

Initiation à la statistique

Département Paramédical Bachelier Infirmier responsable de soins généraux

(UE 45)

Omar A. Belaredj Ph.D.

omar.belaredj@he-ferrer.eu

Année académique 2021-2022

Ouvrages de référence

Luc ALBARELLO • Étienne BOURGEOIS • Jean-Luc GUYOT

STATISTIQUE DESCRIPTIVE

Un outil pour les
praticiens-chercheurs

2^e édition

Préface de Jean-Marie De Ketele

MÉTHODES EN SCIENCES HUMAINES



de boeck

STATISTIQUE POUR LES INFIRMIÈRES

T. ANCELLE • M. ROTHAN-TONDEUR

MALOINE

Objectifs

- Ce cours constitue une introduction à la statistique, destinée à des Bacheliers Infirmiers responsables de soins généraux sans formation scientifique particulière. Son but principal est de donner les éléments de base permettant de comprendre, d'interpréter et de critiquer les informations à caractère statistique dans le domaine de la santé. L'accent est mis sur l'utilisation pratique et non sur la rigueur mathématique. L'emploi de formules est réduit au maximum et celles-ci sont simplifiées autant que possible sans affecter la qualité des résultats.
- Initier le praticien-chercheur aux méthodes et techniques permettant de réaliser les opérations qui relèvent de la statistique descriptive - dénombrement, classement, synthèse et présentation des données, que celles-ci concernent une population finie ou un échantillon.

Exemples : articles de presse

<https://www.lesoir.be/413561/article/2021-12-21/coronavirus-tous-les-chiffres-sont-en-baisse-infographie>

<https://datastudio.google.com/embed/reporting/c14a5cfc-cab7-4812-848c-0369173148ab/page/ZwmOB>

<https://www.lejdd.fr/Societe/exclusif-top-50-du-jdd-goldman-et-marceau-restant-les-personnalites-preferees-des-francais-pesquet-decolle-4084604>

Qu'est-ce que la statistique?

- ❑ Vient du latin status = « état ». Le terme statisticum apparaît à la fin du XVII^e siècle.
- ❑ Statistique = ensemble de méthodes permettant de décrire et d'analyser des observations (ou données). Ces observations consistent généralement en la mesure d'une ou plusieurs caractéristiques communes sur un ensemble de personnes ou d'objets équivalents.
- ❑ Remarque : une statistique = grandeur calculée à partir des observations recueillies (ex : moyenne d'âge des élèves d'une même classe, balance commerciale de la France, etc..)

Les statistiques sont aujourd'hui utilisées dans tous les secteurs d'activité :

- Industrie : fiabilité, contrôle qualité,
- Economie et finance: sondages, enquête d'opinion, assurance, marketing
- Santé, environnement,...
- Partout où l'on dispose de données
- ont connu un grand essor avec l'arrivage des ordinateurs performants

Statistique descriptive et statistique inférentielle

- La statistique se divise classiquement en deux parties :
 - La statistique **descriptive**
 - La statistique **inférentielle** : ensemble des méthodes et des théories permettant de généraliser à une population de référence des conclusions obtenues à partir de l'étude d'un échantillon extrait de cette population

Statistique descriptive et statistique inférentielle

La statistique descriptive

Operations:

- de **dénombrement**,
- de **classement/rangement** (en catégories),
- de **synthèse** (en pourcentages), et
- de **présentation des données** (sous forme de tableaux de données et de graphiques)

Repartition de la population du Royaume de Belgique au premier janvier 2005, selon la nationalité : chiffres bruts et pourcentages (Source : I.N.S, 2006)		
Belges	Etrangers	Population totale
9574990 91,7%	870862 8,3%	10445852 100,0%

Belges		Etrangers	
9574990	91,7%	870862	8,3%

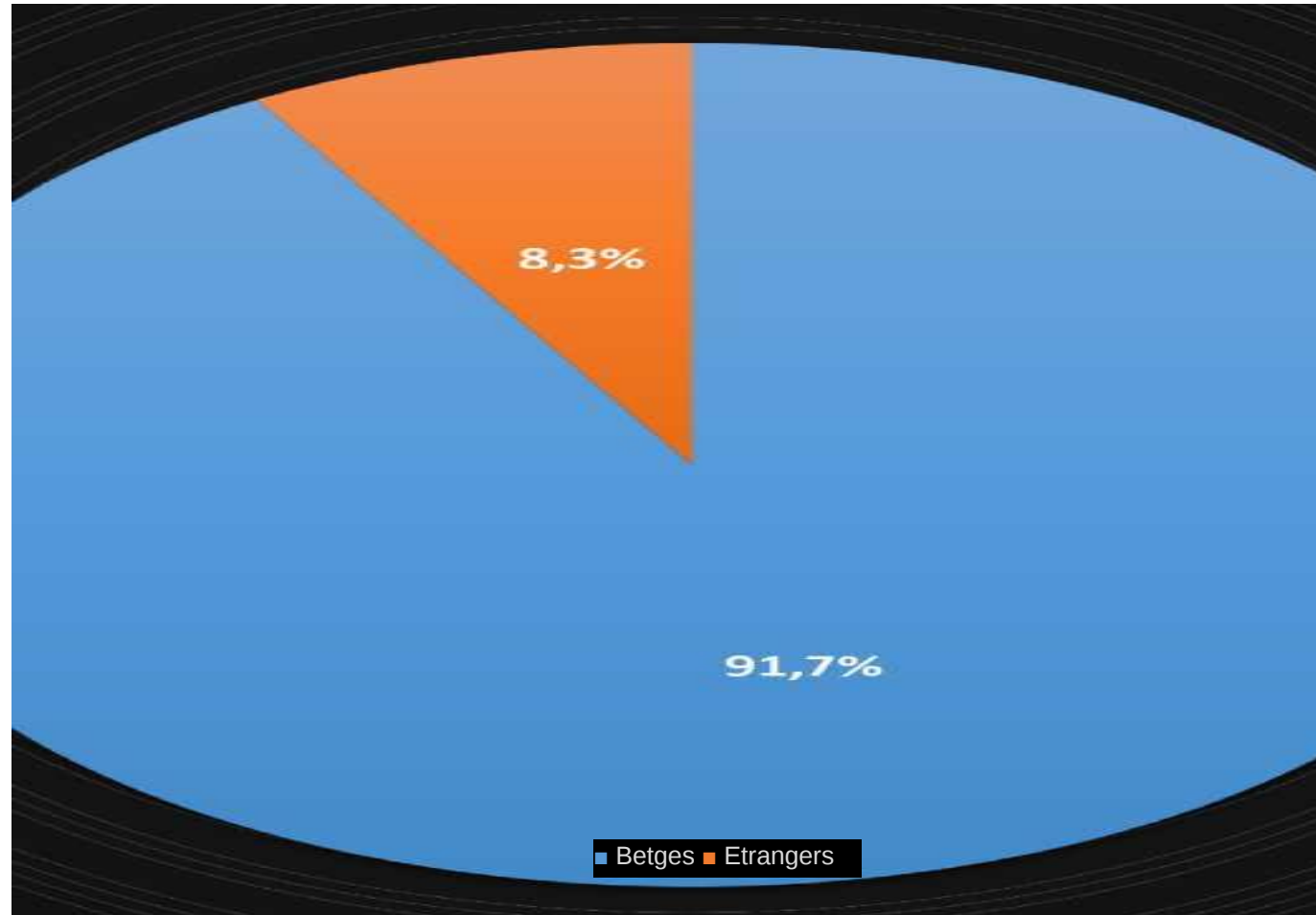
Belges	Etrangers
91,7%	8,3%

Belges	Etrangers
91,7%	8,3%

Répartition de la population du Royaume de Belgique au premier janvier 2005

Distribution par nationalité

Source : INS 2006



Exercice 1

Parmi les énonces suivants, quels sont les objets de recherche qui relèvent de la statistique descriptive et quels sont ceux qui relèvent de la statistique inférentielle ?

Etablir les proportions de chômeurs complets indemnisés de moins de 25 ans selon le genre. Situation au 31 décembre 2004 en Belgique. Source : Office National de l'Emploi.

Statistique descriptive. L'ensemble des chômeurs complets indemnisés est étudié dans sa totalité : les données disponibles concernent l'ensemble de cette population et celle-ci est dénombrable de façon exhaustive

Exercice 2

Parmi l'énoncé suivant, quels sont les objets de recherche qui relèvent de la statistique descriptive et quels sont ceux qui relèvent de la statistique inférentielle ?

Vérifier l'hypothèse que les jeunes garçons ont une attitude vis-à-vis de la politique différente de celle des jeunes filles dans les pays de l'UE.

Statistique inférentielle. Vu la taille de la population de référence (l'ensemble des jeunes hommes et jeunes filles de l'UE), elle n'est **pas exhaustivement dénombrable** et on aura donc très vraisemblablement recours à un échantillon.

Exercice 3

Parmi l'énoncé suivant, quels sont les objets de recherche qui relèvent de la statistique descriptive et quels sont ceux qui relèvent de la statistique inférentielle ?

Etudier la répartition de toutes les personnes de plus de 18 ans suivant un enseignement de plein exercice selon le niveau d'enseignement en Belgique en 2005.

Statistique descriptive. C'est la totalité de la population de référence qui est étudiée. Celle-ci est finie, il est possible de comptabiliser l'ensemble des individus constituant cette population.

Chapitre 1

Le triptyque de la statistique descriptive

Le triptyque de la statistique descriptive

□ **Population** statistique

□ **Unité** statistique (ou individu)

□ **Observation** (ou caractère ou variable)

Le triptyque de la statistique descriptive

□ **Population statistique** = univers de référence lors de l'étude d'un problème statistique.

Ensemble constitue de :

- Citoyens, patients...
- Objets (matériels ou immatériels)
- Actions ou de situations
- Animaux
- Etc...

Le triptyque de la statistique descriptive

□ **Unité statistique (ou individu):** un élément d'un ensemble (population) dont on mesure (ou observe) la valeur qu'il a pour la variable étudiée.

- Citoyens d'un pays (unité statistique = citoyen)
- Patients dans le service pneumologie (unité statistique = patient)
- Ensemble des prix d'articles de consommation (unité statistique = prix)

Le triptyque de la statistique descriptive

□ Population

□ Individu

□ **Echantillon**

Sous-ensemble d'une population sur lequel on effectue une étude statistique.

Son étude vise généralement à tirer des conclusions relatives à la population dont il est issu.

Exemples :

- Fraction de la population d'un pays (1000 personnes / 60 millions)
- Prélèvement d'eau (éprouvette) dans un bassin
- Etc...

Le triptyque de la statistique descriptive

❑ Population

❑ Individu

❑ **Echantillon représentatif**

- S'il possède les mêmes caractéristiques que la population que l'on souhaite étudier.
- Cette représentativité doit surtout se faire sur les caractéristiques pouvant influencer les réponses.

•Faute de représentativité, les résultats obtenus sur un échantillon ne peuvent être généralisés à la population étudiée. l'échantillon est biaisé.

Exemple:

Pour me faire une idée du niveau de la 3^{ème} BIRSG, je prends les étudiants des 2 premières rangées et je les questionne.

Est-ce un échantillon représentatif ?

Le triptyque de la statistique descriptive

□ Observation (ou caractère ou variable) :

- Caractéristique qui est observée sur chaque individu constituant l'ensemble
- Caractéristique qui varie au sein de la population
- **Modalités** de la variable:

Exemples:

homme vs femme pour la variable «sexe»

Insatisfait vs satisfait vs très satisfait pour la variable «assouplissement COVID»

« Il convient que le praticien-chercheur adopte systématiquement le reflexe de distinguer explicitement, dans un ensemble de données numériques ou dans un tableau »

- (1) ce qui se réfère à la population,
- (2) qui ou quels sont les individus qui composent celle-ci,
- (3) est ce que l'échantillon est représentative, et
- (4) quelle est ou quelles sont les observations(s) faite(s) sur chacun de ces individus

Dans chacun des **énoncés** ci-dessous se rapportant à l'étude qui suit, identifiez la **population de référence**, les **individus**, la ou les **variables** (observations) considérées.

L'étude:

Dans le cadre d'une recherche européenne sur l'accès des adduites à l'enseignement supérieur, une équipe de chercheurs s'est penchée sur la composition de la population étudiante d'une université. L'étude est basée sur le fichier des étudiants inscrits dans les différentes facultés de cette université pour une année académique donnée. Voici quelques-unes des hypothèses qui ont été formulées à ce propos.

Enoncé 1

La proportion d'étudiants « adultes » (c'est-à-dire de 25 ans ou plus) par rapport aux « jeunes » (c'est-à-dire de moins de 25 ans) est plus élevée dans la faculté de psychologie et des sciences de l'éducation que dans la faculté de sciences appliquées.

Enoncé 1

La proportion d'étudiants « adultes » (c'est-à-dire de 25 ans ou plus) par rapport aux « jeunes » (c'est-à-dire de moins de 25 ans) est plus élevée dans la faculté de psychologie et des sciences de l'éducation que dans la faculté de sciences appliquées.

Population : l'ensemble des étudiants de cette université (tous âges confondus) inscrits dans les programmes d'études soit de la faculté de psychologie et des sciences de l'éducation soit de la faculté de sciences appliquées au cours de l'année académique considérée,

Individu : chaque étudiant apparaissant dans le fichier étudie comme ayant pris son inscription principale dans un de ces programmes d'étude d'une de ces facultés dans cette université cette année-là (en cas d'inscription d'un individu dans plusieurs programmes d'études au cours de la même année, seule l'inscription principale est prise en compte).

Variables : 1) l'âge (révolu) des étudiants (moins de 25 ans, ou 25 ans ou plus) ;
2) la faculté d'inscription (psychologie et sciences de l'éducation, ou sciences appliquées).

Enoncé 2

On trouve d'avantage d'étudiants adultes dans les facultés de sciences humaines que dans celles de sciences (comprenant notamment la faculté de sciences appliquées).

Enoncé 2

On trouve d'avantage d'étudiants adultes dans les facultés de sciences humaines que dans celles de sciences (comprenant notamment la faculté de sciences appliquées).

Population : l'ensemble des étudiants adultes (25 ans ou plus) inscrits dans les programmes d'études considérés de toutes les facultés de sciences humaines et de sciences, au cours de l'année académique considérée.

Individu : chaque étudiant de 25 ans ou plus apparaissant dans le fichier étudie comme ayant pris son inscription principale dans un de ces programmes d'étude d'une de ces facultés dans cette université cette année-là,

Variables : une seule est pris en compte ici : la faculté d'inscription (faculté de sciences humaines ou faculté de sciences)

Enoncé 3

La proportion d'étudiants adultes (par rapport aux jeunes) varie d'une faculté à une autre au sein de cette université.

Enoncé 3

La proportion d'étudiants adultes (par rapport aux jeunes) varie d'une faculté à une autre au sein de cette université.

Population : l'ensemble des étudiants (tous âges confondus) inscrits dans les programmes d'études considérés de l'ensemble des facultés de cette université au cours de l'année académique considérée

Individu : chaque étudiant apparaissant dans le fichier étudie comme inscrit dans un des programmes d'étude considérés de cette université cette année-là (seule l'inscription principale est prise en compte)

Variables : 1) l'Age des étudiants (moins de 25 ans, ou 25 ans ou plus) et
2) la faculté d'inscription (une des 11 facultés de cette université)

Enoncé 4

L'âge moyen et le genre des étudiants varient d'une faculté à l'autre au sein de cette université

Enoncé 4

L'âge moyen et le genre des étudiants varient d'une faculté à l'autre au sein de cette université

Population : l'ensemble des étudiants (tous âges confondus) inscrits dans les programmes d'études considérés de l'ensemble des facultés de cette université au cours de l'année académique considérée.

Individu : chaque étudiant apparaissant dans le fichier étudie comme inscrit dans un des programmes d'étude considérés de cette université cette année-là (seule l'inscription principale est prise en compte)

Variables : 1) l'âge des étudiants (valeur réelle),
2) le genre (homme ou femme) et
3) la faculté d'inscription (une des 11 facultés de cette université)

Enoncé 5

Les étudiants adultes ayant une charge de famille sont proportionnellement plus nombreux (par rapport à ceux qui n'en ont pas) dans les programmes à horaire décalé que dans les programmes de jour.

Enoncé 5

Les étudiants adultes ayant une charge de famille sont proportionnellement plus nombreux (par rapport à ceux qui n'en ont pas) dans les programmes à horaire décalé que dans les programmes de jour.

Population : l'ensemble des étudiants adultes (25 ans ou plus) inscrits dans les programmes d'études considérés (tous types d'horaire confondus) de toutes les facultés de cette université au cours de l'année académique considérée

Individu : chaque étudiant de 25 ans ou plus apparaissant dans le fichier étudie comme ayant pris son inscription principale dans un des programmes d'étude considérés d'une de ces facultés dans cette université cette année-là.

Variables : 1) le fait d'avoir ou non une charge de famille (sans enfant, avec enfants(s)),
2) le type d'horaire du programme d'inscription (horaire décalé, horaire de jour)

Enoncé 6

Il y a proportionnellement davantage de programmes à horaire décalé (par rapport aux programmes « de jour ») dans les facultés de sciences humaines que dans les facultés de sciences.

Enoncé 6

Il y a proportionnellement davantage de programmes a horaire décalé (par rapport aux programmes « de jour ») dans les facultés de sciences humaines que dans les facultés de sciences.

Population : l'ensemble des programmes d'études considérés (tous types d'horaire confondus) de toutes les facultés de sciences humaines et de sciences de cette université au cours de l'année académique considérée,

Individu : chaque programme d'étude considéré (identifie a partir du catalogue officiel de l'université) dans les facultés de sciences et de sciences humaines de cette université cette année-là,

Variables : 1) le type d'horaire du programme d'étude (horaire décalé, horaire de jour),
2) la faculté d'appartenance du programme (faculté appartenant au secteur des sciences humaines ou faculté appartenant au secteur des sciences)

Chapitre 2

Types de variables

Variable: Caractéristique mesurable à laquelle on peut attribuer plusieurs valeurs différentes

Il existe 2 différent types de variables:

- Les variables **Quantitatives**
- Les variables **Qualitatives**

- Les variables **Quantitatives**

- ❖ Variables **Quantitatives continues**

- ❖ Variables **Quantitatives discrètes**

- ❖ Variables **Quantitatives temporelles**

Les variables Quantitatives continues

- Exemples : « Pression Artérielle »
- Définition:

Ce sont des variables qui peuvent prendre plusieurs valeurs numériques dans l'intervalle des observations.

Champ d'application:

Ce type de variable est particulièrement utilisé en biologie médicale car il permet de quantifier objectivement les observations.

Les variables Quantitatives continues

- Exemples :

le poids: 45,6 kg

le taux de cholestérol: 1,98 g/L

La glycémie: 4,9 mmol/L

La taille: 1,72 m

Les variables Quantitatives continues

Remarque:

Parfois l'observateur décide, par exemple, à partir de sujets dont le poids a été mesuré en Kg avec 2 décimales, **d'arrondir** les résultats.

Alors la variable quantitative continue se transforme en variable quantitative discrète.

Cee procédé est appelé « **discrétisation d'une variable continue** »

Les variables Quantitatives **discrètes**

- Exemple : « **Nombre de rappel d'un vaccin** »
- Définition:

Ce sont des variables qui sont numériques discrètes. Généralement des entiers.

Autres exemples:

La **dentition** : 32 dents

Nombre de patients covid dans le service: 18

Les variables Quantitatives temporelles

- Exemple : « 09-02-2022 »
- Définition:

Ce sont des variables quantitatives particulières qui utilisent les unités de mesure du temps.

Elles se divisent en 2 groupes:

- Variable de durée:
- Variable à un instant donné

Les variables Quantitatives **temporelles**

Variable de durée: secondes, minutes, heure, jours, mois, ans, etc. elles sont de nature continue

Exemple :

Age de la grossesse: 5 mois et 2 semaines

Durée d'incubation d'une toxi-infection: 2h30 min

Les variables Quantitatives **temporelles**

- **Variable à un instant donné**: elles servent à définir un instant donné (début ou fin d'un évènement)
- Ils en existe 2 types principaux:
 - * le type de la date sous la forme **jj-mm-aaaa**
 - * le type horaire sous la forme **hh-mm-ss**

Exemple :

Date de début de la grossesse: 15/04/2013

Heure de début d'une gastroentérite: 22h15

- Les variables **Qualitatives**

- ❖ Variables **Qualitatives** **ordinales**

- ❖ Variables **Qualitatives** **nominales**

- ❖ Variables **Qualitatives** **binaires**

Les variables **Qualitatives ordinales**

Exemple : « **forme clinique d'une maladie** »

- Modalités de la variable : **modérée, moyenne, sévère**
- La forme clinique d'une maladie est un **caractère** de nature **qualitative** : chaque forme clinique correspond à un état de la maladie,
- Il existe ici un ordre hiérarchique objectivement repérable entre les différentes modalités de la variable, ce qui justifie que l'on parle de niveau pour les qualifier : les formes de niveau supérieur de maladie se distinguent des maladies de niveau inférieur par le fait qu'elles possèdent certaines caractéristiques en plus.
- Pour toutes ces raisons, on qualifiera la variable « **forme clinique d'une maladie** » de variable ordinale

• Définition : une **variable ordinale** est une **variable**, le plus fréquemment **de nature qualitative** dont les **modalités sont hiérarchisées** (une variable ordinale peut, dans certains cas, se formuler sous forme numérique : l'ordre d'arrivée, un score sur une échelle, etc.)

Les variables **Qualitatives ordinales**

Exemples :

- Stade de gravité d'une escarre:

Stade 0 : Patient à risque, phase de prévention. Une rougeur réversible apparaît sur le point d'appui. ...

Stade 1 : Apparition d'une rougeur persistante. Une rougeur persistante s'est formée sur le point d'appui. ...

Stade 2 : Dermabrasion, cloques. ...

Stade 3 : Nécrose tissulaire. ...

Stade 4 :

Niveau d'études: universitaire, supérieur non-universitaire

Remarque: en informatique, on associe à chaque modalité un code numérique

Les variables qualitatives **nominales**

Ces variables ne peuvent être hiérarchisées. Elles sont nommées mais pas ordonnées, l'ordre de présentation des modalités est donc **arbitraire**.

Autres types de variables nominales (a titre d'exemple) :

Le **sexe**: hommes vs femmes

Le **groupe sanguin**: A, B, AB, O

La **catégorie de personnel paramédical**: infirmier, kiné, sage-femme, psychologue, ...

- Il peut arriver que l'on désigne les différentes **modalités d'une variable nominale** par un **code numérique**.

Il s'agit en fait d'une pratique très courante destinée à faciliter les procédures de traitement des données.

On notera cependant que **l'assignation d'un code numérique aux modalités d'une variable qualitative ne modifie pas pour autant la nature de la variable**.

Les variables qualitatives **binaires**

- Exemple : « la survie d'un patient covid »
- Modalités de la variable : vivant, décédé

Ces variables qui ne peuvent prendre que **2 valeurs**.

-> variables **dichotomiques**: Malade ou Sain

-> variables **booléennes**: VRAI ou FAUX

-> variables de **Bernoulli** : 0 ou 1

Exercice 1

Pour chaque recherche proposée ci-dessous, examinez la ou les variables étudiées et pour chacune d'entre-elles, précisez s'il s'agit d'une variable nominale, ordinale ou métrique

Etude de la répartition des étudiants inscrits à l'université selon le genre, l'âge et la faculté.

- Genre (qualitative nominale);
- âge (quantitative continue);
- Faculté d'appartenance (qualitative nominale)

Exercice 2

Pour chaque recherche proposée ci-dessous, examinez la ou les variables étudiées et pour chacune d'entre-elles, précisez s'il s'agit d'une variable nominale, ordinale ou métrique

Etude de la répartition du nombre d'échecs, d'abandons et de réussites par niveau d'étude, par type d'enseignement et par réseau d'enseignement en Communauté française de Belgique pour l'année 2003-2004

- Performance (échecs, abandons et réussites) de l'étudiant (qualitative nominale);
- Niveau d'étude (qualitative ordinale);
- Type d'enseignement et réseau d'enseignement (qualitative nominales)

Exercice 3

Pour chaque recherche proposée ci-dessous, examinez la ou les variables étudiées et pour chacune d'entre-elles, précisez s'il s'agit d'une variable nominale, ordinale ou métrique

Comparaison des attitudes des chefs d'entreprise de la Région-Nord Pas de Calais a l'égard des partis d'extrême-droite (enquête par questionnaire utilisant des échelles d'attitude) selon le niveau d'études du chef d'entreprise, et le secteur d'activité de son entreprise.

- Attitudes (qualitative ordinale) ;
- Niveau d'études (qualitative ordinale);
- Secteur d'activité (qualitative nominale)

Population

```
graph TD; Population[Population] --> Echantillon[Echantillon]; Echantillon --> Variable[Variable]; Variable --> Quantitative[Quantitative]; Variable --> Qualitative[Qualitative]; Quantitative --> Discrete[Discrète]; Quantitative --> Continue[Continue]; Quantitative --> Temporelle[Temporelle]; Qualitative --> Ordinale[Ordinale]; Qualitative --> Nominale[Nominale]; Qualitative --> Binaire[Binaire];
```

Echantillon

Variable

Quantitative

Discrète

Continue

Temporelle

Qualitative

Ordinale

Nominale

Binaire

Exercice

Déterminer quelle type de variable s'agit-il?

N°	Variable	LOUISE	MICHEL	JURGEN
1	Date de naissance	08-04-1977	30-06-1978	12-07-1990
2	Age au 1 ^{er} janvier 2013	35 ans	34 ans	22 ans
3	Sexe	F	M	M
4	Taille (m)	1,68	1,78	1,82
5	Fratric	2	1	5
6	Niveau d'études	Infirmière	kiné	Dentiste
7	Etat civil	Célibataire	Marie	Divorcé
8	Groupe sanguin	A	AB+	O-
9	Tabagisme	EX fumeur	Fumeur	Non-fumeur
10	Couleur des yeux	Noir	Vert	Marron
11	Couleur des cheveux	Bold	Noir	Châtain
12	Nombre de pièces du logement	2	2	3

Chapitre 3

Présentation des données

Après avoir recueilli des données dans une étude statistique, l'étape suivante consiste à décrire l'ensemble sous une forme synthétique.

Trier et ensuite regrouper les données afin de les visualiser sous une forme condensée qui permet d'appréhender globalement leur distribution

Tri des données

Supposons que l'on réalise un sondage dont l'unique question est la suivante:

« *Quelle est la boisson que vous consommez le plus fréquemment avec le repas du soir ?* »

Les réponses peuvent être choisies dans la liste :

eau	E
limonade	L
bière	B
vin	V
café	C
thé	T
alcool	A
autre (divers)	D

Les résultats **bruts** de l'enquête sont notés sur des formulaires nominatifs:

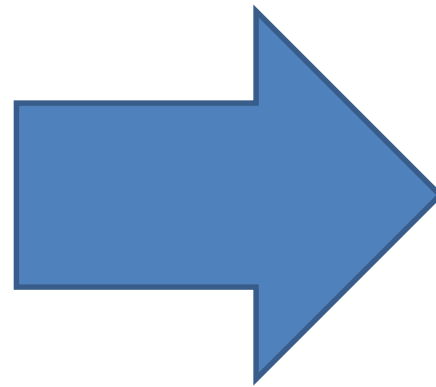
Delphine	L
Rose	V
Jean-Philippe	L
Marylin	C
Maude	E
Stéphanie	B
Julie	B
Olivier	D
Johanne	B
Julien	E
Sandrine	V
Justine	V
Anita	L
Stéphanie	B
Christine	B
Kristel	V
Aurore	A
Jean-Yves	B
François-Michael	E
Fabian	B
Louise	L
Stéphane	A
Anthony	E
Barbara	B
Macha	B

Remarque:

La liste détaillée des résultats ne se prête pas bien à une interprétation globale.

Données bruts

Delphine	L
Rose	V
Jean-Philippe	L
Marylin	C
Maude	E
Stéphanie	B
Julie	B
Olivier	D
Johanne	B
Julien	E
Sandrine	V
Justine	V
Anita	L
Stéphanie	B
Christine	B
Kristel	V
Aurore	A
Jean-Yves	B
François-Michael	E
Fabian	B
Louise	L
Stéphane	A
Anthony	E
Barbara	B
Macha	B



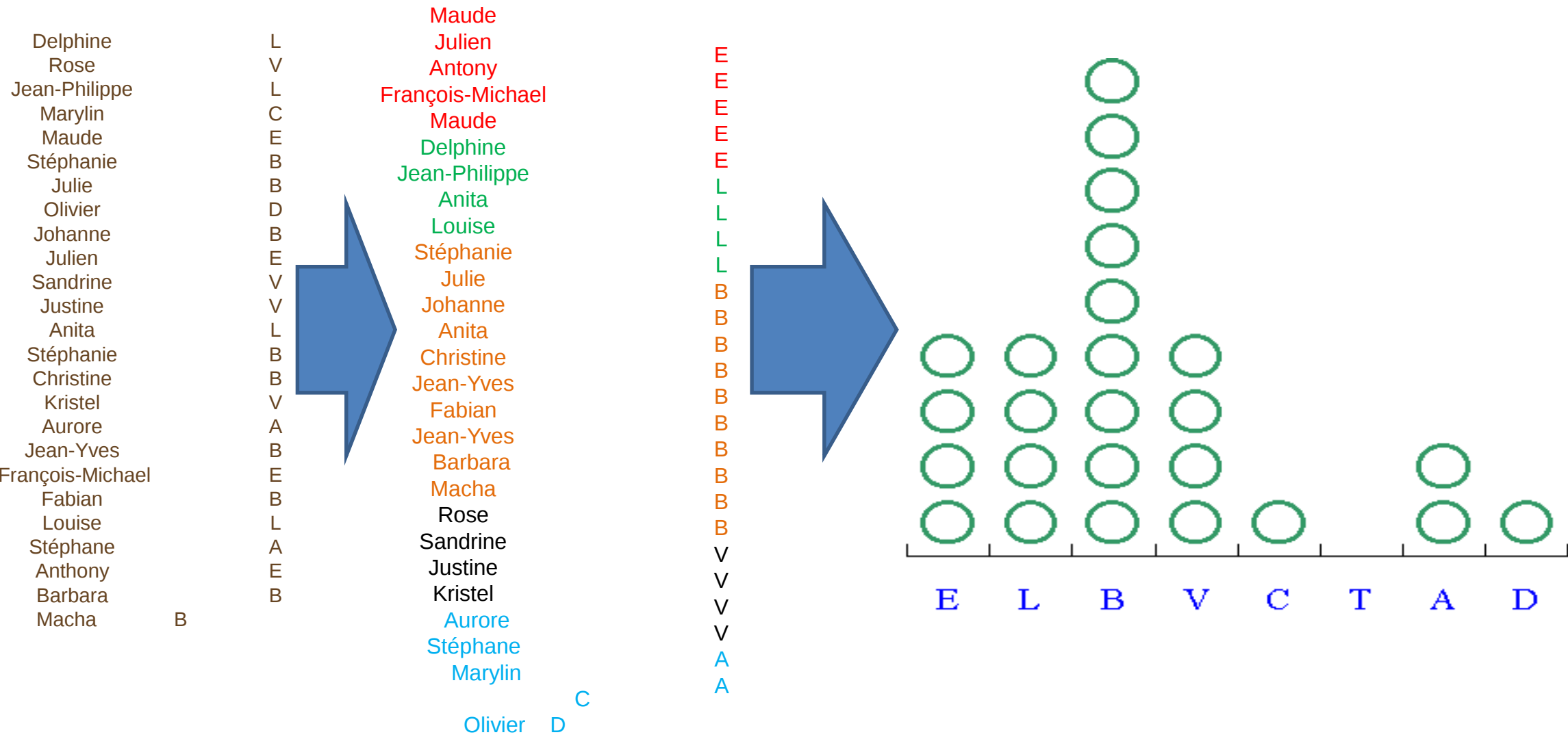
données ordonnées

Maude	E
Julien	E
Antony	E
François-Michael	E
Maude	E
Delphine	E
Jean-Philippe	L
Anita	L
Louise	L
Stéphanie	L
Julie	B
Johanne	B
Anita	B
Christine	B
Jean-Yves	B
Fabian	B
Jean-Yves	B
Barbara	B
Macha	B
Rose	B
Sandrine	V
Justine	V
Kristel	V
Aurore	V
Stéphane	V
Marylin	A
Olivier	A

données bruts

données ordonnées

données groupées



Processus du tri des données

Etape 1: Tableau **Brut**

Etape 2: Tableau **Ordonné**

Etape 3: Tableau **Groupé**

Etape 4: Tableau **Recensé**

Exemple 1: variable discrète

Modalités X_i	Effectifs n_i	Fréquences $f_i = \frac{n_i}{N}$	Effectifs Cumulés	Fréquences Cumulées
A	4			
B	4			
C	9			
D	4			
E	1			
F	2			
G	1			

Exemple 1: variable continues

Classe d'âge X_i	0-19	20-39	40-59	60-79	80-99
Effectifs n_i	8	7	3	3	4

Classe Modalités	Valeurs centrales x_i	Effectifs n_i	Fréquences $f_i = \frac{n_i}{N}$	Effectifs Cumulés	Fréquences Cumulées
0-19		8			
20-39		7			
40-59		3			
60-79		3			
80-99		4			

EXERCICE:

Nous avons demandé aux étudiants de 5 -ème année en sciences de l'université Libre de Bruxelles, combien de vaccin anti-covid vous avez reçu.

Les résultats de l'enquête sont représentés dans le **tableau brut** suivant:

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

Tableau ordonnée

Effectifs = nombre d'étudiants =

Modalités de la variable (Vaccin)=

Tableau groupé

Modalités (Nombre de vaccin) x_i	1	2	3	4
Effectifs (Nombre d'étudiants) n_i				

TOTAL N =

Tableau recensé

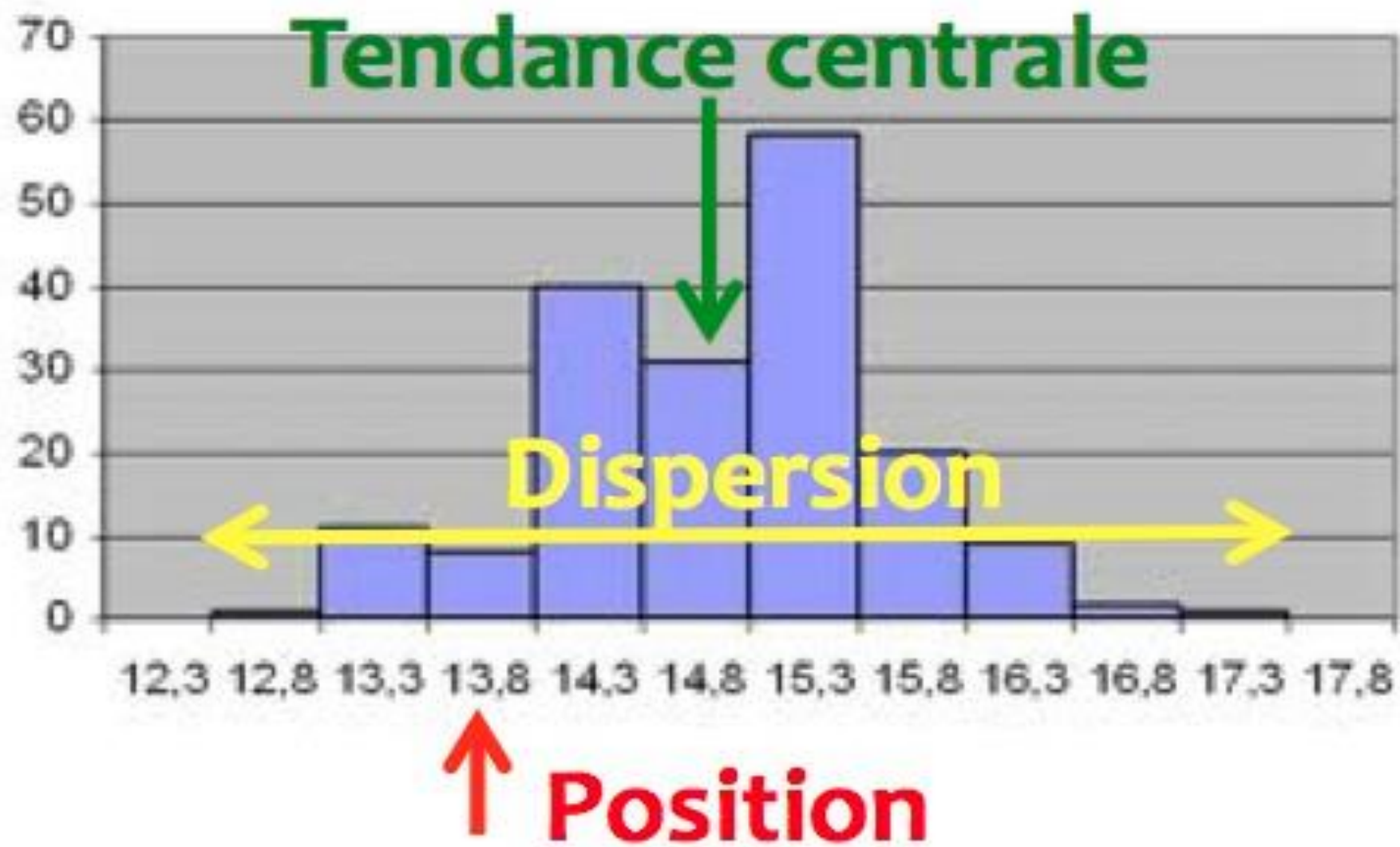
Modalités X_i	Effectifs n_i	Fréquences $f_i = \frac{n_i}{N}$	Effectifs Cumulés	Fréquences Cumulées

TOTAL =

Paramètres de distribution

□ Paramètres de position

□ Paramètres de dispersion



Paramètres de distribution

Remarque:

Deux cas peuvent être traitées en statistique:

- ❖ Série de valeurs
- ❖ Tableau de valeurs

Paramètres de distribution

❖ Série de valeurs

Taille (cm) : 145 166 188 147 110 180

❖ Tableau de valeurs

Modalité x_i	3	5	1	8	10
Effectif n_i	2	2	2	4	6

□ Paramètres de position:

Modalité x_i	x_1	x_2	...	x_p	Total
Effectif n_i	n_1	n_2	...	n_p	$N = n_1 + \dots + n_p$

$i = 1 \dots p$

□ Paramètres de position:

La **moyenne** de la série statistique est le nombre $\bar{x}$, tel que :

$$\bar{X} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_p \cdot x_p}{n_1 + n_2 + \dots + n_p}$$

❑ Paramètres de position:

Exemple:

Modalité x_i	3	5	1	8	10
Effectif N_i	2	2	2	4	6

$$\text{Moyenne } \overline{X} = \frac{\dots\dots\dots}{\dots\dots\dots}$$

$$\overline{X} =$$

❑ Paramètres de position:

Exemple:

Les notes d'un étudiant en 3ème année Bachelier infirmier:

10 12 10 13 09 07 15 12 14 17

$$\bar{X} = \frac{X_1 + X_2 + X_3 \dots + X_p}{p} \quad (p = 10)$$

$$\bar{X} = \frac{\dots\dots\dots}{\dots\dots}$$

□ Paramètres de position:

La médiane

On considère une liste de N données rangées par ordre croissant.

Si la série est de taille impaire ($N=2n+1$), la *médiane* est la donnée de rang $n+1$.

Si la série est de taille paire ($N=2n$), la *médiane* est la demi-somme des données de rang n et de rang $n+1$.

❑ Paramètres de position:



La **médiane**

10 12 10 13 09 07 15 12 14 17 11

Il faut tout d'abord **ordonner** la série:

Le nombre de valeurs est égal à, qui est

La médiane $Me =$

❑ Paramètres de position:



La **médiane**

80 120 110 113 92 77 115 115 80 117 111 121

Il faut tout d'abord **ordonner** la série:

Le nombre de valeurs est égal à, qui est

La médiane $Me=$

❑ Paramètres de position:

La **médiane**

Longueur (en m)	37	39	40	41	42	43	44	48
Effectif	4	3	4	3	2	3	5	2

Effectif totale N = qui est **pair** ➔

La médiane est la modalité correspondante à la 13 -ème valeur de l'effectif.

37 37 37 37 39 39 39 40 40 40 40 41 41 41 42 42 43 43 43

 Me = 41

Exercice:

Calculer la moyenne et la médiane

Nombres de patients COVID	8	11	17	22	29	33	44	66
Effectif	5	1	9	3	10	9	9	2

Exercice:

Calculer la moyenne et la médiane

Taille (cm) : 145 166 188 147 110 180

❑ Paramètres de position:

Les **quartiles**

Soit donnée une liste de N données.

Le **premier quartile** est la plus petite donnée **Q1** de la liste telle qu'**au moins un quart des données** de la liste sont inférieures ou égales à $Q1$.

Le **troisième quartile** est la plus petite donnée **Q3** de la liste telle qu'**au moins les trois quarts des données** de la liste sont inférieures ou égales à $Q3$.

Le **deuxième quartile** est la plus petite donnée **Q2** qui est la **médiane**

❑ Paramètres de position:

Les **quartiles**

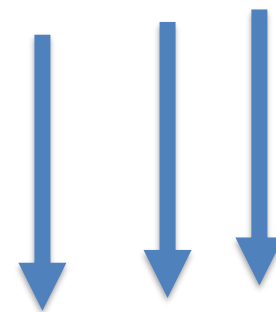
10 12 10 13 09 07 15 12 14 17 11

Le nombre est

Q2=

Q1=

Q3 =



❑ Paramètres de position:

Les **quartiles**

80 120 110 113 92 77 115 115 80 117 111 121

Le nombre est

Q2=

Q1=

Q3 =



❑ Paramètres de dispersion:

Etendue

L'étendue d'une série statistique à caractère quantitatif est la différence entre la plus grande valeur de la série et la plus petite.

Ecart interquartile

L'écart interquartile d'une série statistique à caractère quantitatif est la différence entre le 3ème quartile et le premier quartile. $IQ = Q3 - Q1$

❑ Paramètres de dispersion:

80 120 110 113 92 77 115 115 80 117 111 121

Etendue = Valeur maximale – Valeur Minimale
=

Q1=

Q3 =

IQ = Q3 – Q1 =

❑ Paramètres de dispersion:

Longueur (en m)	37	39	40	41	42	43	44	48
Effectif	4	3	4	3	2	3	5	2

Etendue = Valeur maximale – Valeur Minimale
=

IQ = Q3 – Q1
=

Paramètres de dispersion

la **variance**:

la variance est la moyenne des carrés des écarts à cette moyenne :

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Paramètres de dispersion

la **variance**:

la variance est la moyenne des carrés des écarts à cette moyenne :

Dans le cas d'une série de valeurs:

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Dans le cas d'un tableau de valeurs:

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

Paramètres de dispersion

Ecart type:

L'**écart-type** d'une série statistique est le nombre réel, noté σ , défini par:

$$\sigma = \sqrt{v}$$

La variance et l'écart-type sont des **paramètres de dispersion** d'une série statistique.

L'avantage de l'écart-type est de s'exprimer dans la même unité que les X_i

Exercice: Série de valeurs

Calculer la variance et l'écart type:

120 110 113 92 77 115 115 80 117 111 121

Exercice: Tableau de valeurs

Calculer la variance et l'écart type

Longueur (en m)	37	39	40	41	42	43	44	48
Effectif	4	3	4	3	2	3	5	2

Diagramme en bâtons

Lorsque la variable étudiée est **quantitative discrète**, on peut représenter la série statistique par un diagramme en bâtons : La **hauteur de chaque bâton est proportionnelle à l'effectif** (ou la fréquence) associée à chaque valeur.

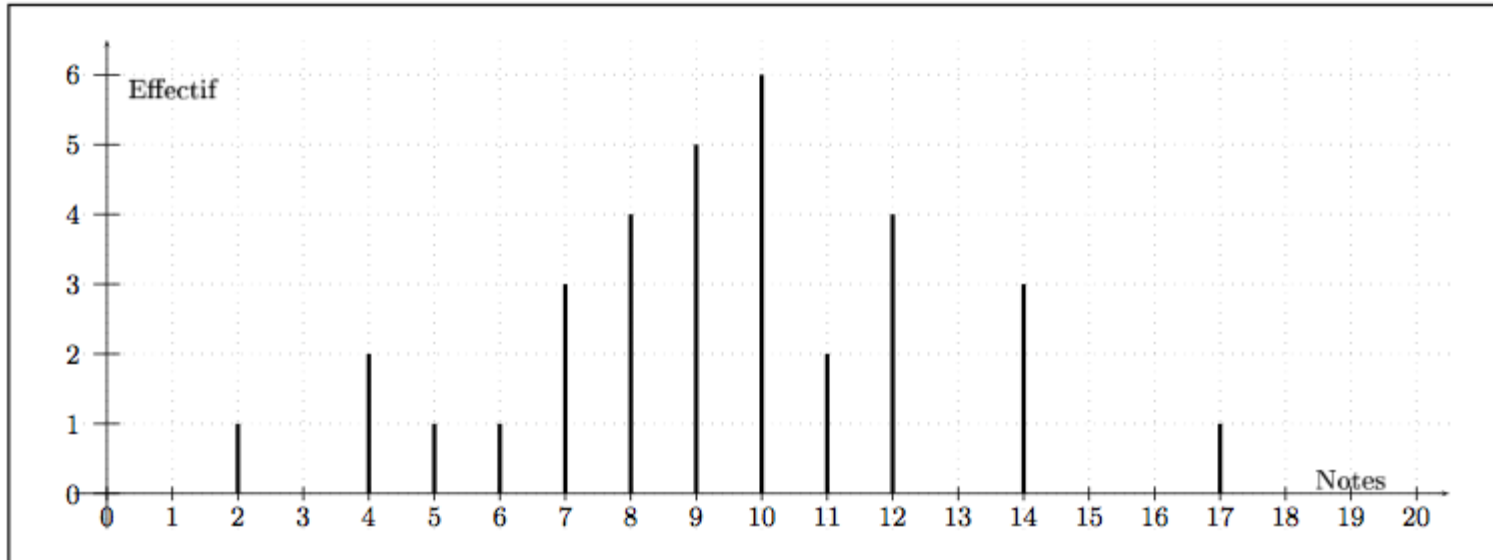
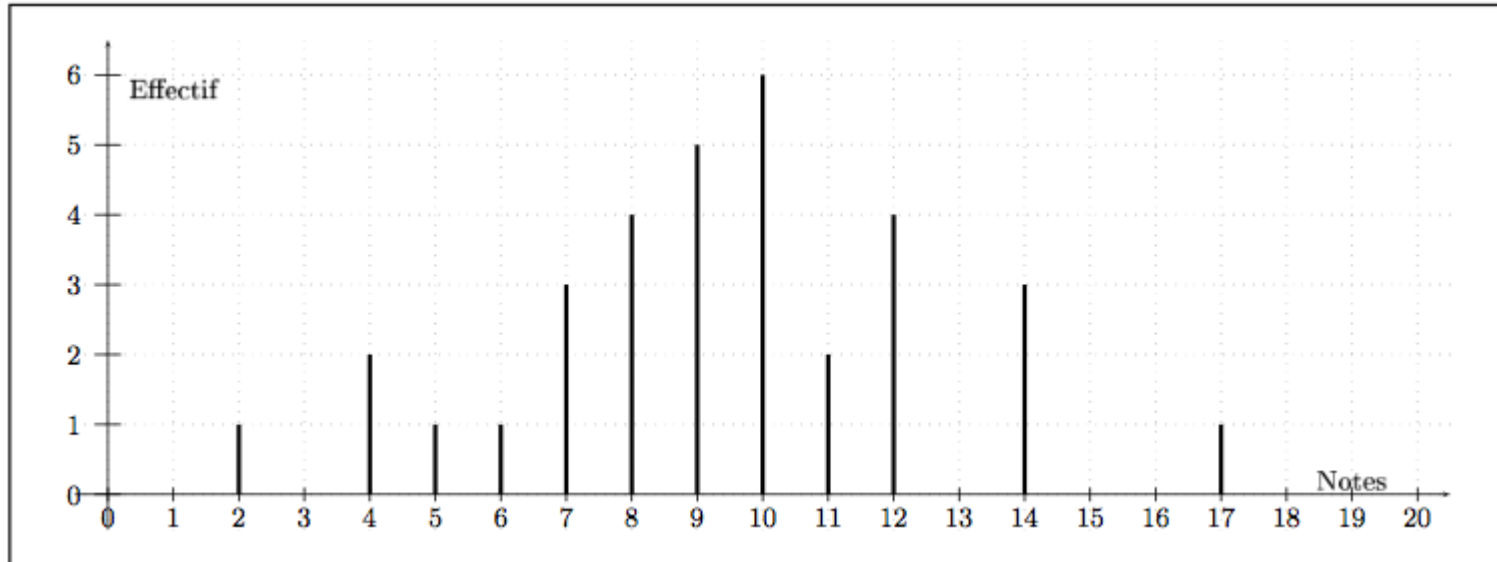


Diagramme en bâtons

[illegible]

Histogramme

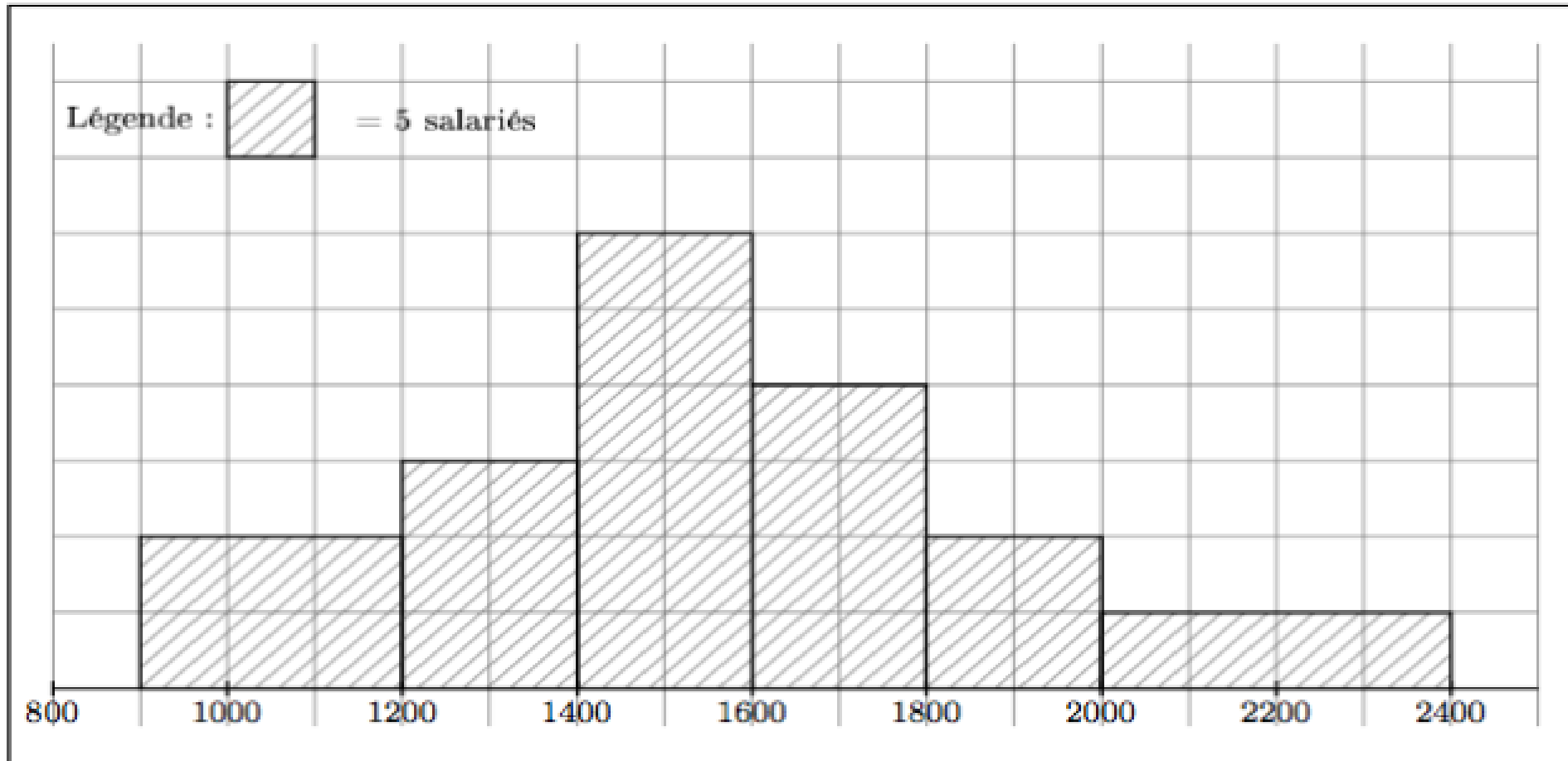
Lorsque la variable étudiée est **quantitative continue**, et lorsque les données sont **regroupées en classes**, on peut représenter la série par un *histogramme* :

L'aire de chaque rectangle est alors proportionnelle à l'effectif (ou à la fréquence) associée à chaque classe.

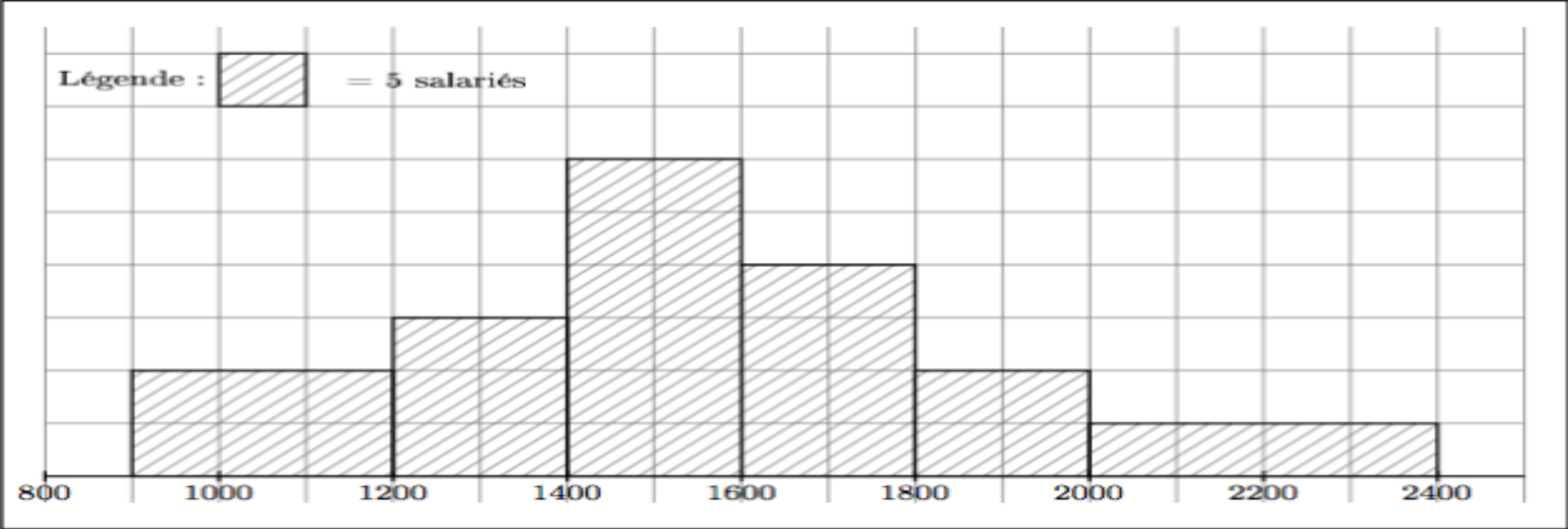
Lorsque les classes ont la **même amplitude**, c'est la **hauteur de chaque rectangle qui est proportionnelle à l'effectif**.

Histogramme

Voici un histogramme représentant la répartition des salaires dans une entreprise :



Histogramme



Salaires	[900;1200[	[1200;1400[	[1400;1600[	[1600;1800[	[1800;2000[	[2000;2400[	Total
Effectif							
Fréquence							

diagramme circulaire

Lorsque la **variable est qualitative**, on représente la série par un *diagramme circulaire ou semi-circulaire (camembert)* :

La mesure de chaque secteur angulaire est proportionnelle à l'effectif (ou à la fréquence) associée à la valeur du caractère.

Le **diagramme sectoriel** ou "camembert" se prête très bien à la représentation des pourcentages.

On dessine un disque découpé en secteurs ou "morceaux de tarte". L'angle au centre de chaque secteur est proportionnel au pourcentage d'individus dans la catégorie correspondante.

