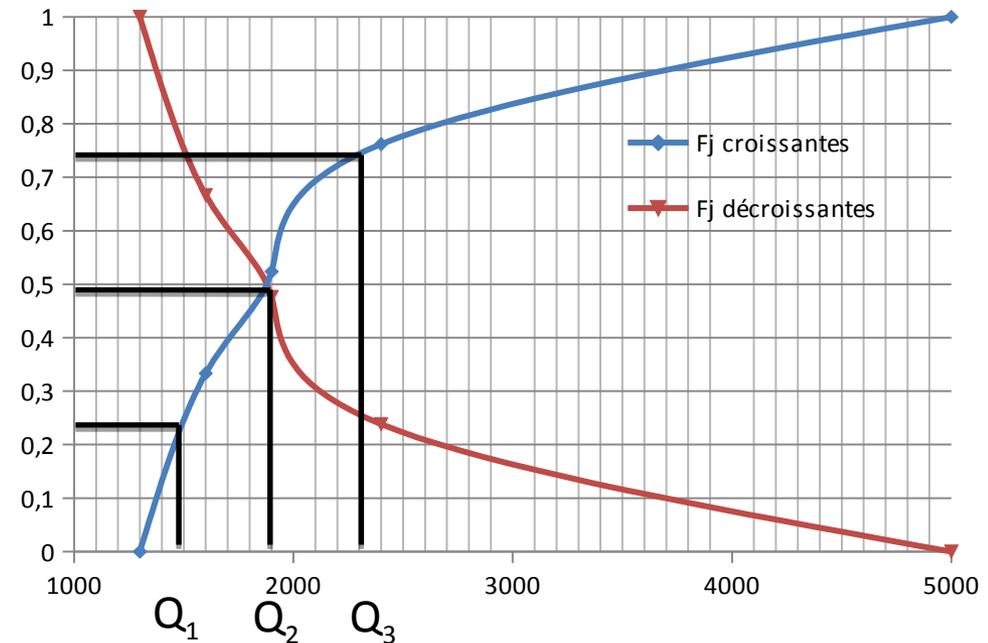
Salaire	Effectifs (n_j)	Amplitude (a_j)	Fréquence f_j	Densité d'effectifs (d_j) en %	Densité de fréquences (df_j) en %
[1300 ; 1600[7	300	0.33	2,3	0.12
[1600 ; 1900[4	300	0.19	1,3	0.06
[1900; 2400[5	500	0.24	1	0.05
[2400; 5000]	5	2600	0.24	0,19	0.009
Total	21		1		



La largeur des « barres » est variable (amplitude des classes) et la hauteur correspond à la densité d'effectifs ou de fréquences

➤ COURBE DES FRÉQUENCES CUMULÉES CROISSANTES ET DÉCROISSANTES

Salaire	Effectifs (n_i)	Fréquence f_j	Fréquences cumulées croissantes	Fréquences cumulées décroissantes
[1300 ; 1600[7	0,33	0,33	1
[1600 ; 1900[4	0,19	0,52	0,67
[1900 ; 2400[5	0,24	0,76	0,48
[2400 ; 5000]	5	0,24	1	0,24
Total	21	1		



1^{er} quartile (Q_1) : 25 % des salariés gagnent strictement moins de 1500 euros

2^{ème} quartile (Q_2) ou médiane : 50 % des salariés gagnent strictement moins de 1900 euros

3^{ème} quartile (Q_3) ou médiane : 75 % des salariés gagnent strictement moins de 2300 euros

➤ HISTOGRAMME : CLASSES D'ÉGAL EFFECTIF

Le critère visé est l'équipartition, c'est à dire le même nombre de données par classe.

A partir du nombre de l'effectif total (N) et du nombre de classes (n), on en déduit le nombre F d'individus par classe.

Pour cela on classe les données par ordre croissant et on met dans la classe 1 les F premières données, dans la classe 2 les F suivantes etc.

Dénomination des quantiles selon la valeur de n :

4 classes : quartiles

5 classes : quintiles

6 classes : sextiles

10 classes : déciles

100 classes : centiles