

STATISTIQUES

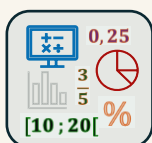
Collecter, traiter, analyser, interpréter les données

- Statistiques descriptives – 2^{nde} professionnelle**
Statistiques à une variable – partie 1 – septembre octobre
- Statistiques descriptives – 2^{nde} professionnelle**
Statistiques à une variable – partie 2 – mars avril

Sommaire	Page
Mots clés	3
Liste des connaissances et capacités mathématiques associées	3
Introduction	4
Les étapes de traitement des données	4
Les tableaux	5
→ Structure – Caractère - Effectifs	5
→ Fréquences	5-6
Le caractère statistique – exemples	7
Les diagrammes possibles - exemples	8
Les paramètres de position	10
→ Le mode	10
→ La moyenne	11
→ La médiane	12
→ Les quartiles – schéma de synthèse	13



<https://ladigitale.dev/digislides/s/671d295348e6c/>



Connaissances et capacités mathématiques associées

Recueillir et organiser des données statistiques.

- Organisation sous forme de tableaux – effectifs, catégories, variable
- Présentation de données en pourcentage – **fréquences***
 - Calculer, appliquer un **pourcentage*** / Associer et utiliser les écritures – **fractionnaire – décimale - en pourcentage*** / Calculer avec **les puissances de dix*** / Effectuer un **arrondi***
- Regroupement par classes d'une série statistique
 - Lire et interpréter un intervalle de valeurs

Organiser des données statistiques en choisissant un mode de représentation adapté à l'aide des fonctions statistiques d'une calculatrice ou d'un tableur.

- Diagramme en secteurs, en bâtons, en colonnes, à lignes brisées.

Extraire des informations d'une représentation d'une série statistique.

Interpréter des séries statistiques à l'aide d'indicateurs de position

- Mode, classe modale, **moyenne***, médiane, quartiles.

***Automatismes à acquérir et entretenir**

Introduction

Il s'agit d'étudier et résumer un ensemble d'observations collectées sur une **population*** en les rendant éclairantes et significatives.

L'objectif principal est donc **la description de données** au travers :**

- De leur **présentation** la plus commode et synthétique possible – souvent des tableaux,
- De leur **représentation graphique** – diagrammes,
- Du **calcul de caractéristiques numériques** – moyenne, mode etc.

Pour une population, l'ensemble des données relève d'un seul critère d'observation appelé **caractère statistique** et associé à une variable mathématique.

Chaque catégorie ou valeur de ce caractère est associé à un **effectif** de la population totale.



*Une **population** au sens statistique désigne un ensemble d'humains, d'animaux, d'objets, de bactéries, d'organisations, etc.

Chaque élément de cette population est un **individu**.

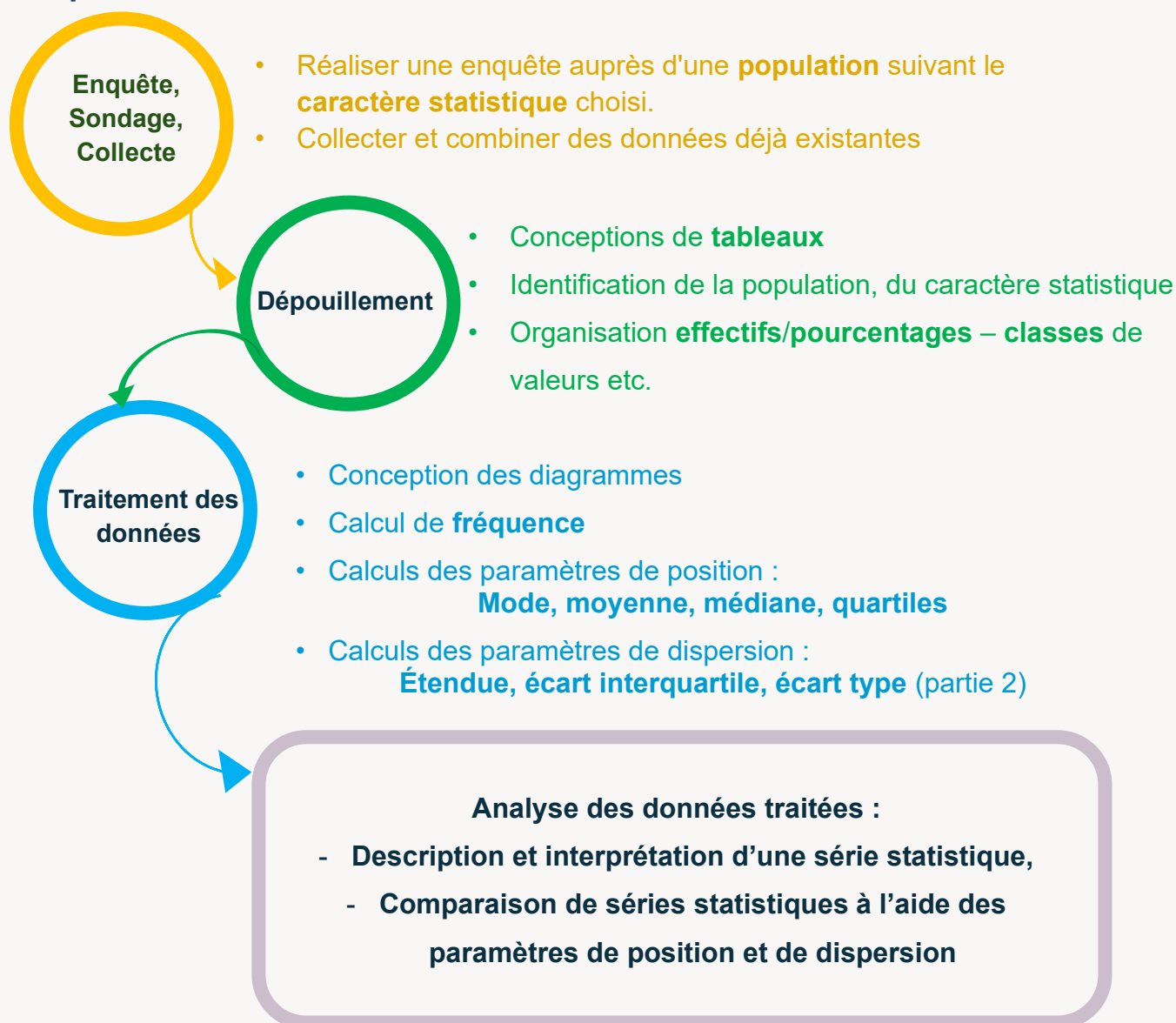


** Le terme « **données** », très utilisé en statistique désigne,

- **L'ensemble** des **individus** observés,
- La **variable** considérée,
- Les **valeurs** de cette variable associées aux **effectifs** de population.

*Image Pixabay
Dapple designers*

I. Les étapes du traitement des données



II. Les tableaux

Ils permettent de présenter de façon synthétique un ensemble de données, c'est une étape préparatoire aux graphiques et calculs.

Nombre de bon(s) de commandes par vendeur/vendeuse
60 vendeurs/vendeuses – octobre 2024 – semaine 1

1	6	4	9	2	7	6	4
1	2	5	6	9	0	3	8
8	0	4	3	10	1	4	0
1	7	2	7	7	2	4	8
5	8	1	1	8	7	7	3
4	3	6	2	7	10	3	1
6	3	8	10	1	1	3	7
5	7	8	1				

Deux possibilités de présentation synthétique des données

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nombre de bons de commande	0	1	2	3	4	5	6	7	8	9	10	
2	Effectifs - vendeurs/vendeuses	3	10	5	7	6	3	5	9	7	2	3	60

Caractère statistique
Variable discrète

Effectif total := somme(B2; L2)

Effectifs : nombre de vendeurs/vendeuses

	A	B	C	D	E	F	G	H
1	Nombre de bons de commande	[0 ; 2[[2 ; 4[[4 ; 6[[6 ; 8[[8 ; 10[[10 ; 12[
2	Effectifs - vendeurs/vendeuses	13	12	9	14	9	3	60

Caractère statistique
Variable continue

Effectif total := somme(B2; G2)

La fréquence d'une valeur du caractère

Elle se calcule par **comparaison de son effectif par rapport à l'effectif total** :

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nombre de bons de commande	0	1	2	3	4	5	6	7	8	9	10	
2	Effectifs - vendeurs/vendeuses	3	10	5	7	6	3	5	9	7	2	3	60

Unité de comparaison

La fréquence de « 7 bons de commande » : $\frac{9}{60} = 0,15 = \frac{15}{100}$ soit 15%

Fréquence sous sa forme
numérique

Fréquence en % de
l'effectif total

D'où : 15% des vendeurs/vendeuses ont chacun 9 bons de commandes complétés.

Les fréquences en complément dans le tableau initial

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nombre de bons de commande	0	1	2	3	4	5	6	7	8	9	10	
2	Effectifs - vendeurs/vendeuses	3	10	5	7	6	3	5	9	7	2	3	60
3	Fréquence	0,05	0,17	0,08	0,12	0,10	0,05	0,08	0,15	0,12	0,03	0,05	1
4	Fréquence en %	5	17	8	12	10	5	8	15	12	3	5	100

$$= \frac{B2}{60}$$

$$= \frac{E2}{60} \times 100$$

$$\text{Total} := \text{somme}(B4; L4)$$

III. Le caractère statistique

Il peut être,

QUALITATIF (1)

Caractère statistique non mesurable :

Les types d'activité

	Population (%)
Actifs ayant un emploi	56,7
Chômeurs	11,4
Retraités	5,5
Élèves, étudiants et stagiaires non rémunérés	16,8
Autres inactifs	9,6

Population de 15 à 64 ans par type d'activité en 2021 - La Rochelle
Source : Insee, RP2021 exploitation principale, géographie au 01/01/2024.

QUANTITATIF

- Discret (2)

Caractère statistique mesurable au moyen de la **variable** :

Le nombre d'étoile(s)

La **variable** peut prendre un nombre fini de valeurs – ici : **1, 2, 3, 4 et 5**.

Nombre d'étoiles	Hôtels
1	1
2	7
3	16
4	11
5	1
Ensemble	35

Effectifs

Effectif total

- Continu (3)

Caractère statistique mesurable au moyen de la **variable** : **L'âge en année(s)**

La **variable** peut prendre une infinité de valeurs **entre 0 et 105**, sans rupture et réparties sous forme **d'intervalles** successifs – **ici ce sont des classes d'âges** : [0 ; 15[puis [15 ; 30[etc.

Âges	Effectifs
[0 ; 15 [10 265
[15 ; 30 [18 432
[30 ; 45 [12 281
[45 ; 60 [14 002
[60 ; 75 [14 070
[75 ; 90 [8 535
[90 ; 105 [950
Ensemble	78 535

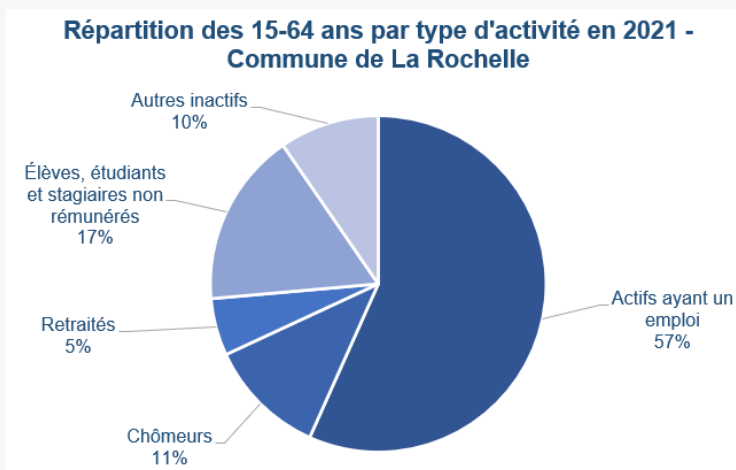
Hôtels - La Rochelle et environs classés par nombre d'étoile(s) – octobre 2024
Source : <https://www.larochelle-tourisme.com>

Population rochelaise par tranches d'âges – 2021
Source : INSEE RP 2021 – exploitations principales; Géographie au 01/01/2024 et ville-data.com

IV. Les diagrammes possibles

(1) LE DIAGRAMME EN SECTEURS

- Il est adapté principalement pour les séries dont le caractère est **qualitatif**.
- Chaque secteur représente une **catégorie – ici, le type d'activité**.
- L'aire du secteur est proportionnelle à la **fréquence** - Ou à l'**effectif**.

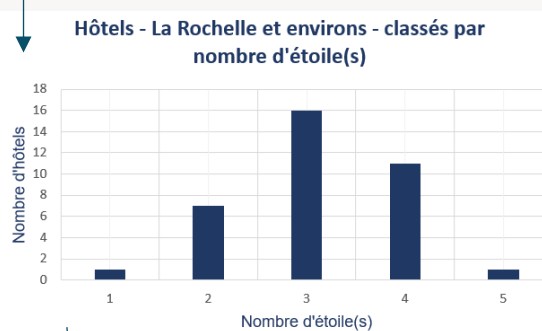


Population de 15 à 64 ans par type d'activité en 2021 - La Rochelle
Source : Insee, RP2021 exploitation principale, géographie au 01/01/2024.

(2) (1) LE DIAGRAMME EN BÂTONS

- Il est adapté principalement pour les séries dont le caractère est **quantitatif discret** ou **qualitatif**
- La hauteur de chaque bâton est proportionnelle à l'**effectif** - ou à la **fréquence**.

Effectifs – Ici le nombre d'hôtels

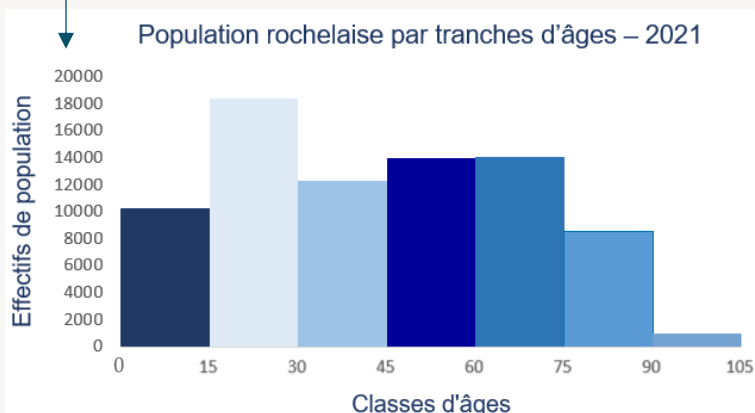


Caractère, **quantitatif discret** – Ici, le nombre d'étoile(s)

(3) LE DIAGRAMME EN COLONNES

- Il est adapté pour les séries dont le caractère est **quantitatif continu**.
- Chaque colonne représente une classe de valeurs du caractère.
- La hauteur de la colonne est proportionnelle à l'**effectif** - ou à la **fréquence**.

Effectifs – Issus de la population rochelaise 2021

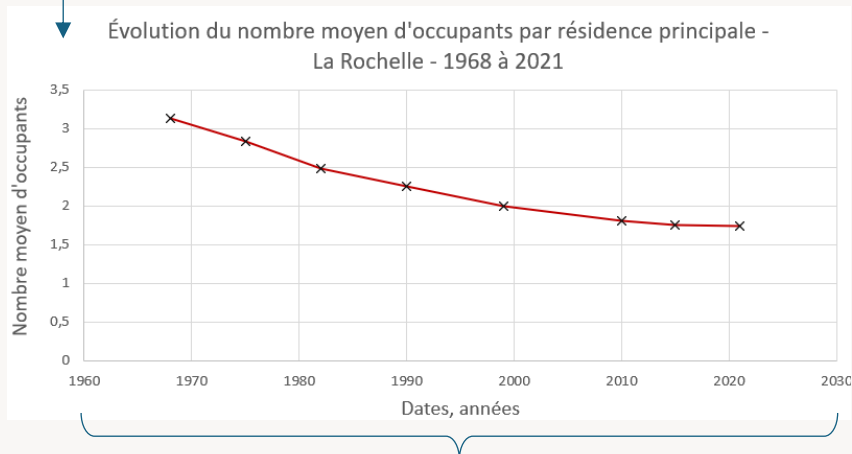


Caractère, **quantitatif continu** – Ici, l'âge des habitants sous forme d'intervalles

LE DIAGRAMME À LIGNES BRISÉES

- Il est adapté pour les séries dont le **caractère varie dans le temps**.
- Les points reliés forment une suite continue de segments. L'ensemble constitue une **ligne brisée**.

Les valeurs du caractère – Ici, le nombre moyen d'occupants par résidence principale



Sources : Insee, RP1967 à 1999 dénombremments, RP2010 au RP2021 exploitations principales.

Le temps – Ici sous forme de dates, l'unité de graduation est de 10 ans. L'axe horizontal est gradué de 1960 à 2030

V. Les paramètres de position - Méthodes

Ils permettent de décrire et résumer une série statistique au **caractère quantitatif continu**.

Le Mode

C'est la **valeur du caractère la plus fréquente**. Elle peut se lire dans **un tableau** ou sur un **graphique**.

Nombre d'étoiles	Hôtels
1	1
2	7
3	16
4	11
5	1
Ensemble	35

Le **mode** de cette série est la valeur **3**.

Âges	Effectifs
[0 ; 15 [10 265
[15 ; 30 [18 432
[30 ; 45 [12 281
[45 ; 60 [14 002
[60 ; 75 [14 070
[75 ; 90 [8 535
[90 ; 105 [950
Ensemble	78 535

La **classe modale** est **[15 ; 30[**.

Le **mode** est **22,5**.

La moyenne

La valeur moyenne d'une série statistique (valeur du caractère ; effectif associé) se note \bar{x} et s'obtient en effectuant :

$$\bar{x} = \frac{\text{Somme des valeurs de la variable}}{\text{Nombre de valeurs}}$$

Les cinq derniers achats d'une famille dans une grande surface sont :

75 € ; 90 € ; 55 € ; 120 € ; 100 €

Chaque valeur est unique.

$$\bar{x} = \frac{75 + 90 + 55 + 120 + 100}{5}$$

$$\bar{x} = 88$$

Le montant moyen des cinq tickets de courses est de 88,00 €

Nombre d'étoiles	Hôtels
1	1
2	7
3	16
4	11
5	1
Ensemble	35

Les valeurs de « Nombre d'étoiles » peuvent concerner plusieurs hôtels.

$$\bar{x} = \frac{1 \times 1 + 2 \times 7 + 3 \times 16 + 4 \times 11 + 5 \times 1}{35}$$

$$\bar{x} = 3,2$$

Le nombre moyen d'étoiles par hôtel est 3,2

Âges	Effectifs
[0 ; 15 [10 265
[15 ; 30 [18 432
[30 ; 45 [12 281
[45 ; 60 [14 002
[60 ; 75 [14 070
[75 ; 90 [8 535
[90 ; 105 [950
Ensemble	78 535

Chaque classe d'âge est représentée par sa valeur centrale :

7,5 pour [0 ; 15[

22,5 pour [15 ; 30[

Etc.

$$\bar{x} = (7,5 \times 10265 + 22,5 \times 18432 + 37,5 \times 12281 + 52,5 \times 14002 + 67,5 \times 14070 + 82,5 \times 8535 + 97,5 \times 950) \div 78535$$

$$\bar{x} = 43,7$$

L'âge moyen de la population rochelaise en 2021 était de 43,7 ans.

La médiane

La **médiane** d'une série statistique se détermine à partir des valeurs du caractère **ordonnées dans l'ordre croissant**.

Elle se note **Med**.

Définition : La médiane est la valeur du caractère pour laquelle,

- Au moins 50 % des valeurs lui sont inférieures ou égales.
- Au moins 50 % des valeurs lui sont supérieures ou égales.

Les cinq derniers achats d'une famille dans une grande surface s'élèvent à :

75 € ; 90 € ; 55 € ; 120 € ; 100 €

Dans l'ordre croissant :

55 ; 75 ; 90 ; 100 ; 120

3 valeurs

Le montant médian des achats est de 90 €.

Med = 90

Les salaires nets des six employés d'une TPE sont :

1 475 € ; 2 540 € ; 1 690 € ; 2055 € ; 3120 € ; 1530 €

Dans l'ordre croissant :

1 475 ; 1 530 ; 1 690 ; 2055 ; 2 540 ; 3120

$Med = (1690 + 2055) \div 2 = 1872,5$

Le salaire médian dans cette TPE est de 1872,50 €

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nombre de bons de commande	0	1	2	3	4	5	6	7	8	9	10	
2	Effectifs - vendeurs/vendeuses	3	10	5	7	6	3	5	9	7	2	3	60

3 valeurs inférieures ou égales à 0

13 valeurs inférieures ou égales à 1

18 valeurs inférieures ou égales à 2

25 valeurs inférieures ou égales à 7

31 valeurs inférieures ou égales à 4

La trentième valeur est un 4.

Dans le cas d'un grand nombre de valeurs, ici 60 nombres de bon de commande :

- 1- **Calculer** : *effectif total* $\div 2$ $60 \div 2 = 30$
- 2- **Ranger** les valeurs du caractère dans l'ordre croissant
- 3- **Rechercher** la trentième valeur

Le nombre médian de nombre de bons de commande par vendeur/vendeuse est 4.

Les quartiles

Les **quartiles Q_1 et Q_3** d'une série statistique :

- **Q_1** : c'est la plus petite valeur du caractère telle que **au moins 25%** des valeurs lui sont inférieures ou égales.
- **Q_3** : c'est la plus petite valeur du caractère telle que **au moins 75%** des valeurs lui sont inférieures ou égales.
- La **médiane** est le **deuxième quartile**.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Nombre de bons de commande	0	1	2	3	4	5	6	7	8	9	10	
2	Effectifs - vendeurs/vendeuses	3	10	5	7	6	3	5	9	7	2	3	60

←
3 valeurs inférieures ou égales à 0

←
13 valeurs inférieures ou égales à 1

←
18 valeurs inférieures ou égales à 2

Pour déterminer la valeur de **Q_1** :

- 1- **Calculer** : *effectif total* $\div 4$ $60 \div 4 = 15$
- 2- **Ranger** les valeurs du caractère dans l'ordre croissant
- 3- **Rechercher** la quinzième valeur

La quinzième valeur est un 2. D'où **$Q_1 = 2$**

Schéma de résumé des valeurs d'une série statistique.

