

## Chapitre 3 : Les tableaux de fréquence

### Tableaux de fréquences, formules classiques.

#### Tableaux de fréquences de variables qualitatives

Vous avez vu lors du premier cours qu'il était préférable de placer les données brutes et les statistiques descriptives dans des feuilles distinctes. Dans le même ordre d'idée, nous placerons les tableaux de fréquences dans une troisième feuille à part.

Ouvrez le classeur *HSB2.xlsx*, à la feuille *Données brutes*.

Combien cette base de données contient-elle de variables (sans compter la colonne i) ? Justifiez .....

**4 variables (car 4 colonnes)**

Combien de sujets sont encodés dans cette base de données ? Justifiez .....

**200 sujets (car 200 lignes)**

Pour répondre à cette dernière question, vous avez dû vous déplacer vers le bas du tableau, afin de compter de combien de lignes d'observations étaient encodées :

	A	B	C	D	E
195	194	F	privé	technique	Wallonie
196	195	M	privé	général	Wallonie
197	196	M	privé	technique	Wallonie
198	197	M	privé	technique	Wallonie
199	198	F	privé	technique	Wallonie
200	199	M	privé	technique	Wallonie
201	200	M	privé	technique	Wallonie
202					
203	Nombre de modalités:	2		3	3
204	N:				
205					
206	Min:				
207	Max:				
208					

Et si je vous demande maintenant, quelle est la variable encodée dans la colonne D ?.....

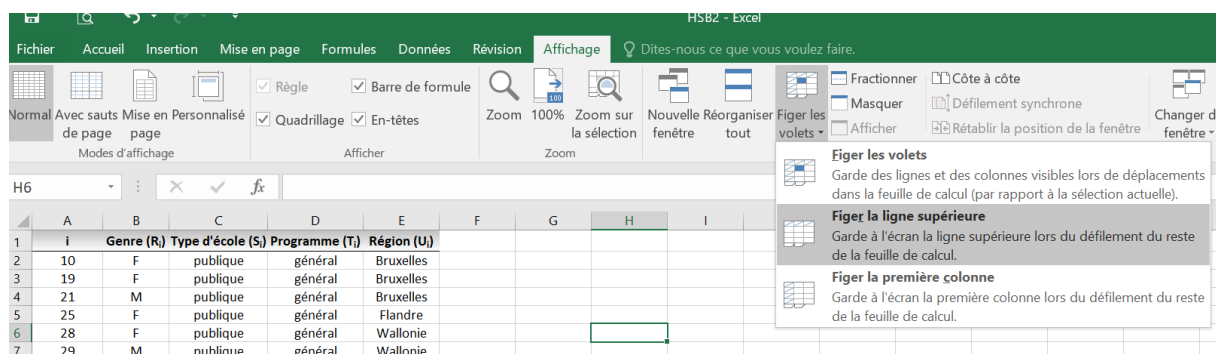
**La variable  $T_i$ .**

Sauf si vous avez vraiment une excellente mémoire visuelle, répondre à cette dernière question a impliqué que vous vous redéplaciez vers le haut du tableau.

Une difficulté avec des bases de données contenant un nombre important d'observations et qu'il n'est pas possible de les afficher dans leur entièreté à l'écran.

Il y a cependant moyen de faire en sorte que certaines parties importantes restent visibles en permanence. Par exemple, il est possible de « figer » la ligne contenant le nom des variables, de sorte qu'elle reste visible même lorsqu'on se déplace vers le bas du tableau.

Remontez tout en haut du tableau, de sorte à ce que la ligne contenant le nom des variables (soit la première ligne) soit visible à l'écran. Ensuite, dans l'onglet affichage, cliquez sur « Figer les volets », et ensuite, sur « Figer la ligne supérieure ».



Redescendez jusqu'en bas du tableau. Vous voyez, à présent, le nom des titres reste visible en permanence.

À la ligne 203, j'ai indiqué pour trois des variables de la base de données le nombre de modalités qu'elle contient. Quelles sont ces trois variables ?

..... **Genre( $R_i$ )**

..... **Programme( $T_i$ )**

..... **Région( $U_i$ )**

Nous allons nous intéresser de plus près à la variable  $Programme(T_i)$ . Notre objectif sera de compléter le tableau de fréquences de cette variable, qui se situe dans la feuille *Tableaux de fréquences*.

Dans la feuille *Données brutes*, commencez par trier vos données par ordre alphabétique, en fonction de la variable *Programme(T<sub>i</sub>)*. Cela permet de repérer plus facilement l'ensemble des modalités d'une variable (de toutes les énumérer en étant sûr(e) de n'en oublier aucune).

Quelles sont les modalités de la variable *Programme(T<sub>i</sub>)* ?

..... Général

..... Professionnel

..... Technique

Vous disposez à présent d'informations suffisantes pour commencer à créer le tableau de fréquences. Déplacez-vous dans la feuille *Tableau de fréquences*.

- À quoi correspond la colonne j ? .....  
C'est un indice qui attribue un numéro à chaque modalité de la variable possible. Pour rappel, on utilise l'indice i pour attribuer un numéro aux sujets, dans les tableaux de données brutes, et l'indice j pour attribuer un numéro aux modalités, dans les tableaux de fréquences.
- Pourquoi la valeur maximale de j est-elle 3 ? .....  
Car 3 modalités à la variable étudiée
- À quoi correspond t<sub>j</sub> ? .....  
Variable « nom des modalités de la variable T<sub>i</sub> ».

Remarquez que cette notation n'a pas été choisie au hasard. Elle consiste à écrire en minuscule la lettre qui décrit la variable *Programme(T<sub>i</sub>)* et de lui associer un petit indice « j ». Cette notation ayant été vue lors du cours théorique, vous devriez savoir ce qu'on va indiquer dans les cellules B4:B6. Faites-le. Faites bien attention de retranscrire parfaitement les valeurs de t<sub>j</sub>.

I4				
	A	B	C	D
1	Tableau de fréquences de la variable "Programme (T <sub>i</sub> )"			
2				
3	j	t <sub>j</sub>	n <sub>j</sub>	f <sub>j</sub>
4	1	général		
5	2	professionnel		
6	3	technique		
7	TOTAL			
8				

Les cellules C4:C6 contiendront respectivement les valeurs de  $n_1$ ,  $n_2$  et  $n_3$

Que représente  $n_1$ ,  $n_2$  et  $n_3$ ? .....

**Les fréquences absolues de chaque modalité de la variable Programme( $T_i$ ).**

Lors des cours antérieurs, nous avons vu que la formule qui permet de déterminer les valeurs de  $n_j$  est la suivante : **=NB.SI(série de données ; critère)**

Elle se traduit, en français par le « nombre de valeurs, au sein de la série de données spécifiée, qui correspondent au critère ».

Commencez par appliquer la formule dans la cellule C4 (vous l'étirerez ensuite dans les cellules C5:C6). Pour vous aider, répondez aux questions suivantes :

Quelle est la série de données que l'on veut étudier ? .....

**'Données brutes'!D2:D201**

De manière générale, à quoi correspond le critère ? .....

**à  $t_j$**

Cet élément est-il une variable ou une constante ? .....

**une variable**

Dans la cellule C4 en particulier, quelle est la valeur du critère ? .....

**« général » (Pour rappel, l'ordre des lignes n'a pas d'importance quand on représente le tableau de fréquences d'une variable nominale, il n'est donc pas impossible que vous ayez une autre réponse. Cependant, si vous avez suivi mon conseil consistant à trier les noms des modalités par ordre alphabétique, « général » est la réponse attendue).**

A-t-on intérêt à introduire directement la valeur de l'élément dans la cellule C4 ?

Justifiez : .....

**Non. L'élément étant une variable il prendra par définition une valeur différente pour chaque ligne du tableau de fréquence. En introduisant directement la valeur « général » dans la cellule E2, on se priverait de la possibilité d'étirer la formule dans les cellules C5:C6.**

N'oubliez pas que vous allez devoir étirer la formule introduite dans la cellule C4, dans les cellules C5 et C6. Dans tous les cas, la série de données que l'on veut étudier sera exactement la même. Que faudra-t-il faire pour le faire comprendre à Excel ? .....

**Introduire des \$ devant les numéros de lignes.**

Introduisez la formule appropriée dans la cellule C4.

Avant de continuer, nous allons vérifier que les valeurs introduites dans les cellules C4:C6 sont crédibles. Pour ce faire, dans la cellule C7, calculez la somme des cellules C4 :C6. Quelle valeur obtenez-vous ?.....

200

Retournez dans la feuille *Données brutes*. Dans la cellule D204, introduisez une formule qui vous permettra de connaître le nombre de sujets pour lequel on connaît le programme d'études (= variable  $T_i$ ).

**ATTENTION**, ce nombre ne sera pas forcément égal au nombre de sujets de la base de données (souvenez-vous qu'il peut y avoir des données manquantes). Vous avez vu au cours 2 une formule qui permet de déterminer ce nombre d'observations (voire la liste p.9 du cours 2). Quelle valeur obtenez-vous ?.....

200

Si vous avez obtenu la même réponse dans les cellules D204 de la feuille *Données brutes* et C7 de la feuille *Tableaux de fréquences*, c'est normalement que vous n'avez pas fait d'erreurs.

Passons à la suite. Les cellules D4:D6 contiendront respectivement les valeurs de  $f_1$ ,  $f_2$  et  $f_3$ .

Que représente  $f_1$ ,  $f_2$  et  $f_3$ ?.....

Les fréquences relatives de chaque modalité de la variable Programme( $T_i$ ).

Comment les calcule-t-on ?..... $n_j/n$

**Un indice** : la réponse à la question précédente implique d'appliquer l'un des opérateurs mathématiques vu aux cours antérieurs : la division.

L'objectif sera d'appliquer ce calcul dans les cellules D4:D6. Commencez par introduire la formule adéquate dans la cellule D4. Étirez la formule dans les cellules D5 et D6. **Attention** : de même que pour la colonne  $n_j$ , vérifiez bien que la série de données spécifiée dans **chaque** cellule soit correcte.

Avant de continuer, nous allons vérifier que les valeurs introduites dans les cellules D4:D6 sont crédibles. Pour ce faire, dans la cellule D7, calculez la somme des cellules D4 :D6.

Trouvez-vous « 1 » comme réponse ? Si oui, c'est que tout va bien ! Essayez de justifier pourquoi : .....

### Exercices

Toujours dans la feuille *Tableaux de fréquences* du fichier *HSB2.xlsx*, complétez le tableau de fréquence de la variable *Genre(R<sub>i</sub>)*.

1. L'une des colonnes n'a pas été nommée. À quoi correspond-elle ?  
.....

#### Aux modalités de la variable *Genre(R<sub>i</sub>)*.

2. Quel serait le nom adéquat pour cette variable ? ..... **r<sub>i</sub>**
3. Entrez en *C11* une formule que vous pourrez étirer dans la cellule *C12*. Vérifiez ensuite la crédibilité de vos réponses en calculant, dans la cellule *C13*, la somme des cellules *C11:C12*.
4. Entrez en *D11* une formule que vous pourrez étirer dans la cellule *D12*. Vérifiez ensuite la crédibilité de vos réponses en calculant, dans la cellule *D13*, la somme des cellules *D11:D12*.

### Tableaux croisés dynamiques.

Créer un tableau de fréquences manuellement, comme vous venez de le faire pour les variables *Programmes(T<sub>i</sub>)* et *Genre(R<sub>i</sub>)* n'est pas particulièrement fastidieux, essentiellement parce que ces variables n'ont que peu de modalités. Cela peut devenir nettement plus pénible dans certains cas :

- Lorsque l'on souhaite créer un tableau de fréquences pour une variable qui contient beaucoup de modalités ;
- Lorsque la base de données est extrêmement volumineuse (plusieurs centaines, voire plusieurs milliers d'observations).

Heureusement, Excel fournit un outil très puissant pour pallier ce problème : les **tableaux croisés dynamiques**. Ils permettent d'effectuer des calculs relativement complexes en quelques clics, à condition de respecter certaines contraintes :

1. La première ligne de la base de données brutes doit obligatoirement contenir les noms des variables.
2. Il ne peut y avoir de ligne (ou colonne) vide dans la base de données brutes.

3. Chaque colonne ne peut comporter qu'un seul type de données (on ne peut y avoir, dans une seule colonne, certaines cellules de type texte, et d'autres cellules de type nombre).

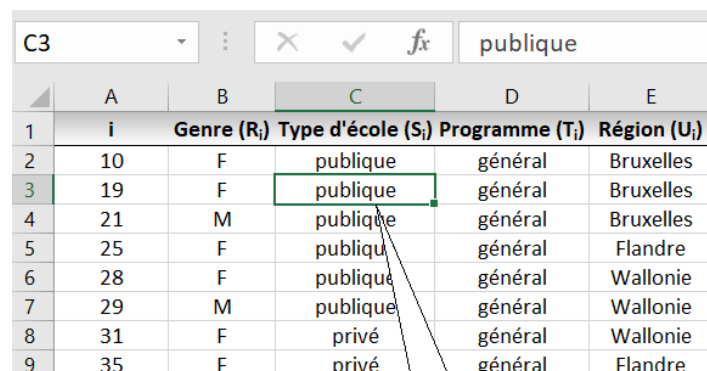
Bien que le tableau croisé dynamique puisse servir à diverses analyses, nous allons l'illustrer par la création de tableaux de fréquences.

Dans les exercices ci-dessous, vous serez amené(e)s à constituer des tableaux de fréquences pour divers types de variables :

- Variables qualitatives nominales
- Variables quantitatives discrètes
- Variables quantitatives continues

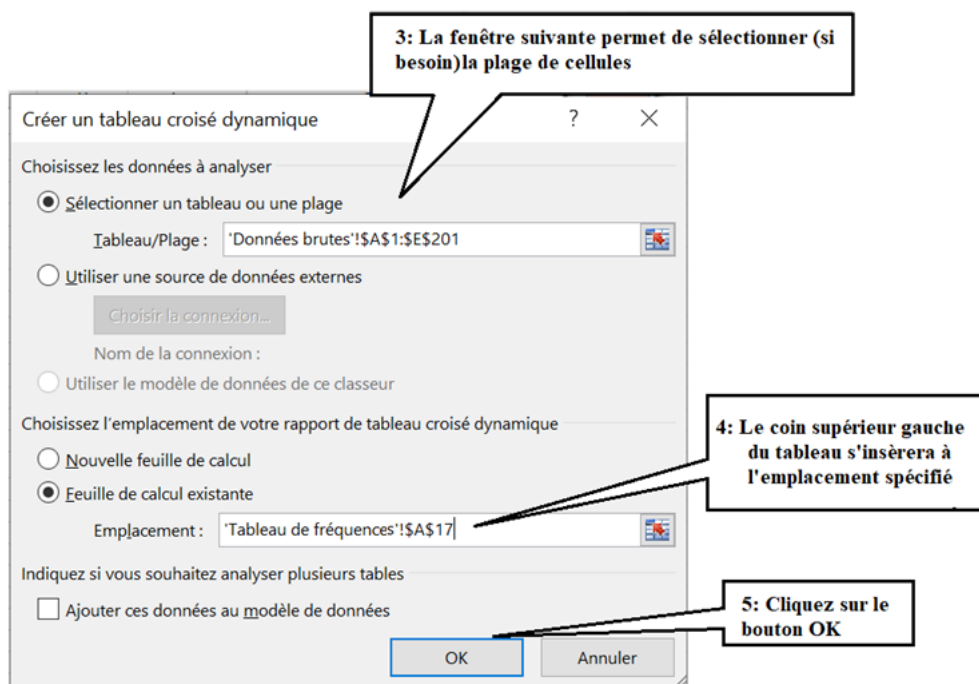
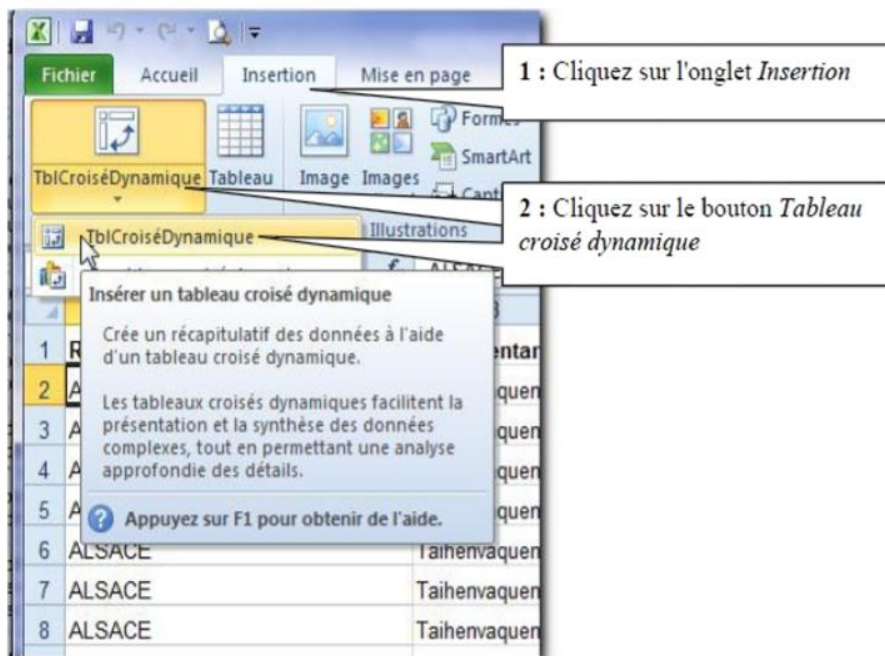
### Tableaux de fréquences de variables qualitatives

Rendez-vous dans la feuille *Données brutes* du fichier *HSB2.xlsx*, et activez n'importe quelle cellule du tableau. Nous allons créer un tableau de fréquences pour la variable *Région*( $U_i$ ). Nous ferons en sorte que le coin supérieur gauche du tableau se situe dans la cellule *A17* de la feuille *Tableau de fréquences*



	A	B	C	D	E
1	i	Genre ( $R_i$ )	Type d'école ( $S_i$ )	Programme ( $T_i$ )	Région ( $U_i$ )
2	10	F	publique	général	Bruxelles
3	19	F	publique	général	Bruxelles
4	21	M	publique	général	Bruxelles
5	25	F	publique	général	Flandre
6	28	F	publique	général	Wallonie
7	29	M	publique	général	Wallonie
8	31	F	privé	général	Wallonie
9	35	F	privé	général	Flandre

Activez une cellule du  
tableau (n'importe  
laquelle)





La fenêtre est prête pour le tableau croisé

Le volet permettra de sélectionner les champs de notre tableau

6: cochez le champ suivant (la variable pour laquelle on veut constituer le tableau de fréquences)

Notez que normalement, le champ coché se place automatiquement dans la zone « Lignes » du tableau croisé dynamique. S'il en est autrement, déplacez le champ de cette zone.

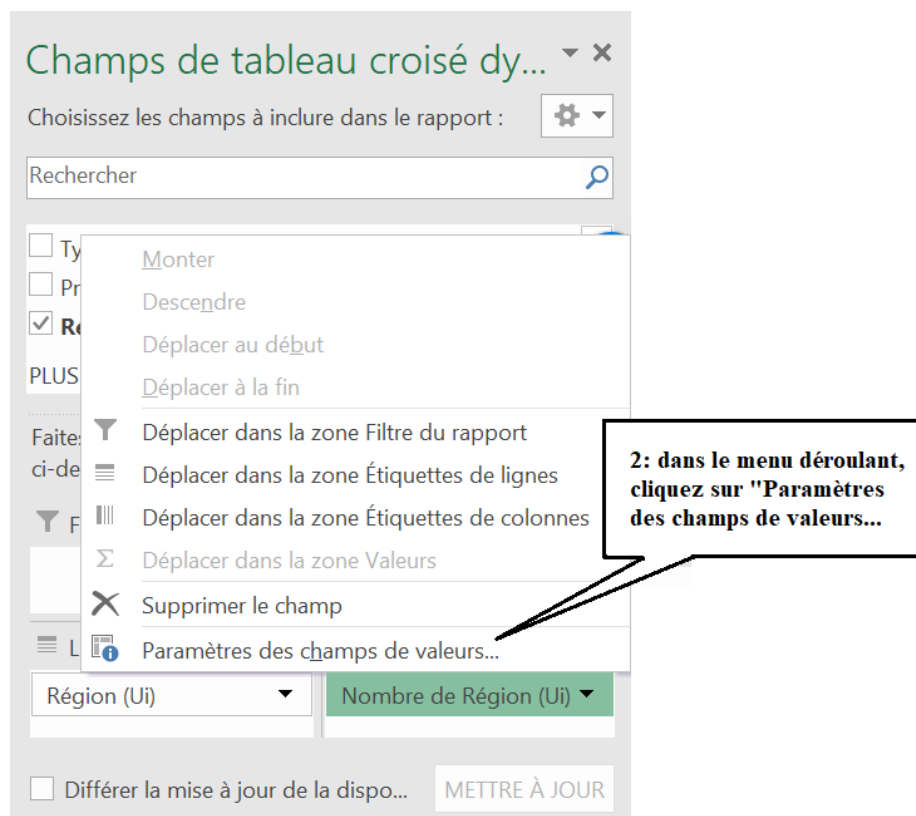
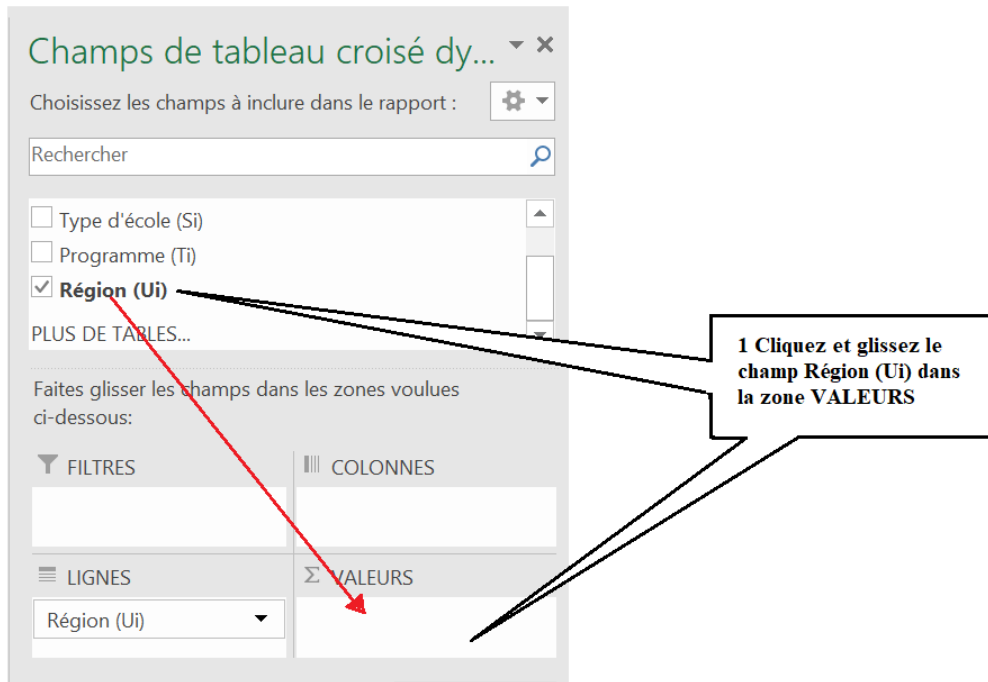
Résultat :

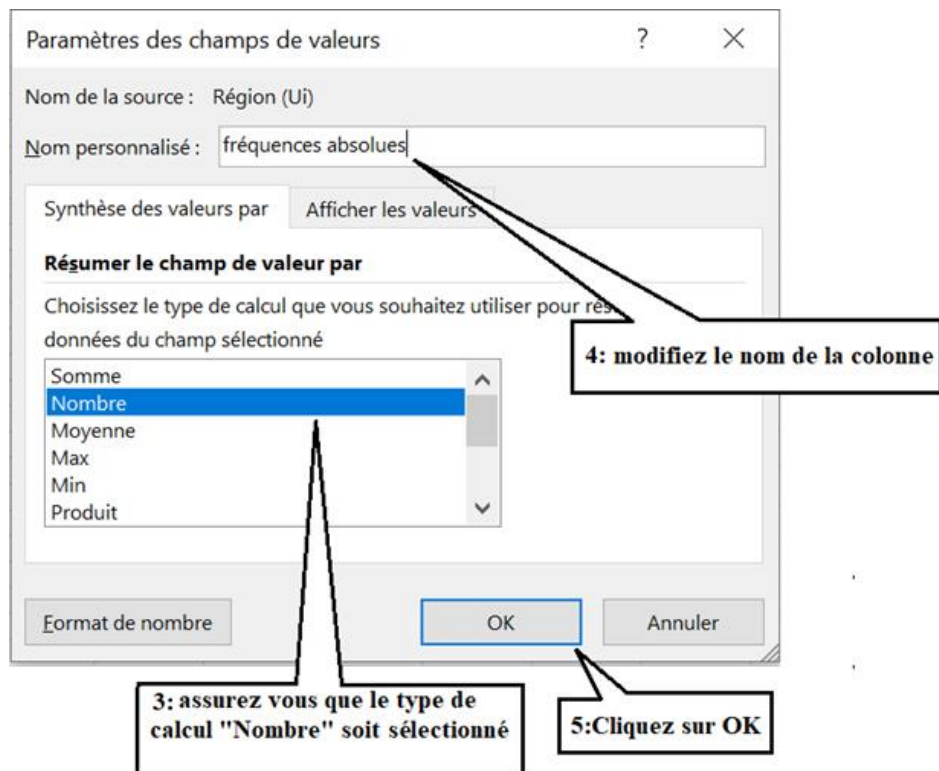
15				
16	Tableau de fréquences de la variable "Région (U <sub>i</sub> )"			
17	Étiquettes de lignes			
18	Bruxelles			
19	Flandre			
20	Wallonie			
21	Total général			
22				
23				

## Ajouter des colonnes au tableau

Chaque colonne du tableau de fréquences servira à calculer un type de fréquences pour les modalités de la variable *Région*( $U_i$ ).

- Première colonne :  $n_j$

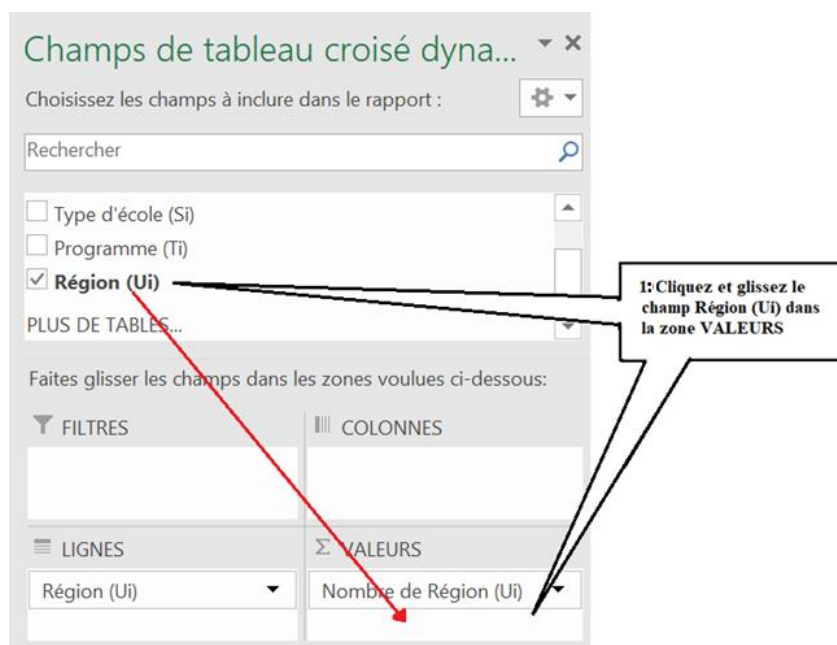


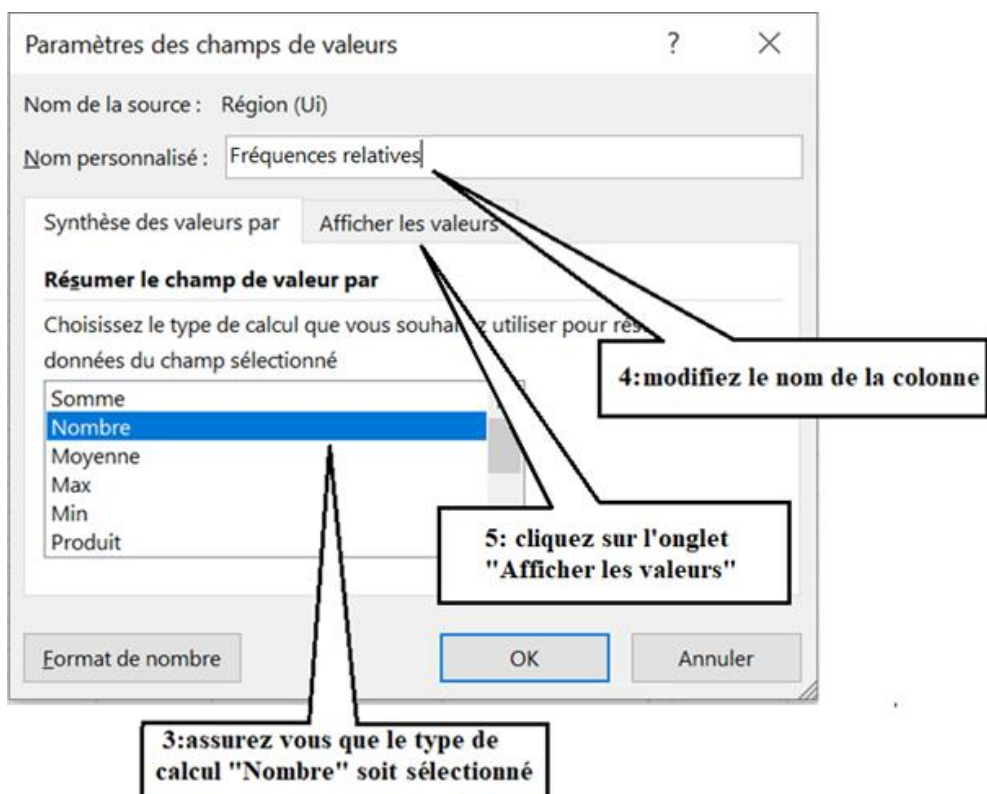
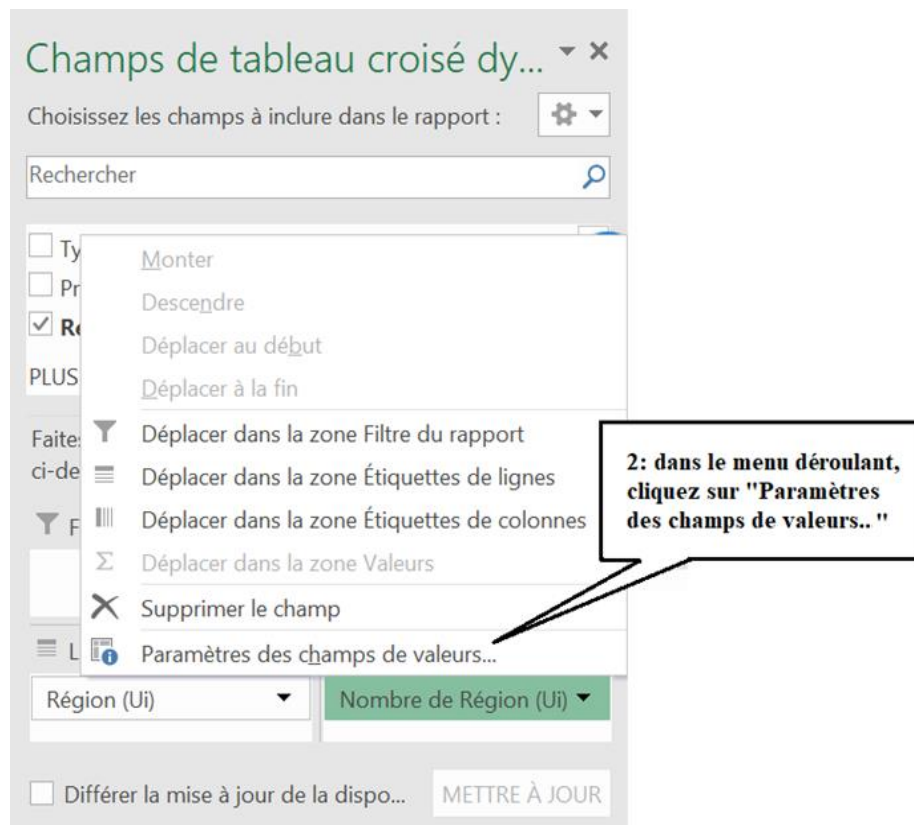


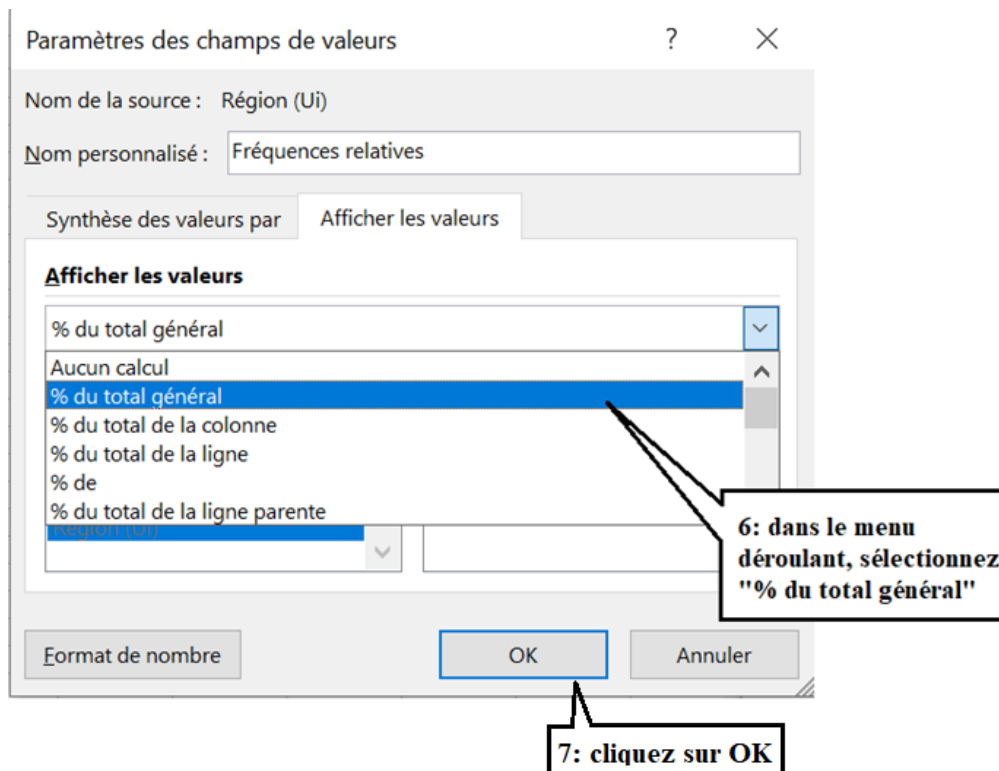
Résultat attendu :

Tableau de fréquences de la variable "Région (U <sub>i</sub> )"	
Étiquettes de lignes	Fréquences absolues
Bruxelles	91
Flandre	38
Wallonie	71
Total général	200

- Deuxième colonne :  $f_j$







Résultat attendu :

15			
16	Tableau de fréquences de la variable "Région (U <sub>i</sub> )"		
17	Étiquettes de lignes	Fréquences absolues	Fréquences relatives
18	Bruxelles	91	45,50%
19	Flandre	38	19,00%
20	Wallonie	71	35,50%
21	<b>Total général</b>	<b>200</b>	<b>100,00%</b>
22			

Votre tableau de fréquences est terminé.

### Tableaux de fréquences de variables quantitatives discrètes

Lorsque l'on souhaite réaliser un tableau de fréquences pour une variable quantitative discrète, bien que le principe soit assez similaire à celui des variables qualitatives, il y a malgré tout quelques différences :

- Il est possible de calculer des types de fréquences supplémentaires : les fréquences **absolues cumulées** et les fréquences **relatives cumulées** ;
- Il est impératif de maintenir un intervalle régulier entre les catégories (essentiellement pour pouvoir ultérieurement créer des graphiques, sur base du tableau de fréquences ; même si nous ne le ferons pas dans le cadre de ce cours, acquérir de bons réflexes est une bonne chose).

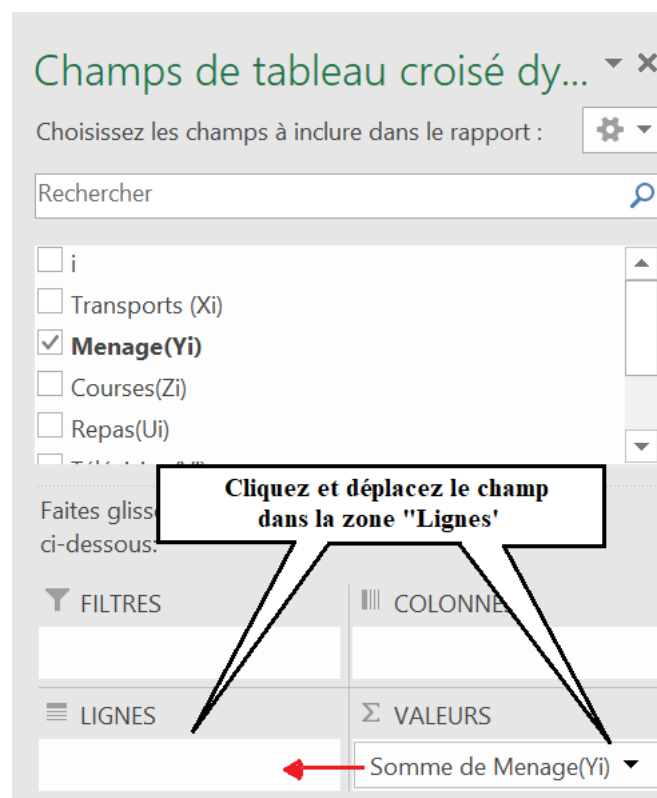
Ouvrez le fichier *temps.activités.xlsx*. Pour 25 sujets, on étudie le temps (en heures, réponse systématiquement arrondie à l'unité la plus proche) passé quotidiennement :

- Dans les transports en commun ( $X_i$ )
- À faire du ménage ( $Y_i$ )
- À faire les courses ( $Z_i$ )
- À manger et préparer à manger ( $U_i$ )
- À regarder la télévision ( $V_i$ )
- À faire du sport ( $W_i$ )

Nous allons créer un tableau de fréquences pour la variable *Menage*( $Y_i$ ).

Insérez un nouveau tableau croisé dynamique, tout en vous assurant que le coin supérieur gauche du tableau croisé se trouve dans la cellule A2 de la feuille *Tableaux de fréquences*.

**ATTENTION** : une fois le tableau créé, sélectionnez la variable *Menage*( $Y_i$ ). Les modalités de cette variable étant des chiffres, et non du texte comme précédemment, il est vraisemblable que par défaut, la variable ne se place pas dans la bonne zone. Pour délimiter les différentes lignes d'un tableau de fréquences (= les modalités de la variable étudiée), il faut nécessairement placer la variable dans la zone « Ligne ». Déplacez-là, si nécessaire.



Reproduisez ensuite les étapes vues précédemment pour calculer les fréquences absolues et relatives.

Résultat attendu :

	A	B	C	D	E
1	Tableau de fréquences de la variable "Menage(Y <sub>i</sub> )"				
2	Étiquettes de lignes ▼	Fréquences absolues	Fréquences relatives		
3	1	6	24,00%		
4	2	5	20,00%		
5	3	3	12,00%		
6	4	5	20,00%		
7	5	6	24,00%		
8	Total général	25	100,00%		
9					

Notez que l'intervalle entre les étiquettes est régulier. Ceci est une condition **nécessaire** lorsqu'on travaille avec des variables quantitatives discrètes, comme rappelé précédemment !

Nous allons à présent ajouter les colonnes propres aux fréquences absolues cumulées et aux fréquences relatives cumulées.

- Fréquences absolues cumulées :

### Champs de tableau croisé dy...

Choisissez les champs à inclure dans le rapport :

Rechercher

- ☐ i
- ☐ Transports (Xi)
- ☒ **Menage(Yi)**

Faites glisser les champs dans les zones voulues ci-dessous:

FILTRES

LIGNES

Menage(Yi)

COLONNES

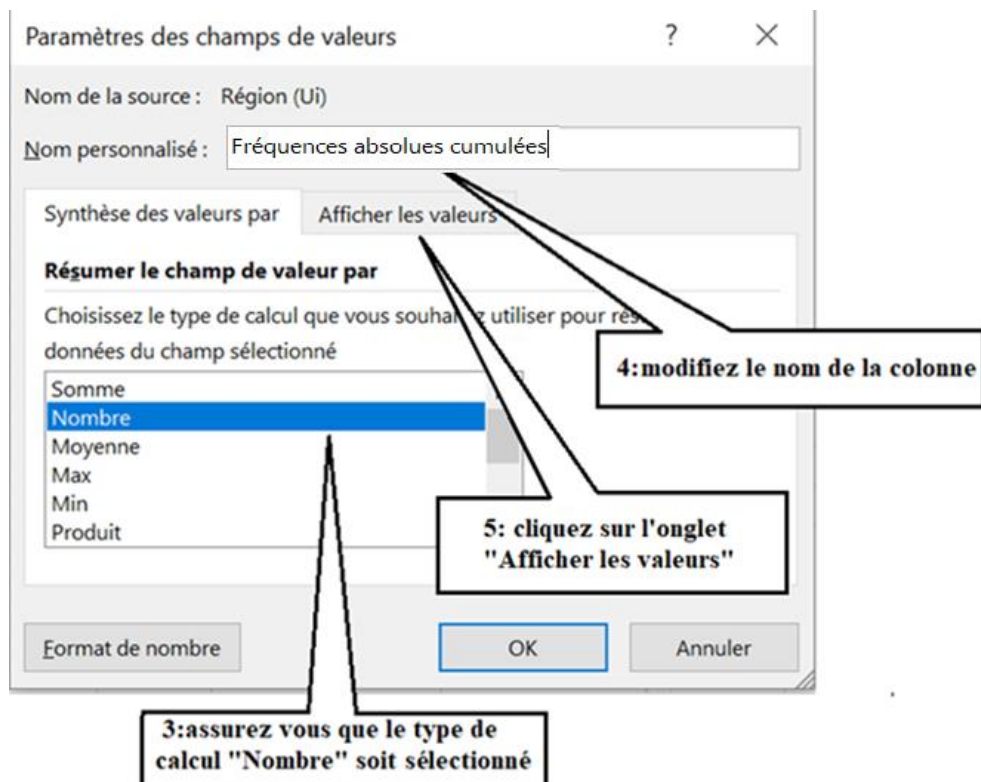
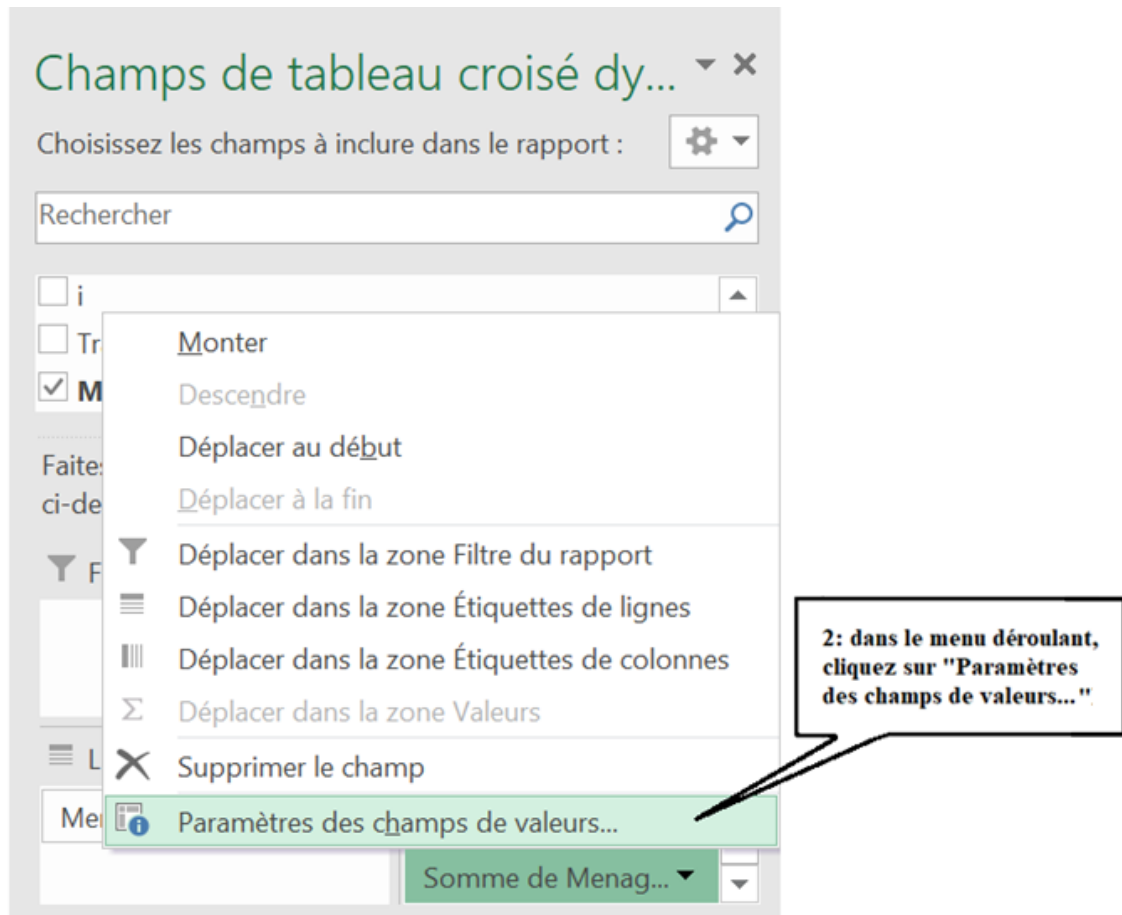
Σ Valeurs

VALEURS

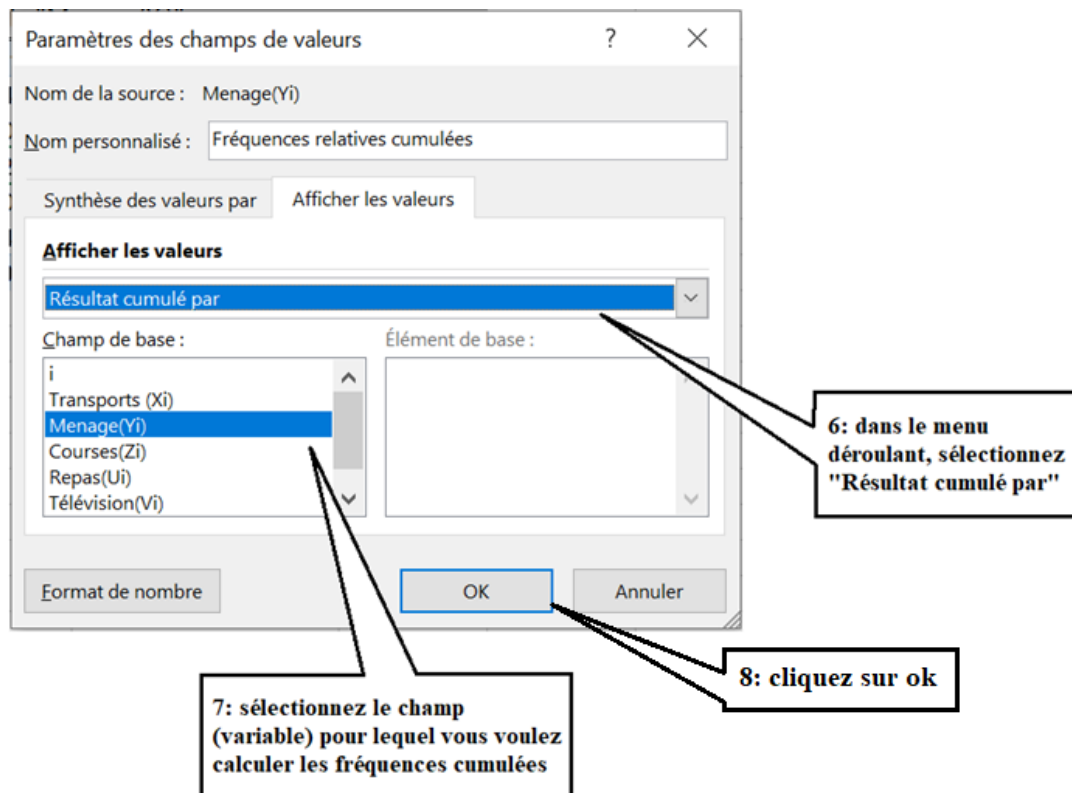
Fréquences relatives ▼

Somme de Menag... ▼

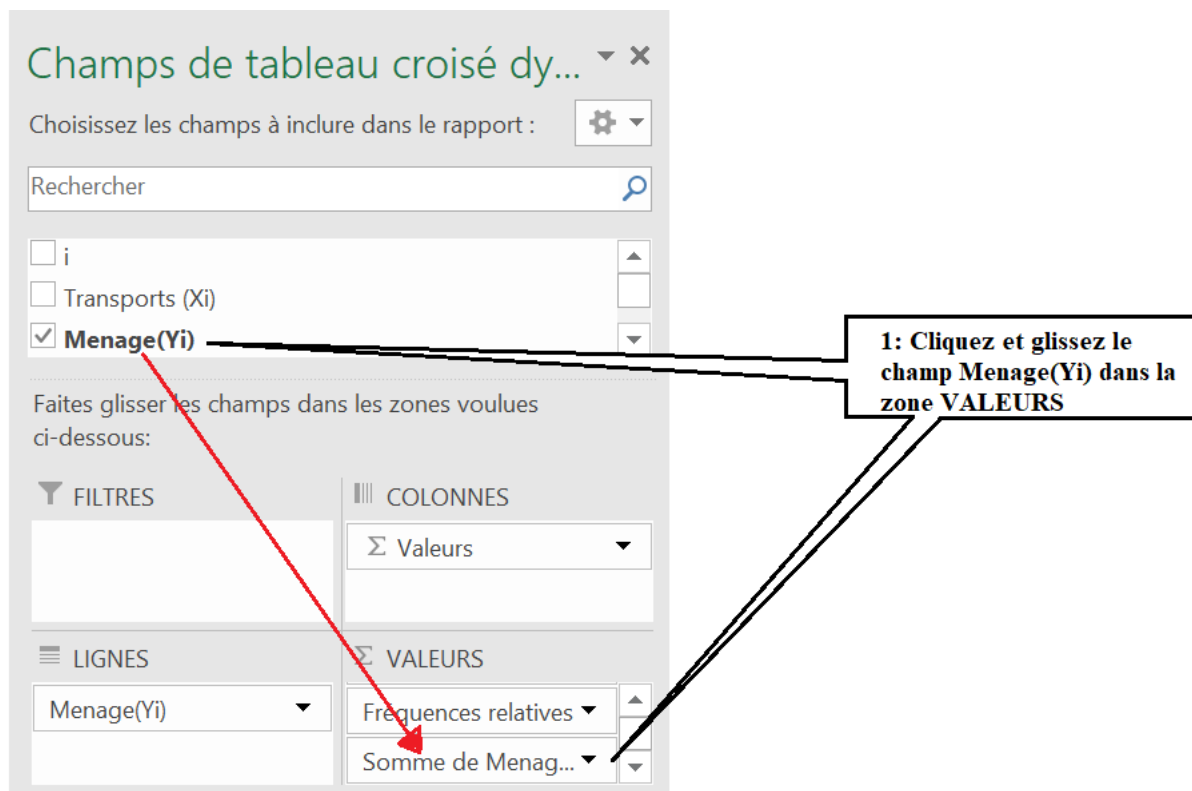
1: Cliquez et glissez le champ Menage(Yi) dans la zone VALEURS

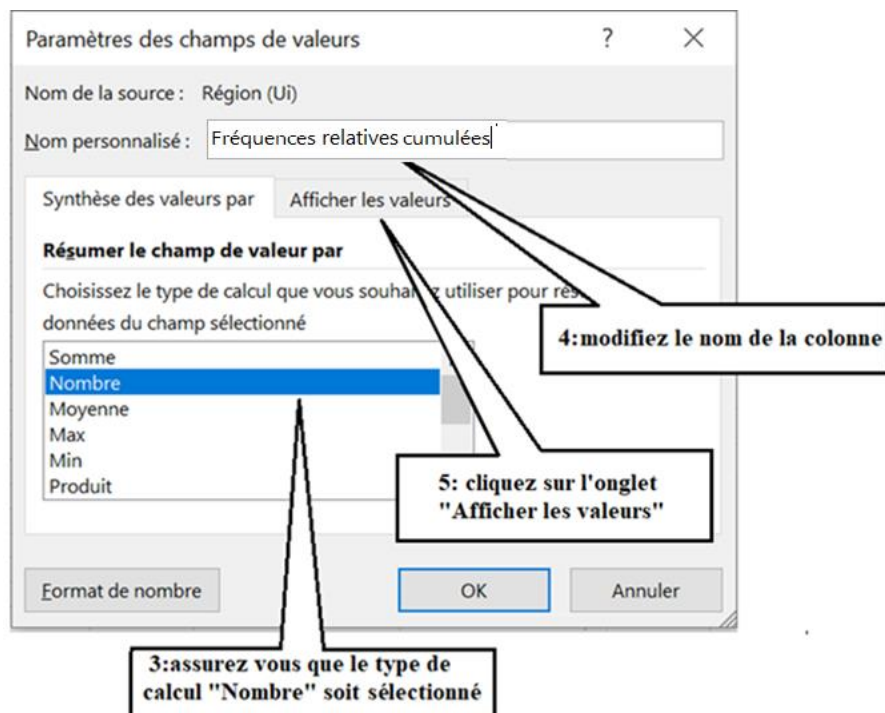
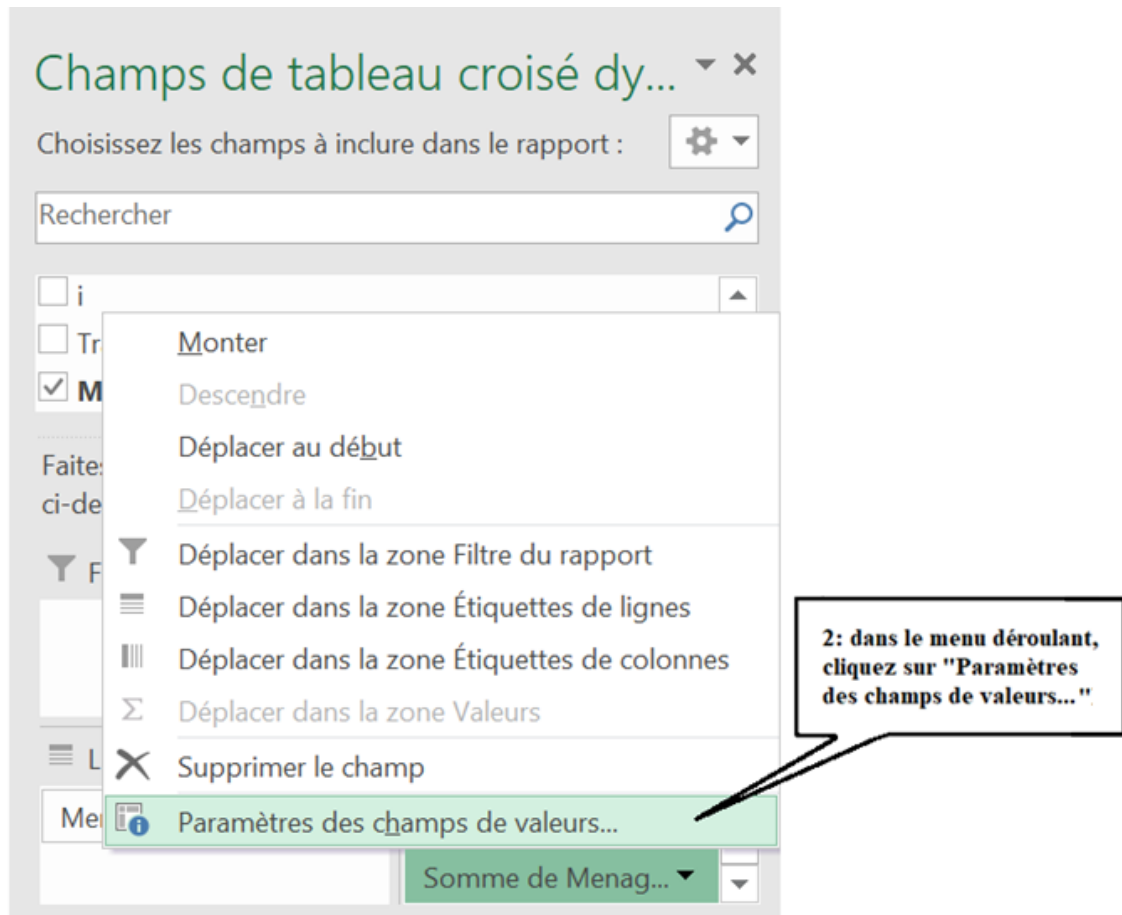


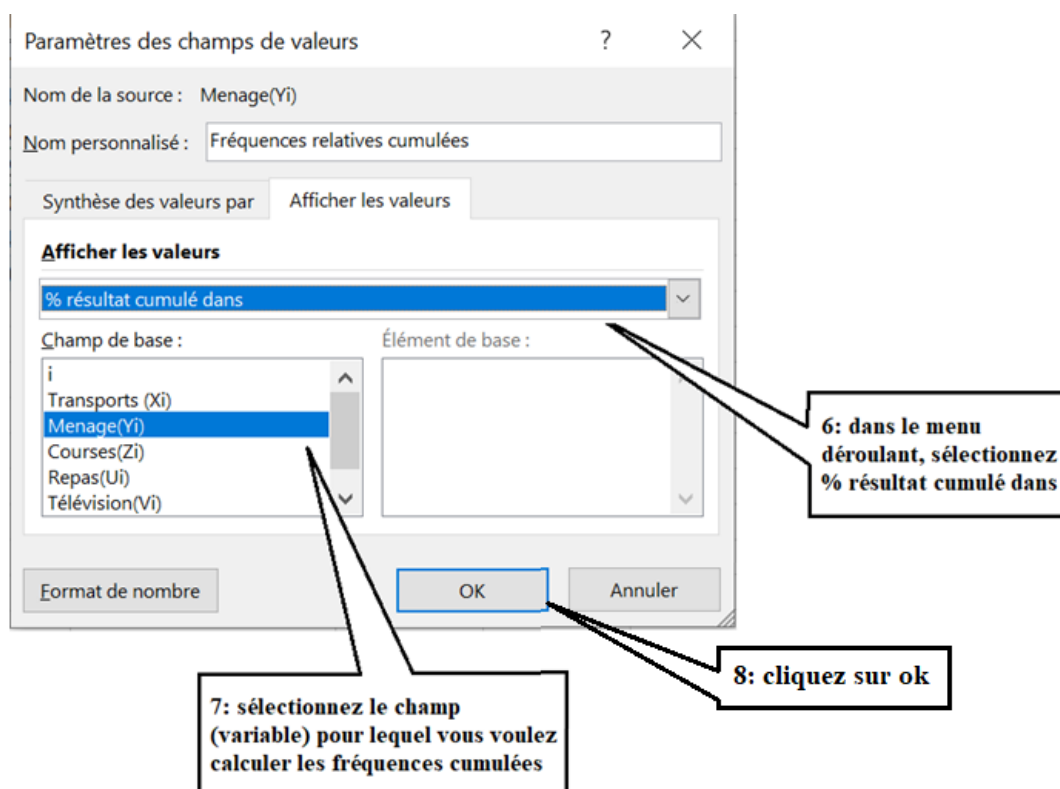




- Fréquences relatives cumulées :







Résultat attendu :

	A	B	C	D	E
1	Tableau de fréquences de la variable "Menage(Y <sub>i</sub> )"				
2	Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées
3	1	6	24,00%	6	24,00%
4	2	5	20,00%	11	44,00%
5	3	3	12,00%	14	56,00%
6	4	5	20,00%	19	76,00%
7	5	6	24,00%	25	100,00%
8	Total général	25	100,00%		

Vous allez à présent créer un tableau de fréquences pour la variable *Courses(Z<sub>i</sub>)*.

Insérez un nouveau tableau croisé dynamique, tout en vous assurant que le coin supérieur gauche du tableau croisé se trouve dans la cellule A11 de la feuille *Tableaux de fréquences*.

### Tableaux de fréquences de variables quantitatives continues

Ouvrez le fichier *Taille\_age.xlsx*. Rendez-vous dans la feuille *Données brutes* du fichier et activez n'importe quelle cellule du tableau. Nous allons créer un tableau de fréquences pour la variable *Région(U<sub>i</sub>)*.

Insérez un tableau de fréquences, en vous assurant que le coin supérieur gauche se trouve dans la cellule A2 de la feuille *Tableaux de fréquences*. Cochez le champ *taille(X<sub>i</sub>)*. Voici ce que vous devez obtenir :

A2						
	A	B	C	D	E	F
1	Tableau de fréquences de la variable "Taille(X <sub>i</sub> )"					
2	Étiquettes de lignes					
3	135,74					
4	135,86					
5	136,13					
6	138,25					
7	138,98					
8	141,03					
9	142,19					
10	142,33					
11	144,51					
12	144,79					
13	145,16					
14	145,24					
15	146,80					
16	147,21					
17	147,40					
18	147,71					
19	147,72					

- Quelle est la première modalité de la variable Taille(X<sub>i</sub>) ? .....  
**135,74**
- Quelle est la dernière modalité de la variable Taille(X<sub>i</sub>) ? .....  
**198,11**
- Combien de valeurs différentes sont envisagées dans le tableau de fréquences ? .....  
**300**

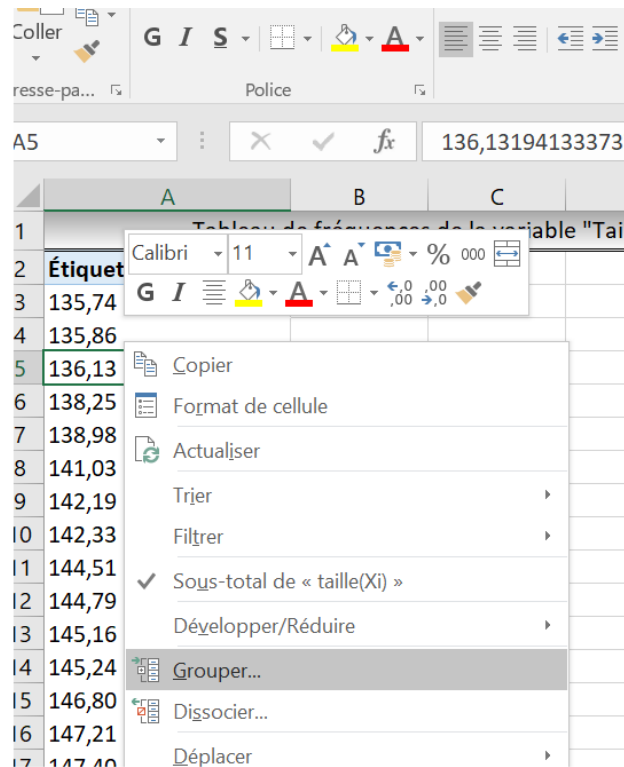
Vous devez avoir trouvé autant de lignes qu'il n'y a de sujets dans le tableau de données brutes. Cela est dû au fait que la variable *Taille(X<sub>i</sub>)* est une variable continue, et que les réponses n'ont pas été arrondies à l'unité. Si on laisse le tableau en l'état, quelle sera la fréquence absolue associée à chaque ligne du tableau de fréquence ? .....

**1. Cela signifie que le tableau de fréquences ne résume pas du tout le tableau de données brutes.**

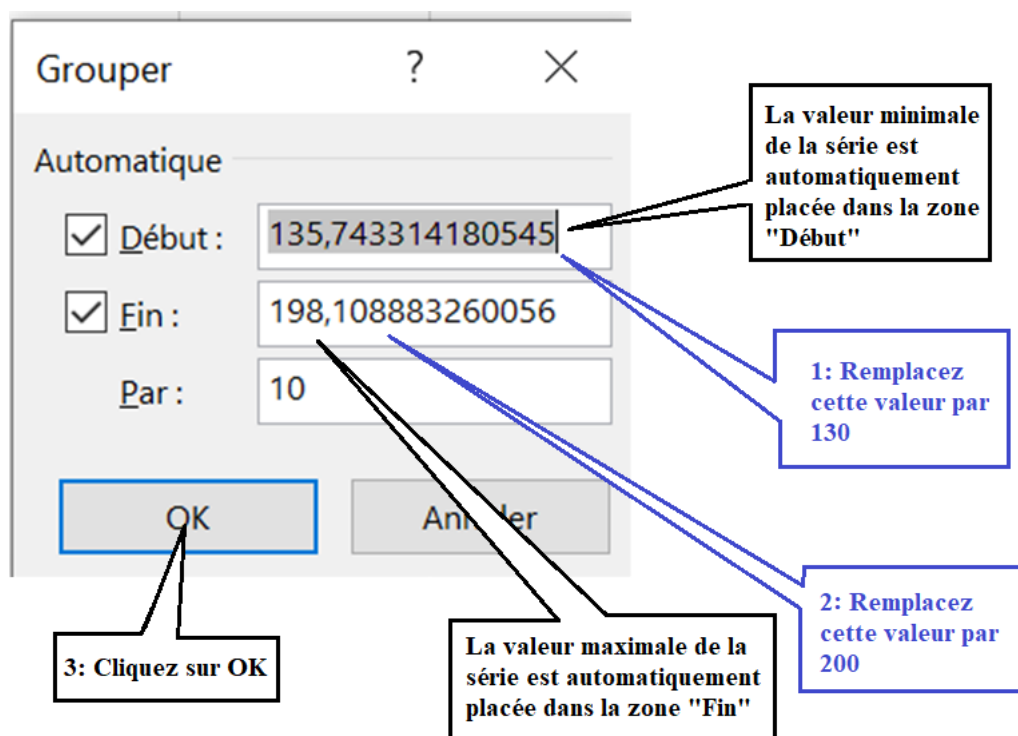
Lorsqu'on travaille avec des variables continues, dans la mesure où il y a une infinité de réponses possibles, plutôt que d'envisager une ligne par valeur possible de la variable *Taille(X<sub>i</sub>)*, on regroupe les valeurs par classes (mais vous le saviez déjà, puisque nous l'avons vu lors du cours théorique).

Nous allons envisager des classes d'amplitude 10, en s'assurant que la borne inférieure de la première classe soit un multiple de 10.

Cliquez sur l'une des cellules du tableau. Ensuite, appuyez sur le clic droit de votre souris, et cliquez sur l'option « Grouper... »



Un interface va s'ouvrir :



**Remarque :** la catégorie « Par » est celle dans laquelle on spécifie l'amplitude des classes désirée. S'il est indiqué 10, il n'est pas nécessaire de la modifier pour cet exemple. S'il est indiqué une autre valeur, remplacez là par 10.

Résultat attendu :

A6					
	A	B	C	D	E
1	Tableau de fréquences de la variable "Taille(X <sub>i</sub> )"				
2	Étiquettes de lignes				
3	130-140				
4	140-150				
5	150-160				
6	160-170				
7	170-180				
8	180-190				
9	190-200				
10	Total général				

La suite est identique à ce que vous avez fait précédemment, pour les tableaux de fréquences de variables discrètes. Reproduisez donc toutes les étapes de sorte à calculer les fréquences absolues, relatives, absolues cumulées et relatives cumulées.

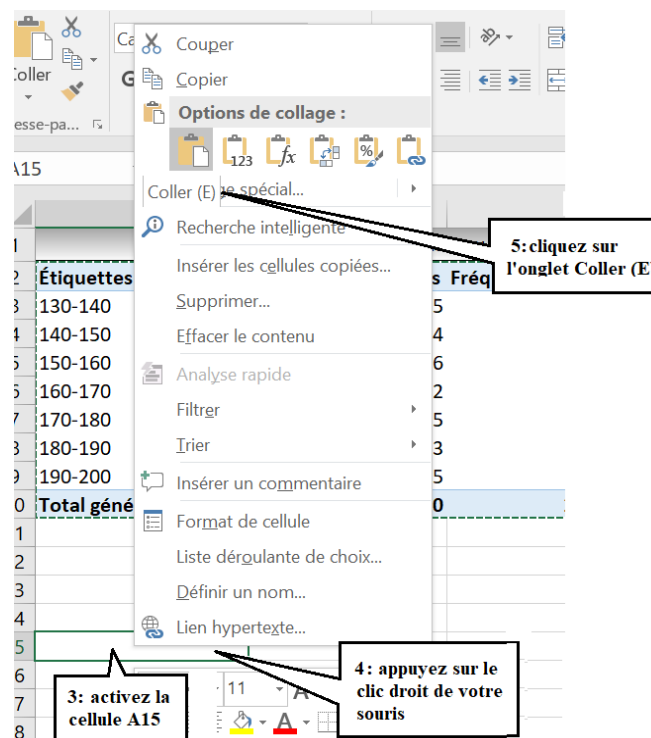
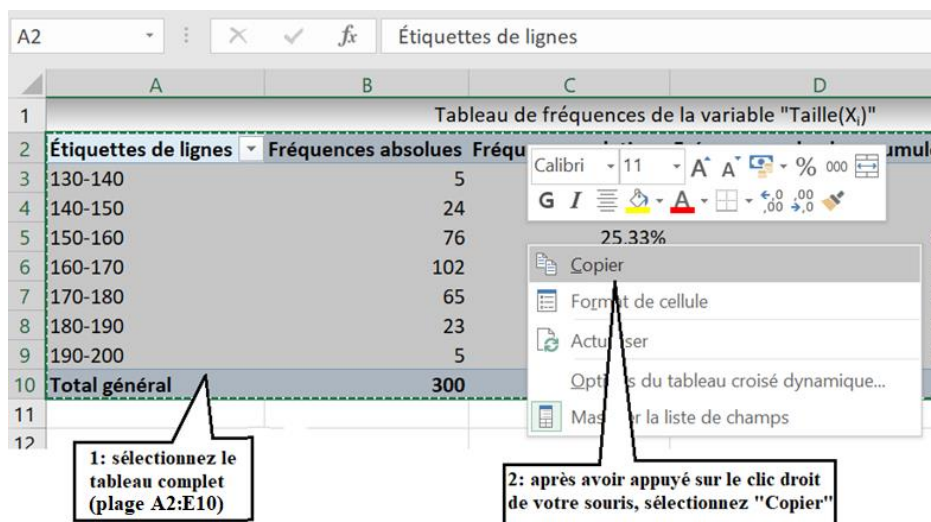
Résultat attendu :

E14					
	A	B	C	D	E
1	Tableau de fréquences de la variable "Taille(X <sub>i</sub> )"				
2	Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées
3	130-140	5	1,67%	5	1,67%
4	140-150	24	8,00%	29	9,67%
5	150-160	76	25,33%	105	35,00%
6	160-170	102	34,00%	207	69,00%
7	170-180	65	21,67%	272	90,67%
8	180-190	23	7,67%	295	98,33%
9	190-200	5	1,67%	300	100,00%
10	Total général	300	100,00%		

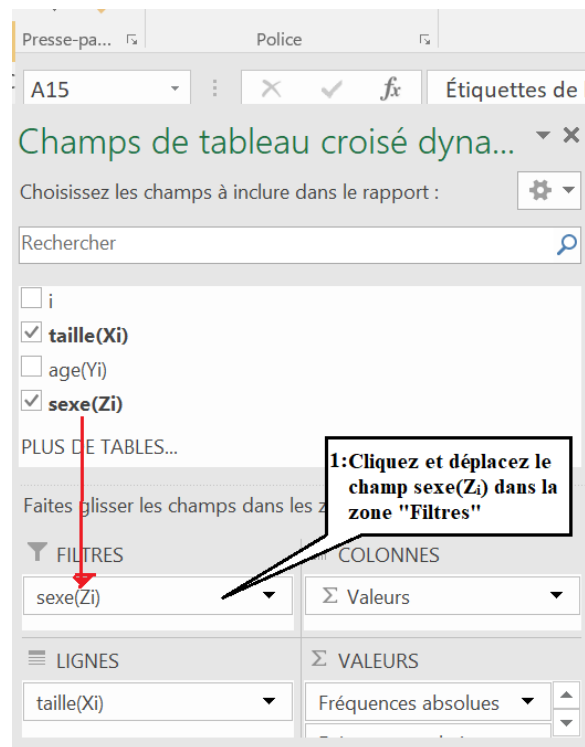
## Inclusion d'un filtre

Il est possible de spécifier que l'on ne veut pas prendre toutes les observations du tableau de données brutes en compte. Nous allons à nouveau réaliser un tableau de fréquences de la variable  $Taille(X_i)$ , mais en considérant seulement les hommes de la base de données.

Reproduisez à l'identique le tableau de fréquences de la variable  $Taille(X_i)$ , en vous assurant que le coin supérieur gauche se situe dans la cellule A15. Vous pouvez soit reproduire toutes les étapes apprises précédemment (si vous voulez vous entraîner), ou beaucoup plus simplement, sélectionner le tableau complet, et le coller dans la cellule A15.



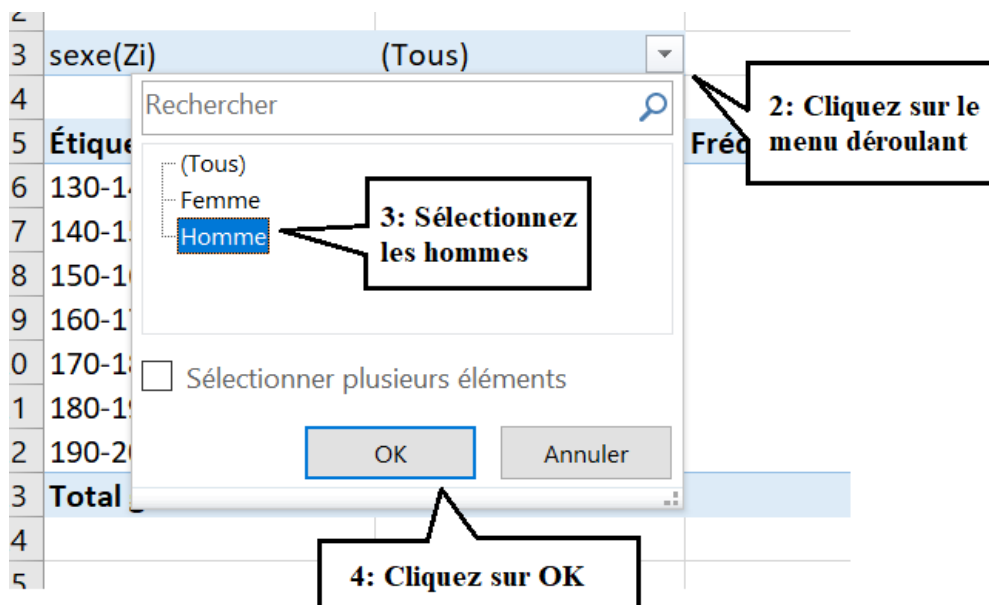
Nous allons maintenant spécifier que seules les lignes qui se rapportent aux hommes doivent être considérées dans le tableau de fréquences.



Un filtre sera placé au-dessus du tableau de fréquences :

12						
13	sexe(Zi)	(Tous)				
14						
15	Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées	
16	130-140	5	1,67%	5	1,67%	
17	140-150	24	8,00%	29	9,67%	
18	150-160	76	25,33%	105	35,00%	
19	160-170	102	34,00%	207	69,00%	
20	170-180	65	21,67%	272	90,67%	
21	180-190	23	7,67%	295	98,33%	
22	190-200	5	1,67%	300	100,00%	
23	Total général	300	100,00%			





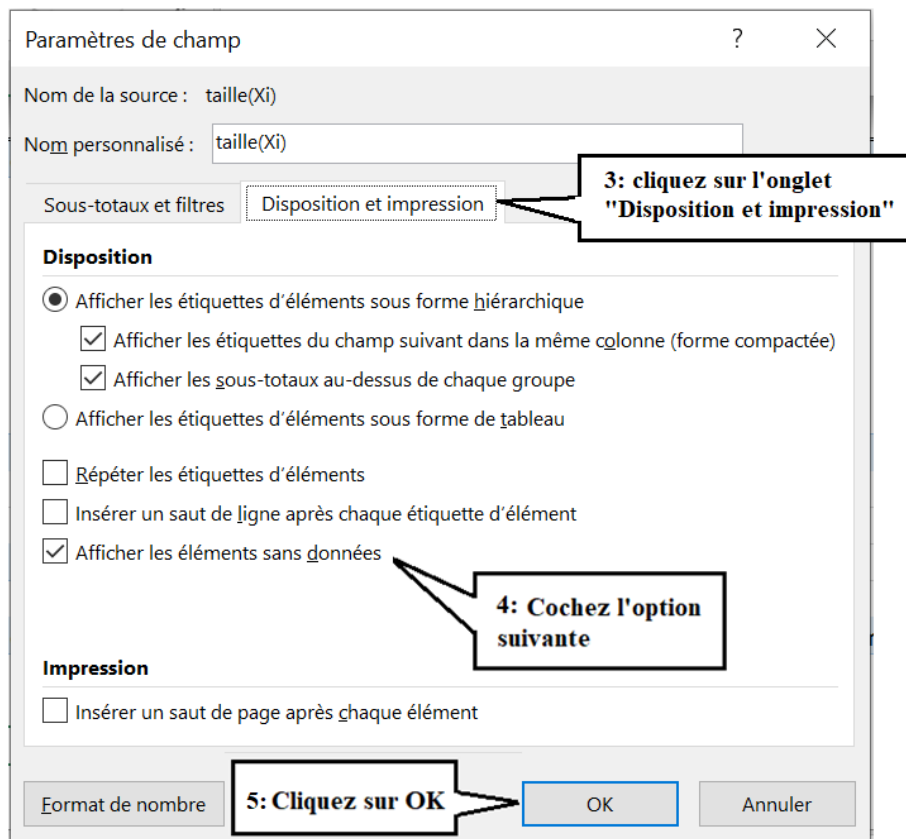
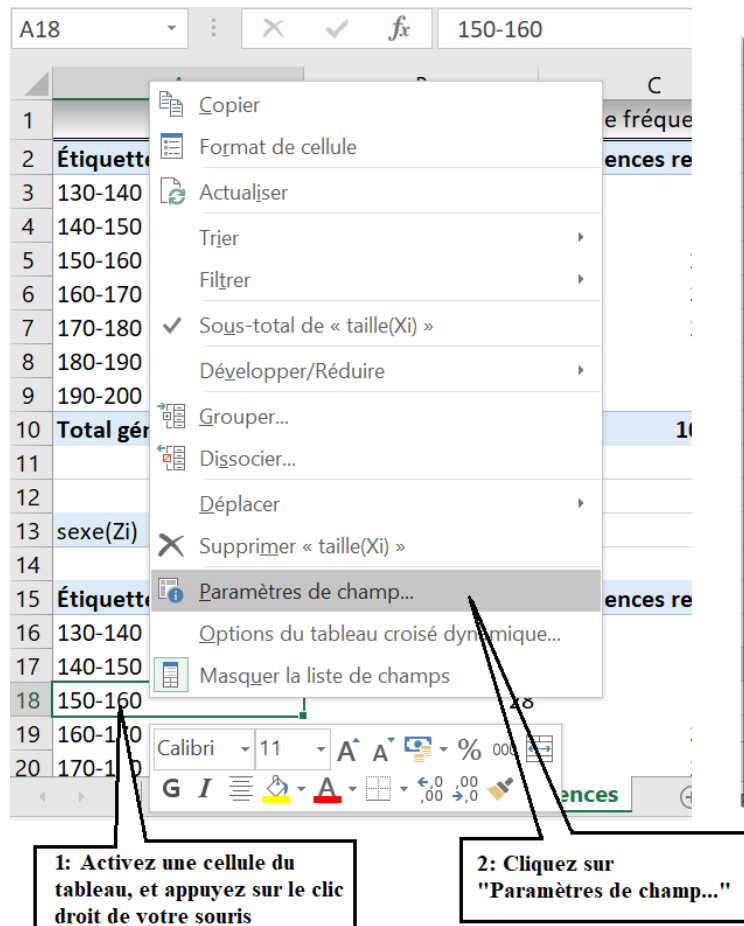
Combien d'hommes de la base de données mesurent entre 150 et 160 centimètres ?

28

Une des classe d'âge qui apparaît dans la base de données complète n'est représentée par aucun homme (et n'apparaît dès lors pas par défaut dans le tableau de fréquences des hommes). De quelle classe s'agit-il ?

La classe 130-140.

Il serait préférable de voir apparaître toutes les classes de taille possibles. Imaginons que l'on souhaite comparer le tableau de fréquences des hommes et celui des femmes (à l'aide de graphiques comparatifs, par exemple). La comparaison sera plus aisée si l'on s'assure que toutes les classes soient représentées (même celles qui sont associées à une fréquence nulle). Nous allons rendre apparente la catégorie associée à une fréquence nulle, dans le tableau de fréquences de la taille des hommes :



Résultat attendu :

	A	B	C	D	E
12					
13	sexe(Zi)	Homme			
14					
15	Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées
16	<130		0,00%	0	0,00%
17	130-140		0,00%	0	0,00%
18	140-150	4	2,67%	4	2,67%
19	150-160	28	18,67%	32	21,33%
20	160-170	52	34,67%	84	56,00%
21	170-180	42	28,00%	126	84,00%
22	180-190	19	12,67%	145	96,67%
23	190-200	5	3,33%	150	100,00%
24	>200		0,00%	150	100,00%
25	Total général	150	100,00%		

À votre avis, pourquoi la cellule associée à la fréquence absolue de la catégorie 130-140 est-elle vide ? .....

Parce que personne, dans le groupe des hommes, ne mesure entre 130 et 140 cm.

Notez que par défaut, deux catégories sont rajoutées : « <130 » et « >200 ». Elles seront toujours associées à une fréquence nulle. Ces catégories sont inutiles (du moins pour l'usage qu'on veut en faire), et nous allons dès lors les supprimer du tableau de fréquences.

1: Ouvrez le menu déroulant en cliquant sur la flèche

2: décochez la case "<130"

3: Déroulez le menu, jusqu'à voir la case ">200", de sorte à la décocher également

Options de tri supplémentaires...

Effacer le filtre de « taille(Xi) »

Filtres s'appliquant aux étiquettes

Filtres s'appliquant aux valeurs

Rechercher

☒ 160-170

☒ 170-180

☒ 180-190

☒ 190-200

☐ >200

Résultat attendu :

sexe(Zi)	Homme			
Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées
130-140		0,00%	0	0,00%
140-150	4	2,67%	4	2,67%
150-160	28	18,67%	32	21,33%
160-170	52	34,67%	84	56,00%
170-180	42	28,00%	126	84,00%
180-190	19	12,67%	145	96,67%
190-200	5	3,33%	150	100,00%
Total général	150	100,00%		

Reproduisez toutes ces étapes à l'identique pour réaliser un tableau de fréquences de la taille des femmes, en considérant les mêmes classes d'âge. Assurez-vous que le coin supérieur gauche du tableau de fréquences des femmes se trouve dans la cellule A30.

**Remarque :** il est à nouveau possible de copier le tableau de fréquences calculé sur l'ensemble des sujets, de le coller dans la cellule A30 et de rajouter ensuite le filtre pour ne sélectionner que les femmes. Cela étant fait, il ne reste qu'à vérifier si toutes les classes modales de la variable sont visibles (et le cas échéant, de les rendre visibles), et de désélectionner les catégories « <130 » et « >200 »

Résultat attendu :

27					
28	sexe(Zi)	Femme			
29					
30	Étiquettes de lignes	Fréquences absolues	Fréquences relatives	Fréquences absolues cumulées	Fréquences relatives cumulées
31	130-140	5	3,33%	5	3,33%
32	140-150	20	13,33%	25	16,67%
33	150-160	48	32,00%	73	48,67%
34	160-170	50	33,33%	123	82,00%
35	170-180	23	15,33%	146	97,33%
36	180-190	4	2,67%	150	100,00%
37	190-200		0,00%	150	100,00%
38	Total général	150	100,00%		

## Exercices

1. Ouvrez le fichier *Entraînement.tennis.xlsx*.

Dans la feuille *Tableau de fréquences*, construisez un tableau de fréquences complet de la série de la feuille *Données brutes*, en utilisant la technique des tableaux croisés dynamiques.

2. Juste avant une course à pied de 12 km, il a été demandé à 500 participants de spécifier le temps qu'ils passent par semaine à s'entraîner à la course. Ces informations ont été encodées dans la feuille *Données brutes* du fichier *temps.course.xlsx*. Dans la base de données, figurent également des informations sur leur temps moyen lors de ladite course ainsi que sur leur genre (masculin vs. Féminin).

2.a . Dans la feuille *Tableaux de fréquences*, réalisez un tableau de fréquences de la variable  $Min/km(Y_i)$ . Le coin supérieur gauche du tableau devra se trouver dans la cellule A3). Réalisez des classes d'amplitude égale à 1 minute (la borne inférieure de la première classe doit valoir exactement 3, et la borne supérieure de la dernière classe doit valoir 10).

**Remarque :** n'oubliez pas d'adapter le nom de chaque colonne. Une fois que vous avez terminé, écrivez le titre suivant dans la ligne 1 : « Tableau de fréquence du rythme de course, exprimé en minutes par km ». Centrez le titre et mettez-le en forme (faites-en sorte qu'il ne soit pas plus large que le tableau de fréquences, en sélectionnant l'option « renvoyer à la ligne automatiquement » de l'onglet Accueil, puis Alignement. Pensez à votre cours d'introduction à Excel pour ce faire).

2.b. Toujours dans la feuille *Tableau de fréquences*, réalisez un tableau de fréquences de la variable  $Min/km(Y_i)$ , incluant uniquement les individus de genre masculin (le coin supérieur gauche du tableau devra se situer dans la cellule G3). Pour ce faire, placez la variable  $Genre(X_i)$  dans la zone FILTRES, et ne sélectionnez que les individus de genre masculin (codés M). Assurez-vous que votre tableau soit constitué exactement des mêmes classes que le tableau de fréquences calculé pour l'ensemble de la base de données.

2.c. Refaites exactement la même chose, dans un tableau qui commencera à la cellule M3, mais en sélectionnant cette fois uniquement les femmes (codées F). Le coin supérieur gauche du tableau devra se situer dans la cellule M3.

## Calculs au départ des tableaux de fréquences : usage de la fonction SOMMEPROD

Pour rappel, lors du cours théorique, nous avons vu que certains paramètres pouvaient être calculés :

- Soit au départ de données brutes
- Soit au départ des fréquences (absolues ou relatives).

Pour rappel, nous avons vu trois paramètres pour lesquels je vous ai demandé de pouvoir appliquer les formules au départ des fréquences (absolues et relatives). Les voici :

### La moyenne

$$\text{Au départ des fréquences absolues: } \bar{X} = \frac{\sum_{j=1}^k x_j \times n_j}{n}$$

$$\text{Au départ des fréquences relatives: } \bar{X} = \sum_{j=1}^k x_j \times f_j$$

### L'écart moyen absolu

$$\text{Au départ des fréquences absolues: } \text{EMA}_X = \frac{\sum_{j=1}^k |x_j - \bar{X}| \times n_j}{n}$$

$$\text{Au départ des fréquences relatives: } \text{EMA}_X = \sum_{j=1}^k |x_j - \bar{X}| \times f_j$$

### La variance

$$\text{Au départ des fréquences absolues: } S_X^2 = \frac{\sum_{j=1}^k (x_j - \bar{X})^2 \times n_j}{n}$$

$$\text{Au départ des fréquences relatives: } S_X^2 = \sum_{j=1}^k (x_j - \bar{X})^2 \times f_j$$

Pouvoir appliquer ces formules signifie que vous devez savoir, au départ d'un tableau de fréquences, calculer la moyenne, l'EMA ou la variance d'une variable.

Pour **chaque** formule, remarquez qu'il y a systématiquement une partie de la formule qui implique de :

- a) Effectuer un produit entre plusieurs valeurs, pour chaque modalité ;
- b) Effectuer la somme des produits calculés pour chaque modalité.

Vous avez vu précédemment que la formule adéquate pour ce genre de calcul est la fonction **SOMMEPROD()**.

Ouvrez le fichier *Nb\_chambres.xlsx*. Vous y trouverez deux tableaux de fréquences du nombre de chambres dans la maison d'un certain nombre de propriétaires, vivant en ville ou à la campagne.

1. Combien de propriétaires vivent en ville ?  
.....500
2. Combien vivent à la campagne ?  
.....500

Nous allons commencer par calculer, dans la cellule H6, le **nombre moyen de chambres dans la maison des propriétaires à la campagne**, en utilisant la formule au départ des fréquences relatives :

$$\bar{X} = \sum_{j=1}^k x_j \times f_j$$

- 1.1. Quelle est la multiplication que nous souhaitons appliquer pour chaque ligne du tableau de fréquences ? .....  
 *$x_j \times f_j$*
- 1.2. Pour rappel, voici la manière dont on applique une multiplication dans Excel :

**Application** : = élément1 \* élément2

Complétez le tableau ci-dessous, en indiquant pour chaque ligne du tableau de fréquences, la référence de la cellule qui contient respectivement les valeurs des éléments 1 et 2 de la multiplication à appliquer (pour illustration, la première ligne a déjà été complétée).

Ligne du tableau de fréquence	Élément 1	Élément 2
1	A4	C4
2	..... A5	..... C5
3	..... A6	..... C6
4	..... A7	..... C7
5	..... A8	..... C8
6	..... A9	..... C9
7	..... A10	..... C10

- 1.3. Quelle est la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les valeurs de l'élément 1 (= l'argument *Matrice1* dans le rappel de fonction ci-dessous) ? .....  
**C4 :C10**

1.4. Appliquez correctement la fonction SOMMEPROD, dans la cellule H6.

**Application : = SOMMEPROD(Matrice1 ; Matrice2)**

- Matrice1 = la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les scores qui correspondent à l'élément 1 de la multiplication.
- Matrice2 = la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les scores qui correspondent à l'élément 2 de la multiplication.

Nous allons à présent recalculer le **nombre moyen de chambres dans la maison des propriétaires à la campagne**, dans la cellule I6, en utilisant la formule au départ des fréquences absolues :

$$\bar{X} = \frac{\sum_{j=1}^k x_j \times n_j}{n}$$

- 1.1. Quelle est la partie de la fraction qui pourra être calculée en utilisant la fonction **SOMMEPROD** (numérateur ou dénominateur) ? .....  
.....  
**numérateur**
- 1.2. Quelle est la multiplication que nous souhaitons appliquer pour chaque ligne du tableau de fréquences ? .....  
 **$x_j \times n_j$**
- 1.3. Quelle est la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les valeurs de l'élément 1 (= l'argument *Matrice1* dans le rappel de fonction ci-dessous) ? .....  
.....  
**A4:A10**
- 1.4. Quelle est la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les valeurs de l'élément 2 (= l'argument *Matrice2* dans le rappel de fonction ci-dessous) ? .....  
.....  
**B4 :B10**
- 1.5. Appliquez correctement la fonction SOMMEPROD, dans la cellule I6.
- 1.6. Vous venez de calculer le numérateur de la fraction qui permet de calculer la moyenne, au départ des fréquences absolues. Celui-ci doit encore être divisé par n. Quelle est la référence de la cellule contenant la valeur de n ? .....  
.....  
**B11 (ou D10 au choix)**



Il ne vous reste plus qu'à appliquer la formule adéquate dans la cellule I6.

Nous allons à présent calculer, dans la cellule H7, l'écart moyen absolu de la variable  $chambres(X_i)$ , en utilisant la formule au départ des fréquences relatives :

$$EMA_X = \sum_{j=1}^k |x_j - \bar{X}| \times f_j$$

1.1. Quelle est la multiplication que nous souhaitons appliquer pour chaque ligne du tableau de fréquences ? .....

$$|x_j - \bar{X}| \times f_j$$

1.2. À quoi correspond l'élément 1 de la multiplication ?.....  $|x_j - \bar{X}|$

1.2.1. Quel est le seul terme de cet élément dont la valeur va changer d'une ligne du tableau de fréquences à l'autre ? .....

$$x_j$$

1.2.2. Quelle est la plage de données qui contient, pour l'ensemble des lignes du tableau de fréquences, les valeurs du terme variable de l'élément 1 ? .....

A4:A10

La plage que vous avez dû trouver au point 1.2.2 ne constitue pas en l'état le premier terme de la fonction SOMMEPROD. Il va falloir la transformer :

■ Soustraite la moyenne (calculée dans la cellule H6) de la plage :

→ **Plagededonnées - H6**

■ Calculer la valeur absolue de cette soustraction

→ **ABS(Plagededonnées - H6)**

1.3. Quelle est la plage de données qui contient, pour l'ensemble des sujets, les valeurs de l'élément 2 (= l'argument Matrice2 dans le rappel de fonction ci-dessous) ? .....

C4 :C10

1.4. Appliquez correctement la fonction **SOMMEPROD** dans la cellule H7.

Nous allons à présent recalculer, dans la cellule I7, l'écart moyen absolu de la variable  $chambres(X_i)$ , en utilisant la formule au départ des fréquences absolues :

$$EMA_X = \frac{\sum_{j=1}^k |x_j - \bar{X}| \times n_j}{n}$$

1.1. Quelle est la partie de la fraction qui pourra être calculée en utilisant la fonction **SOMMEPROD** (numérateur ou dénominateur)? .....

.....  
numérateur

- 1.2. Quelle est la multiplication que nous souhaitons appliquer pour chaque ligne du tableau de fréquences ? .....

$|x_j - \bar{x}| \times n_j$

- 1.3. Appliquez la fonction SOMMEPROD correctement dans la cellule I7.  
N'oubliez pas ensuite de la diviser par n.

Comme exercice, calcule la variance de la variable *chambres*( $X_i$ ), en utilisant la formule au départ des fréquences absolues (cellule H8), et au départ des fréquences relatives (cellule I8). Pensez bien à toutes les étapes que vous venez de réaliser pour le calcul de l'EMA.