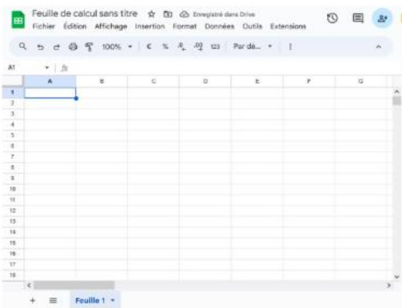
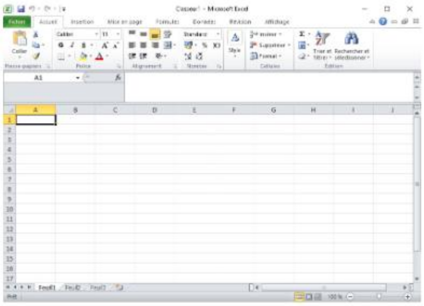
Chapitre 2 – Tableur et statistiques

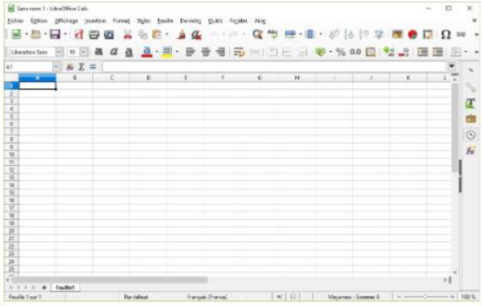
1. Utilisation du tableur

1a. Manipulation

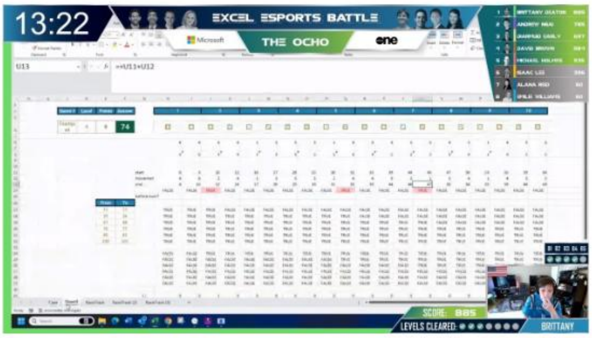
Un tableur est un programme informatique permettant de manipuler des feuilles de calculs, présentées en tableau. Les tableurs existent depuis les années 1980. Actuellement, les trois tableurs les plus utilisés sont Microsoft Excel, Google Sheets et LibreOffice Calc. Les captures d'écran de ce cours sont faits avec Excel, mais les fonctionnalités de base sont généralement les mêmes sur tous les tableurs.



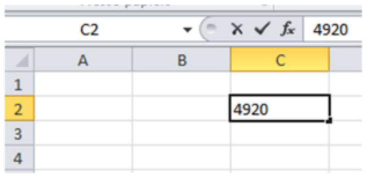




Les tableurs ont de nombreux usages : gestion de bases de données, production de graphiques, analyse statistique…
Ils présentent l'avantage d'organiser les calculs de façon spatiale, plutôt que temporelle comme un langage de programmation.
Il existe également un championnat du monde d'Excel !

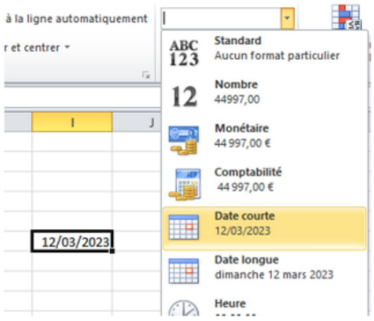


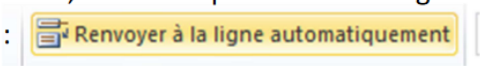
Dans un tableur, chaque casse (on dit plutôt **cellule**) est représentée par une lettre (colonne) et un nombre (ligne). Il est assez facile de se déplacer à l'aide des touches fléchées.



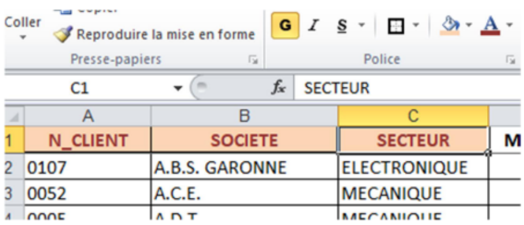
On peut entrer du contenu dans une cellule en tapant directement dedans, ou en utilisant la **barre de formule** pour avoir plus de place.

Chaque cellule peut contenir un **nombre**, du **texte**, une **date** ou d'autres types de données. Le tableur essaie généralement de détecter le type de données entrées, mais on peut changer ce type manuellement (pour les valeurs monétaires, par exemple).



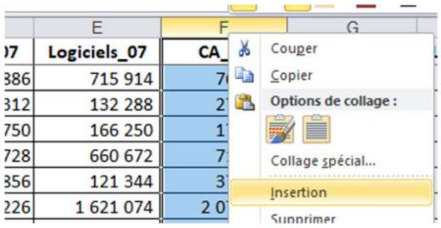
La touche Entrée permet de passer à la cellule en-dessous : généralement, on ne fait pas de saut de ligne dans le texte, mais on peut activer les sauts de ligne avec un bouton : 

Les cellules peuvent être mises en forme : couleurs, taille et surtout la **bordure**, très utile pour faire des en-têtes.

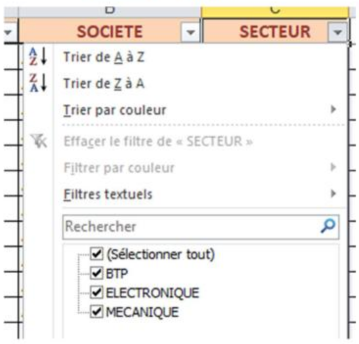


<u>Manipulation des lignes/colonnes</u>

Un **clic droit sur une colonne ou une ligne** permet **d'insérer une colonne/ligne supplémentaire** dans le tableau, à gauche/en haut de la colonne/ligne cliquée.



Lorsqu'une colonne est sélectionnée, un bouton « Trier » permet de la trier, mais surtout, le bouton « Filtrer »  fait apparaître des menus déroulants, qui permettent de ne conserver qu'une partie des données qu'on veut afficher.



1b. Formules

Une formule commence toujours par un signe = .
On peut y utiliser les noms des cellules.

Les formules sont utiles car elles permettent de préparer des calculs qui sont effectués automatiquement, même si on modifie la valeur des cellules. On peut citer :

• les 4 opérations + - * et /.
Par exemple, `=A2*A3` calcule le produit des cellules A2 et A3.

	A
1	
2	5
3	7
4	
5	=A2*A3

	A
1	
2	5
3	7
4	
5	35

Le tableur respecte les priorités opératoires, et on peut aussi se servir des cellules pour faire des calculs intermédiaires.

• `SOMME` pour calculer la somme de plusieurs cellules.
Par exemple, =`SOMME(A1:A4)` calculera A1+A2+A3+A4.
Il est possible d'utiliser la souris après avoir écrit =`SOMME(`
pour délimiter les cellules ajoutée.

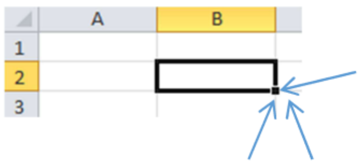
	A
1	4
2	14,2
3	8
4	6,001
5	
6	=SOMME(A1:A4)

	A
1	4
2	14,2
3	8
4	6,001
5	
6	32,201

• `MOYENNE` qui fonctionne comme `SOMME`.

Une formule peut être recopiée en s'adaptant aux cellules.

Lorsque des cellules sont sélectionnées, un petit carré apparaît en bas à droite de cette cellule. Il s'agit de la **poignée de recopie**.



En cliquant dessus, en gardant le bouton de la souris enfoncé et en la **tirant vers le bas ou vers la droite**, il est possible de recopier des cellules, voire des formules qui s'adapteront aux cellules recopiées.

Exemple 1 : commencez à entrer une suite de nombres...

	A	B
1		
2		1
3		3

... et utilisez la poignée pour poursuivre cette suite.

	A	B
1		
2		1
3		3
4		5
5		7
6		9
7		11
8		

Exemple 2 : utilisez une formule dans une cellule...

a	b	a×b
5	7	=B3*C3
10	4	
8	1	
-2	90	
6	2,5	

... puis utilisez la poignée pour reproduire cette formule.

a	b	a×b
5	7	35
10	4	40
8	1	8
-2	90	-180
6	2,5	15

Exemple 3 : calculez une moyenne...

	A	B	C	D	E	F
1		DS 1	DS 2	DS 3	DS 4	Moyenne
2	Élève 1	15	8	11	13	=MOYENNE(B2:E2)

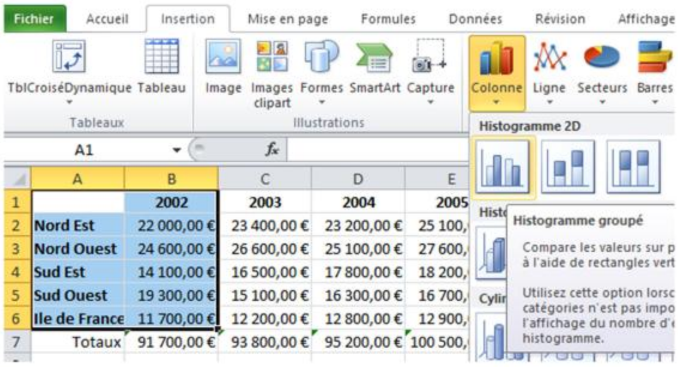
... et recopiez la formule vers le bas pour obtenir toutes les moyennes.

	A	B	C	D	E	F
1		DS 1	DS 2	DS 3	DS 4	Moyenne
2	Élève 1	15	8	11	13	11,75
3	Élève 2	19	11	20	18	17
4	Élève 3	7	14	12	6	9,75
5	Élève 4	5	15	12	11	10,75
6	Élève 5	17	13	15	18	15,75
7	Élève 6	8	12	11	15	11,5

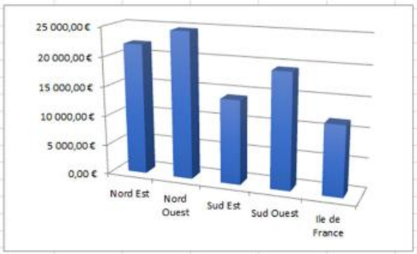
1c. Graphiques

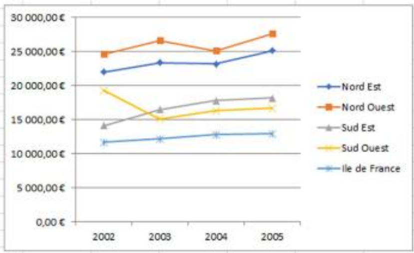
Les tableurs permettent de représenter des séries de valeurs sous forme de différents graphiques.

Pour créer un graphique, il faut d'abord sélectionner dans le tableau, les lignes/colonnes qui contiennent les données ou les légendes.
Ensuite, on peut insérer toutes sortes de graphiques et modifier leur apparence. Certains graphiques sont plus appropriés que d'autres : par exemple, une évolution sera mieux représentée par une courbe, alors qu'un histogramme ou un diagramme circulaire conviendra mieux à une comparaison.



	A	B	C	D	E
1		2002	2003	2004	2005
2	Nord Est	22 000,00 €	23 400,00 €	23 200,00 €	25 100,
3	Nord Ouest	24 600,00 €	26 600,00 €	25 100,00 €	27 600,
4	Sud Est	14 100,00 €	16 500,00 €	17 800,00 €	18 200,
5	Sud Ouest	19 300,00 €	15 100,00 €	16 300,00 €	16 700,
6	Ile de France	11 700,00 €	12 200,00 €	12 800,00 €	12 900,
7	Totaux	91 700,00 €	93 800,00 €	95 200,00 €	100 500,





2. Indicateurs

2a. Définition

Soit une série de n nombres réels x_1 ; x_2 ; … ; x_n .
La moyenne de cette série est :

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

Si de plus, on se donne une série de n coefficients $c_1; c_2; \dots ; c_n$,
la moyenne pondérée de la série est :

$$\frac{x_1 c_1 + x_2 c_2 + \cdots + x_n c_n}{c_1 + c_2 + \cdots + c_n}$$

Exemple 1
Dans un groupe de personnes, on considère le nombre de frères et sœurs. On relève les données statistiques dans le tableau suivant :

Nombre de frères et sœurs	0	1	2	3	4	5	6	7
Effectif	3	6	7	9	5	2	1	1

a. Donner l'effectif total de cette série.
b. Combien de personnes ont au moins trois frères et sœurs ?
c. Calculer le nombre moyen de frères et sœurs.

Exemple 2 Jason a obtenu les notes suivantes en mathématiques :

Note	13/20	7/10	9/10	4/5
Coefficient	4	2	1	4

1. Afin de calculer la moyenne, le professeur ramène les notes sur 20, en tenant compte de la base de notation pour recalculer le coefficient. Compléter le tableau avec les notes ramenées
2. Calculer alors la moyenne de Jason.

Note sur 20				
Coefficient				

Remarque : la moyenne est **linéaire**. Cela signifie que si on multiplie toutes les valeurs d'une série par le même nombre, ou si on ajoute le même nombre à toutes les valeurs d'une série, la moyenne est également multipliée ou additionnée à ce nombre.

3. a. Le professeur décide plutôt d'ajouter 3 points à toutes les notes ramenées sur 20.
Quelle est la nouvelle moyenne de Jason ?
b. Finalement, le professeur décide plutôt d'augmenter toutes les notes des élèves de 20%. Quelle est la nouvelle moyenne de Jason ?

Exemple 1
a. $3 + 6 + 7 + 9 + 5 + 2 + 1 + 1 = \mathbf{34}$ personnes ont été interrogées.
b. $9 + 5 + 2 + 1 + 1 = \mathbf{18}$ personnes ont au moins 3 frères & sœurs.
c. On applique la formule :

$$\frac{0 \times 3 + 1 \times 6 + 2 \times 7 + 3 \times 9 + 4 \times 5 + 5 \times 2 + 6 \times 1 + 7 \times 1}{34} = \frac{90}{34} \approx \mathbf{2,65}$$

Exemple 2

1. *On ramène toutes les notes sur 20, mais par exemple, si une note était initialement sur 10, alors on divise le coefficient par 2 pour tenir compte de l'importance deux fois moindre de la note.*

Note sur 20	**13**	**14**	**18**	**16**
Coefficient	**4**	**1**	**0,5**	**1**

2. La somme des coefficients est maintenant $4 + 1 + 0{,}5 + 1 = 6{,}5$. On calcule :

$$\frac{13 \times 4 + 14 \times 1 + 18 \times 0{,}5 + 16 \times 1}{6{,}5} = \mathbf{14}$$

3a. D'après la linéarité de la moyenne, si on ajoute 3 à toutes les notes ramenées sur 20, la moyenne augmente aussi de 3.
Ainsi, la nouvelle moyenne de Jason est de $14 + 3 = \mathbf{17}$.
3b. Augmenter une note de 20% revient à la multiplier par 1,2. D'après la linéarité de la moyenne, cela revient à multiplier aussi la moyenne par 1,2.
Ainsi, la nouvelle moyenne de Jason est $14 \times 1{,}2 = \mathbf{16{,}8}$.

2b. Médiane et quartiles

La médiane d'une série est la valeur qui la sépare en deux séries d'effectifs égaux.
Les quartiles permettent de séparer la série en 4.

Dans une série statistique **d'effectif n**, rangée dans l'ordre croissant :

- **le 1er quartile**, noté Q_1, est la plus petite valeur de la série telle qu'au moins 25% des valeurs lui soient inférieures ou égales. Il correspond à la valeur de **rang $\frac{1}{4}n$ arrondi à l'entier supérieur**.
- **la médiane** est le « nombre du milieu » : 50% des valeurs de la série sont inférieures à la médiane.

Si l'effectif **n est pair**, la médiane est la **moyenne de la $\frac{n}{2}$ – ième valeur et de la $\frac{n}{2}+1$– ième**.

Si l'effectif **n est impair**, la médiane est la **$\frac{n+1}{2}$ – ième valeur**.

- **le 3ème quartile**, noté Q_3, est la plus petite valeur de la série telle qu'au moins 75% des valeurs lui soient inférieures ou égales. Il correspond à la valeur de **rang $\frac{3}{4}n$ arrondi à l'entier supérieur**.
- l'écart interquartile est la différence $Q_3 - Q_1$.

Remarque : sur un tableur, on utilise `=QUARTILE(... ;1)` , `=MEDIANE(... )` et `=QUARTILE(... ;3)`

Exemple 1 On a relevé les performances, en mètres, obtenues par les élèves d'une classe au lancer du poids :
3,45 ; 5,2 ; 5,35 ; 4,3 ; 6,1 ; 4,28 ; 5,18 ; 4,9 ; 6,21 ; 5,36 ; 5,22 ; 4,9 ; 3,95 ; 4,72 ; 5,5 ; 6,13 ; 5,6
a. Quel est l'effectif ? Déterminer la médiane et les quartiles.
b. Une dernière personne lance un poids à une distance de 5,91 mètres.
Reprendre la question **a** avec cette nouvelle donnée.

Exemple 2 On reprend le tableau précédent dans lequel on a relevé le nombre de frères et sœurs de différentes personnes.
a. Déterminer les effectifs cumulés croissants.
b. En déduire la médiane et les quartiles.

Nombre de frères et sœurs	0	1	2	3	4	5	6	7
Effectif	3	6	7	9	5	2	1	1

Exemple 1

a. Pour trouver la médiane et les quartiles, il faut ranger la série dans l'ordre croissant : 3,45 < 3,95 < 4,28 < 4,3 < 4,72 < 4,9 = 4,9 < 5,18 < 5,2 < 5,22 < 5,35 < 5,36 < 5,5 < 5,6 < 6,1 < 6,13 < 6,21.

- L'effectif (le nombre de valeurs relevées) est **17**.
- Il est impair, donc **la médiane** correspond à la valeur de rang $\frac{17+1}{2} = 9$, c'est-à-dire **la 9ème valeur : 5,2**.
- Le **1er quartile** correspond à la valeur de rang $\frac{17}{4} = 4{,}25 \approx 5$ arrondi à l'entier supérieur, c'est-à-dure à **la 5ème valeur : 4,72**
- Le **3ème quartile** correspond à la valeur de rang $\frac{3\times 17}{4} = 12{,}75 \approx 13$ arrondi à l'entier supérieur, c'est-à-dure à **la 13ème valeur : 5,5**

b. L'ajout de cette nouvelle valeur change tout. On reclasse les valeurs dans l'ordre : 3,45 < 3,95 < 4,28 < 4,3 < 4,72 < 4,9 = 4,9 < 5,18 < 5,2 < 5,22 < 5,35 < 5,36 < 5,5 < 5,6 < **5,91** < 6,1 < 6,13 < 6,21.

• L'effectif est maintenant de **18**.

• Cet effectif est pair, donc **la médiane** correspond à la **moyenne des valeurs de rang 9 et 10**, c'est-à-dire $\frac{5{,}2+5{,}22}{2} = \mathbf{5{,}21}$

• Le **1er quartile** correspond toujours à la valeur de rang $\frac{18}{4} = 4{,}5 \approx 5$ arrondi à l'entier supérieur, c'est-à-dure à **la 5ème valeur : 4,72**

• Le **3ème quartile** correspond toujours à la valeur de rang $\frac{3\times18}{4} = 13{,}5 \approx 14$ arrondi à l'entier supérieur, c'est-à-dure à **la 14ème valeur : 5,6**

Exemple 2

Dans cet exercice, les effectifs sont présentés sous forme de tableau. Le rangement des réponses dans l'ordre croissant est donc :

$$0 \leq 0 \leq 0 \leq 1 \leq 1 \leq 1 \leq 1 \leq 1 \leq 1 \leq 2 \leq 2 \ldots$$

En effet, on compte d'abord 3 personnes qui ont répondu « 0 », puis 6 personnes qui ont répondu « 1 », etc.

Les effectifs cumulés croissants s'obtiennent en additionnant les effectifs de chaque valeur dans l'ordre croissant. Ils permettent de savoir quel est le rang des différentes valeurs.

a.

Frères & sœurs	0	1	2	3	4	5	6	7
Effectifs	3	6	7	9	5	2	1	1
Effectifs cumulés croissants	**3**	**9**	**16**	**25**	**30**	**32**	**33**	**34**

b. • L'effectif total est pair, donc la médiane correspond à la **moyenne des valeurs de rang 17 et 18**. D'après les effectifs cumulés croissants, la 17ème valeur et la 18ème valeur sont toutes les deux égales à 3, donc **la médiane est 3**.

• Le **1er quartile** correspond à la valeur de rang $\frac{34}{4} = 8{,}5 \approx 6$ arrondi à l'entier supérieur, c'est-à-dure à **la 9ème valeur : 1**

• Le **3ème quartile** correspond à la valeur de rang $\frac{3\times34}{4} = 25{,}5 \approx 26$ arrondi à l'entier supérieur, c'est-à-dure à **la 26ème valeur : 4**

2c. Écart-type

L'écart-type est un indicateur de dispersion : plus il est élevé, plus la série comporte des valeurs éloignées de sa moyenne.

Définition : Soit une série d'effectif n, dont les valeurs sont notées x_1 ; x_2; ... ; x_n, et de moyenne m.
L'**écart type** est un autre indicateur de description de cette série, dont la formule est :

$$s = \sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_n - m)^2}{n}}$$

Plus il est élevé, plus les valeurs de la série sont éloignées de la moyenne.

Exemple : Pour la série *10 ; 8 ; 13 ; 11 ; 8* l'écart type est d'environ 1,9
alors que pour la série *17 ; 4 ; 20 ; 1 ; 8* il est d'environ 7,3.
Ces deux séries ont pourtant la même moyenne.

Remarques :
- si la série suit une distribution normale (la distribution « en cloche », la plus fréquente), alors généralement, 95% de ses valeurs sont comprises dans l'intervalle $[m - 2s\ ;\ m + 2s]$.
- sur un tableur, on utilise la formule `=ECARTYPEP(...)`

2d. Regroupement par classes

Dans le cas de données regroupées par classes, on calcule la moyenne en se basant sur le centre des classes.

Exemple 1 Le tableau ci-contre donne la répartition des salaires mensuels des employés d'une petite entreprise.

a. Calculer une valeur approchée du salaire moyen d'un employé.

b. Dans quelle classe est situé le salaire médian ? Que signifie-t-il ?

Salaire (en €)	1 400 à 1 600	1 600 à 1 800	1 800 à 2 000	2 000 à 2 500	2 500 à 3 000
Fréquence (en %)	6,5	9,5	38,5	25,5	20

Exemple 2 Le tableau suivant a été obtenu après avoir relevé la vitesse de 70 véhicules sur un axe routier.

Vitesse (en km/h)	[50 ; 70[	[70 ; 80[	[80 ; 90[	[90 ; 100[	[100 ; 120[
Effectifs	15	36	46	64	70

a. En choisissant un centre approprié pour les classes, déterminer la vitesse moyenne.

b. Déterminer la classe médiane, ainsi que les quartiles.

Exemple 1 a. *Pour calculer la moyenne, sans plus d'informations sur la répartition des salaires à l'intérieur des classes, on prend le centre de chaque classe.*
Par exemple, la première classe (entre 1 400 et 1 600) a pour centre 1 500, la suivant a pour centre 1 700, la troisième 1 900, la quatrième 2 250
et la dernière 2 750.
Les fréquences étant données en pourcentage, leur somme est 100.

$$\frac{1\,500 \times 6{,}5 + 1\,700 \times 9{,}5 + 1\,900 \times 38{,}5 + 2\,250 \times 25{,}5 + 2\,750 \times 20}{100} = \mathbf{2\,114,25€}$$

b. Le salaire médian est le plus petit salaire supérieur à 50% des valeurs, ce qui correspond à la classe « 1 800 à 2 000 ».

Exemple 2 a. *On choisit le centre des classes comme dans l'exemple 1.*
L'effectif total est $15 + 36 + 46 + 64 + 70 = \mathbf{231}$.

$$\frac{60 \times 15 + 75 \times 36 + 85 \times 46 + 95 \times 64 + 110 \times 70}{231} = \frac{21\,290}{231} \approx \mathbf{92,16}\ \mathrm{km/h}$$

b. • L'effectif est impair, donc la médiane correspond à la classe de rang $\frac{231+1}{2} = 116$, c'est-à-dire **la 116ème valeur**, qui appartient à **la classe médiane** $[\mathbf{90,100}[$.

• Le **1er quartile** correspond à la classe de rang $\frac{231}{4} = 57{,}75 \approx 58$, c'est-à-dire à **la 58ème valeur** qui appartient à **la classe médiane** $[\mathbf{80};\mathbf{90}[$.

• Le **3ème quartile** correspond à la classe de rang $\frac{3\times 231}{4} = 173{,}25 \approx 174$ c'est-à-dire à **la 174ème valeur** qui appartient à **la classe médiane** $[\mathbf{100};\mathbf{120}[$.

3. Tableaux croisés

Un tableau croisé d'effectifs présente les valeurs de deux caractères d'une population, l'une en ligne et l'autre en colonne.

Définitions : dans un tableau croisé :

- l'**effectif marginal** d'une valeur x est le nombre d'individus présentant la valeur x.
- la **fréquence marginale** d'une valeur x est le quotient de l'**effectif marginal** de la valeur x sur l'**effectif total**.
- la **fréquence conditionnelle** de la valeur x sachant la valeur y est le quotient du **nombre d'individus correspondant à x et y** sur l'**effectif marginal** de la valeur x.

Exemple 1 Le tableau ci-contre donne la répartition des participants d'une course à pieds.

1. Quel est l'effectif marginal des coureurs français ?

2. Quelle est la fréquence marginale des 25 ans et plus ?

3. Quelle est la fréquence conditionnelle des moins de 25 ans parmi les étrangers ?

4. a. Quelle est la fréquence conditionnelle des 25 ans ou plus parmi les français ?

b. Quelle est la fréquence conditionnelle des français parmi les 25 ans ou plus ?

Âge / Nationalité	Moins de 25 ans	25 ans ou plus	Total
française	10	30	40
étrangère	20	140	160
Total	30	170	200

Exemple 2 Lors d'une enquête portant sur les 2 000 salariés d'une entreprise, on a obtenu les informations suivantes : 30% des salariés ont 40 ans ou plus, 40% des salariés de 40 ans ou plus sont des cadres, et 25% des salariés de moins de 40 ans sont des cadres.

a. Compléter le tableau ci-contre.

b. Déterminer la fréquence conditionnelle des moins de 40 ans parmi les cadres.

c. Déterminer la fréquence conditionnelle des 40 ans ou plus parmi les non-cadres.

	Moins de 40 ans	40 ans ou plus	Total
Cadres	...	...	...
Non-cadres	...	...	...
Total	...	...	2 000

Exemple 1

1. L'effectif marginal des coureurs français est 40.

2. L'effectif des 25 ans et plus est égal à 30, donc leur fréquence marginale est $\frac{30}{200} = 0{,}15 = \mathbf{15}\%$

3. On compte 20 moins de 25 ans parmi les 160 étrangers, donc leur fréquence conditionnelle est $\frac{20}{160} = 0{,}125 = \mathbf{12,5}\%$

4a. Parmi les 40 français, on compte 30 coureurs de 25 ans et plus, donc leur fréquence conditionnelle est $\frac{30}{40} = 0{,}75 = \mathbf{75}\%$

4b. Parmi les 170 coureurs de 25 ans et plus, on compte 30 français, donc leur fréquence conditionnelle est $\frac{30}{170} \approx 0{,}18 = \mathbf{18}\%$

Exemple 2

a. On calcule d'abord le nombre de salariés de 40 ans ou plus : $2\,000 \times 0{,}3 = 600$.
On en déduit qu'il y a 1 400 salariés de moins de 40 ans.
40% des salariés de 40 ans ou plus sont des cadres : $600 \times 0{,}4 = 240$.
On en déduit le nombre de salariés non-cadres de 40 ans ou plus : 360.
Enfin, 25% des salariés de moins de 40 ans sont des cadres : $1\,400 \times 0{,}25 = 350$.
On trouve tout le reste par addition et soustraction.

	Moins de 40 ans	40 ans ou plus	Total
Cadres	**350**	**240**	**590**
Non-cadres	**1 050**	**360**	**1 410**
Total	**1 400**	**600**	2 000

b. Parmi les 590 cadres, on compte 350 moins de 40 ans, donc la fréquence conditionnelle est $\frac{350}{590} \approx 0{,}59 \approx \mathbf{59}\%$

c. Parmi les 1 410 non-cadres, on compte 360 personnes de 40 ans ou plus, donc la fréquence conditionnelle est $\frac{360}{1\,410} \approx 0{,}26 \approx \mathbf{26}\%$