

Statistique

Plan du chapitre

I	Vocabulaire	2
II	Effectifs et fréquences	3
	A - Effectifs et effectifs cumulés.....	3
	B - Fréquences et fréquences cumulées	3
III	Moyenne et écart-type	4
	A - Moyenne (caractéristique de position).....	4
	B - Écart-type et variance (caractéristique de dispersion)	6
	C - Résumé d'une série statistique	8
IV	Médiane, quartiles, écart inter-quartile	8
	A - Médiane	8
	B - Quartiles (caractéristique de position)	9
	C - Écart inter-quartile (caractéristique de dispersion).....	10
	D - Diagramme en boîte	10
	E - Résumé d'une série statistique.....	11
V	Exercices	12

Introduction

Le rôle de la statistique descriptive est de présenter une masse de données sous forme lisible. Puis si possible, de la résumer par quelques valeurs caractéristiques...

Partie I Vocabulaire

Définition 1 : Série statistique

- L'ensemble sur lequel porte l'étude statistique s'appelle la **population** ;
- L'objet d'étude de la série statistique s'appelle le **caractère** de la série ;
- Le caractère de la série peut prendre plusieurs **valeurs numériques** dans le cas d'une étude quantitative, sinon l'étude est qualitative ;
- Un caractère quantitatif peut être **discret** (s'il prend des valeurs isolés comme -1,2,4,5...) ou **continu** (s'il prend toute valeur dans un intervalle qu'on appellera classe) ;

À savoir faire 1 : Utiliser le vocabulaire de base

Dans une boulangerie, le poids affiché de la baguette est de 250 grammes.

Lors d'un contrôle, un agent a prélevé des baguettes et a relevé leurs poids (en gramme) ainsi que leur temps de cuisson (en minute). Les résultats sont donnés dans les tableaux suivant :

Poids (en g)	247	248	249	250	251	252	253
Nombre de baguettes	2	5	11	15	8	6	3

Temps de cuisson (en min)	[10, 11[[11, 12[[12, 13[[13, 14[[14, 15[[15, 16[[16, 17[
Nombre de baguettes	4	2	8	12	9	8	7

1. Quelle est la population étudiée?
2. Les caractères étudiés?
3. Quel caractère est quantitatif?
4. Quel caractère est qualitatif?
5. Quel caractère est discret?
6. Quel caractère est continu?

Partie II Effectifs et fréquences

A - Effectifs et effectifs cumulés

Définition 2 : Série statistique

L'**effectif** d'une valeur est le nombre d'individus ayant cette même valeur.

L'**effectif cumulé croissant** (E.C.C) d'une valeur est la somme des effectifs des valeurs inférieures ou égales à cette valeur.

Exemple :

En reprenant l'exemple de la série statistique d'étude du poids d'une baguette, on a :

Poids (en g)	247	248	249	250	251	252	253
Effectif	2	5	11	15	8	6	3
E.C.C	↓	↗	↓				50

On peut alors dire que l'effectif cumulé croissant de 249 est, c'est-à-dire nous avons baguettes qui ont un poids inférieur ou égal à 249.

De plus on remarque que l'effectif cumulé croissant de la plus grande valeur est l'effectif total.

B - Fréquences et fréquences cumulées

Définition 3 : Fréquence

La **fréquence** d'une valeur est le quotient entre l'effectif de cette valeur et l'effectif total :

$$\text{Fréquence} = \frac{\text{Effectif}}{\text{Effectif total}}$$

La **fréquence cumulée croissante** (F.C.C) d'une valeur est la somme des fréquences des valeurs inférieures ou égales à cette valeur.

Exemple :

Poids (en g)	247	248	249	250	251	252	253
Fréquence	0,04	0,1	0,22	0,3	0,16	0,12	0,06
F.C.C	↓	↗	↓				1

On peut dire que la fréquence cumulée croissante de 249 est, c'est-à-dire nous avons % des baguettes qui ont un poids inférieur ou égal à 249g.

La somme des fréquences, c'est-à-dire la fréquence cumulée croissante de la plus grande valeur, vaut 1.

Partie III Moyenne et écart-type

A - Moyenne (caractéristique de position)

A.1 - Série discrète

Dans cette partie on se donne la série statistique suivante :

Valeur	x_1	x_2	\dots	x_p
Effectif	n_1	n_2	\dots	n_p

On notera alors $N = n_1 + n_2 + \dots + n_p$ l'effectif total de la série statistique.

Propriété 1 : Moyenne d'une série discrète

La **moyenne** de cette série statistique est le nombre réel noté m ou \bar{x} tel que :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

✂ À savoir faire 2 : Déterminer une moyenne

Déterminer le poids moyen d'une baguette de pain de notre série statistique.

.....

.....

.....

.....

.....

.....

.....

Propriété 2 : Linéarité

- Si on multiplie toutes les valeurs d'une série statistique par un même nombre a , sans changer les effectifs, alors la moyenne est multipliée par ce nombre a ;
- Si on ajoute toutes les valeurs d'une série statistique un même nombre b , sans changer les effectifs, alors la moyenne est augmentée par ce nombre b .

On parle de linéarité de la moyenne.

✂ À savoir faire 3 : Utiliser la linéarité de la moyenne

Voici les notes d'une interrogation pour une classe :

Notes	8	9	10	11	12	13	14
Nombre de copies	2	5	3	5	8	7	5

1. Calculer la moyenne de classe.

.....

.....

2. Le professeur choisit d'augmenter toutes les notes d'un point. Quelle est la nouvelle moyenne de classe?

.....

3. Finalement, le professeur change d'avis et décide d'augmenter toutes les notes de 15%. Quelle est la nouvelle moyenne de classe?

.....

A.2 - Série en classe

Intéressons-nous à l'étude du temps de cuisson d'une baguette de notre série statistique.

Temps de cuisson (en min)	[10,11[[11,12[[12,13[[13,14[[14,15[[15,16[[16,17[
Nombre de baguettes	4	2	8	12	9	8	7

Pour calculer la moyenne d'une série regroupée en **classe**, on calcule une **valeur approchée de la moyenne** en remplaçant les valeurs de la série par le **centre de la classe** dans la formule de la moyenne d'une série discrète.

Exemple :

Temps de cuisson (en min)	[10,11[[11,12[[12,13[[13,14[[14,15[[15,16[[16,17[
Centre des classes	10,5	11,5	12,5	13,5	14,5	15,5	16,5
Nombre de baguettes	4	2	8	12	9	8	7

On a alors le temps de cuisson moyen d'une baguette est de :

$$\frac{10,5 \times 4 + 11,5 \times 2 + 12,5 \times 8 + 13,5 \times 12 + 14,5 \times 9 + 15,5 \times 8 + 16,5 \times 7}{50} = 13,94$$

Propriété 3 : Moyenne d'une série en classe

Après avoir donné le centre des classes de notre série statistique :

Classes	$[a_1, b_1[$	$[a_2, b_2[$...	$[a_p, b_p[$
Centre des classes	$c_1 = \frac{a_1 + b_1}{2}$	$c_2 = \frac{a_2 + b_2}{2}$...	$c_p = \frac{a_p + b_p}{2}$
Effectif	n_1	n_2	...	n_p

La **moyenne** de cette série statistique est le nombre réel, noté \bar{x} ou m , tel que :

$$\bar{x} = \frac{n_1 c_1 + n_2 c_2 + \dots + n_p c_p}{N}$$

✂ À savoir faire 4 : Moyenne d'une série en classe

L'étude des retards d'un train, en minutes, sur une certaine période nous est donnée dans le tableau suivant :

Classes	[0, 5[[5, 10[[10, 15[[15, 20[
Centre des classes
Effectif	4	24	15	2

Déterminer le retard moyen de ce train en minutes.

.....

.....

.....

.....

.....

.....

B - Écart-type et variance (caractéristique de dispersion)

Dans cette partie, on se donne la série statistique suivante :

Valeur (ou centre de la classe)	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p

On note $N = n_1 + n_2 + \dots + n_p$ l'effectif total.

Définition 4 : Variance et écart-type

La **variance** est la moyenne des carrés des écarts entre les valeurs de la série et la valeur moyenne.

La variance est le réel noté V tel que :

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

L'**écart-type** est le réel noté σ tel que $\sigma = \sqrt{V}$, est la racine carré de la variance.

💡 À retenir :

- La **variance** mesure la moyenne des carrés des distances entre de chacune des valeurs et la valeur moyenne m . Chaque carré étant pondéré son effectif;
- L'**écart-type** mesure la distance moyenne entre chacune des valeurs et la valeur moyenne m .

Exemple :

En reprenant la série statistique du poids d'une baguette.

Poids (en g)	247	248	249	250	251	252	253
Nombre de baguettes	2	5	11	15	8	6	3

Pour déterminer la variance et l'écart-type on apprendra à obtenir les informations souhaitées à partir d'une calculatrice. Voici des QR-Code vous menant vers un tutoriel pour cela suivant votre calculatrice :



NumWorks

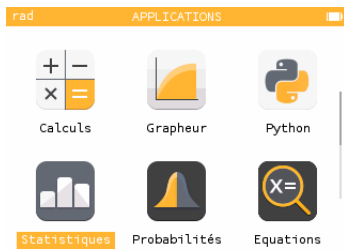


TI

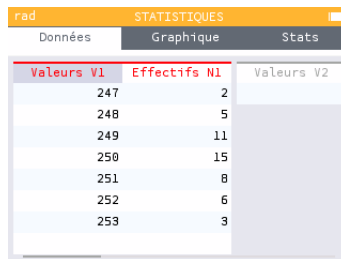


Casio

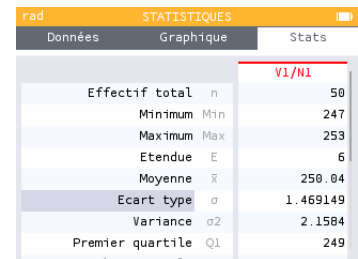
À l'aide de la calculatrice NumWorks voici ce que l'on obtient :



Étape 1 : On sélectionne **Statistiques**



Étape 2 : On complète les données



Étape 3 : On sélectionne **Stats**

On a alors $V \simeq 2,16$ et $\sigma \simeq 1,47$. Cela signifie que le poids des baguettes s'écarte en moyenne d'environ 1,47g du poids moyen qui est de 250,04g.

À savoir faire 5 : Déterminer un écart-type

En reprenant l'étude statistique des notes d'une classe :

Notes	8	9	10	11	12	13	14
Nombre de copies	2	5	3	5	8	7	5

Déterminer l'écart-type de la série statistique représentant les notes de la classe et interpréter ce résultat.

.....

.....

.....

.....

Information : L'écart-type mesure la dispersion des valeurs autour de la moyenne :

- ▶ plus il est grand, plus les valeurs sont dispersées et moins la moyenne représente de façon significative de la série;
- ▶ plus il est petit, plus les valeurs se concentrent autour de la moyenne.

C - Résumé d'une série statistique

On peut ainsi résumer une série statistique par le couple (\bar{x}, σ) .

La moyenne et l'écart-type prennent en compte toutes les valeurs de la série statistique et sont de ce fait influencées par les valeurs extrêmes.

On privilégie l'étude du couple moyenne-écart-type lorsque les valeurs extrêmes ne sont pas des anomalies mais traduisent réellement la dispersion des données.

Partie IV Médiane, quartiles, écart inter-quartile

A - Médiane

Définition 5 : Médiane

La **médiane** est la valeur « théorique » du caractère qui sépare la moitié des valeurs les plus basses, de la moitié des valeurs les plus hautes.

Méthode 1 : Trouver la médiane d'une série statistique

- ▶ **Étape 1** : On range toutes les valeurs par ordre **croissant** ;
- ▶ **Étape 2** : ou bien
 - l'**effectif total est impair** : la médiane est la valeur centrale ;
 - l'**effectif total est pair** : la médiane est la moyenne (ou demi-somme) des valeurs centrales.

À savoir faire 6 : Déterminer la médiane d'une série statistique

Déterminer la médiane de chacune des deux séries statistiques suivantes :

1.

Poids (en g)	247	248	249	250	251	252	253
Nombre de baguettes	2	5	11	15	8	6	3

.....

.....

.....

.....

.....

.....

.....

.....

2.

Notes	8	9	10	11	12	13	14
Nombre de copies	2	5	3	5	8	7	5

.....

.....

.....

.....

.....

.....

.....

B - Quartiles (caractéristique de position)

Définition 6 : Quartiles

- Le **premier quartile** est la plus petite valeur, notée Q_1 , de la série ordonnée telle qu'au moins un quart des valeurs de cette série sont inférieures ou égales à Q_1 ;
- Le **troisième quartile** est la plus petite valeur, notée Q_3 , de la série ordonnée telle qu'au moins trois quarts des valeurs de cette série sont inférieures ou égales à Q_3 .

Information : Au contraire de la médiane, le premier et le troisième quartile sont forcément des valeurs de la série.

Exemple :

En reprenant la série statistique du poids des baguettes :

Poids (en g)	247	248	249	250	251	252	253
Effectif	2	5	11	15	8	6	3
E.C.C	2	7	18	33	41	47	50

↑
8^{ème}, 9^{ème}, ..., 18^{ème}
valeurs de la série

↑
45^{ème} valeur de la série

- On a $\frac{50}{4} = 12,5$, comme Q_1 est la plus petite valeur ayant **au moins** un quart des valeurs de la série qui lui soit inférieure. Q_1 est donc la 13^{ème} valeur de la série. Ainsi $Q_1 = 249$;
- On a $3 \times \frac{50}{4} = 44,25$, comme Q_3 est la plus petite valeur ayant **au moins** trois quarts des valeurs de la série qui lui soit inférieure. Q_3 est donc la 45^{ème} valeur de la série. Ainsi $Q_3 = 252$;

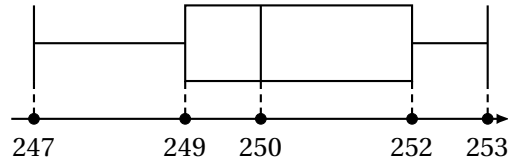
✂ À savoir faire 7 : Quartiles

Déterminer le premier et le troisième quartile de la série statistique des notes d'une classe donnée :

Aux extrémités, on indique la valeur maximale, la valeur minimale de la série. La boîte est définie par les premiers et troisièmes quartiles. La valeur médiane est repérée à l'intérieur de la boîte.

Exemple :

En reprenant la série statistique du poids des baguettes on a le diagramme en boîte suivant :



Information :

- La largeur de la boîte représente l'écart inter-quartile de la série ;
- La largeur du diagramme représente l'étendue de la série ;
- Le diagramme en boîte est aussi appelé boîte à moustaches : les moustaches correspondent aux intervalles $[\text{Min}, Q_1]$ et $[Q_3, \text{Max}]$.

À savoir faire 8 : Diagramme en boîte

Réaliser le diagramme en boîte de la série statistique des notes d'une classe.

E - Résumé d'une série statistique

Une série statistique peut être résumée par le couple (médiane ; écart interquartile).

Lorsque l'on utilise ce couple, la médiane et les quartiles ne dépendent pas des valeurs extrêmes de la série.

En effet, de légères variations des valeurs extrêmes n'entraînent pas de modification de la médiane ni des quartiles.

On dit alors que la médiane est un indicateur « robuste ».

On privilégie l'étude du couple (médiane ; écart interquartile) lorsque les valeurs extrêmes sont peu significatives ou moins importantes que les valeurs centrales de la série.

Partie V Exercices