

Session 4 : Vérifier si le protocole a été respecté : sexe et âge

Dans le classeur de nettoyage des données, on trouve deux feuilles de calcul qui constituent la base de nos vérifications concernant le Sexe et l'Âge, *Vérification sexe* et *Vérification âge*.

On commence par *Vérification sexe*. Le protocole impose qu'un nombre de garçons et de filles à peu près égal soit testé dans chaque école. Les écarts à ce sujet doivent être stockés dans le classeur.

	A	B	C	D	E	F	G	H
1			Cette feuille sert à vérifier que le protocole concernant les sexes à échantillonner a été respecté dans chaque école					
2	nom de l'ensemble de données							
3								
4								
5		Sexe						
6	École	Nombre de garçons	Nombre de filles	Nombre de sexes manquants	Nombre total d'élèves	% de filles	Le protocole a-t-il été respecté dans cette école ? (O/N)	Remarques
7	101	25	25	0	50	50%	O	
8	102	25	25	0	50	50%	O	
9	103	24	26	0	50	52%	O	
10	104	50	0	0	50	0%	O	École de garçons, 2 élèves enregistrés sous '0' au lieu de '1'

Ci-dessus, on trouve un exemple de tableau de *Vérification sexe* pour une enquête de cartographie. Dans chaque école, 50 élèves ont été testés. La première école était une école mixte dans laquelle 25 élèves de chaque sexe ont été échantillonnés. La deuxième école était une école pour filles dans laquelle 50 filles ont été échantillonnées. Dans les deux cas, le protocole a été respecté (en espérant qu'une école pour garçons viendra équilibrer le bilan).

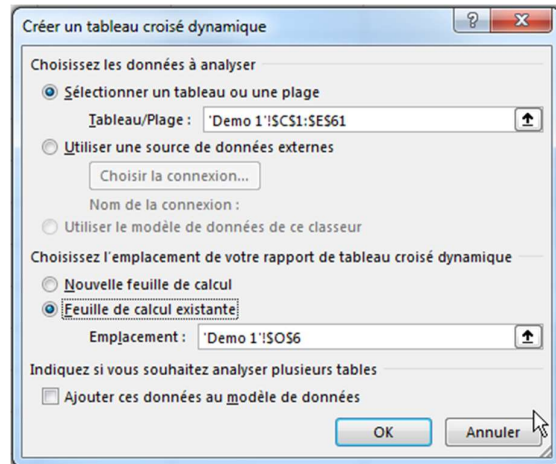
Démonstration Vérification sexe

- Ouvrir le fichier Session_4.xlsx et cliquer sur l'onglet de la première feuille *Demo 1*.
- Préparation du tableau : masquer les colonnes superflues, mettre en forme les en-têtes de la première ligne du tableau, insérer une colonne avec les numéros d'observation.
- En guise de petite vérification préliminaire, placer le curseur dans la première ligne de la variable *Sexe* et appuyer sur Ctrl + ↓ pour vérifier les valeurs manquantes. Ce raccourci placera le curseur à la fin d'un bloc.
[MAJ + Ctrl + ↓ mettra en surbrillance la colonne jusqu'à la fin du bloc]
- À présent, on crée un tableau croisé dynamique qui indique le nombre de garçons et de filles dans chaque école. Mettre

C	D	E
Sexe	Âge	Code école
1	13	1
0	10	1
1		1
0	11	1
1	13	1
0	13	1
1	15	1
0	12	1
1	11	1
0	11	1
1	12	1
0	10	1
1	12	1
0	10	1
1	11	1

en surbrillance les colonnes Sexe, Âge et Code école comme illustré à droite.

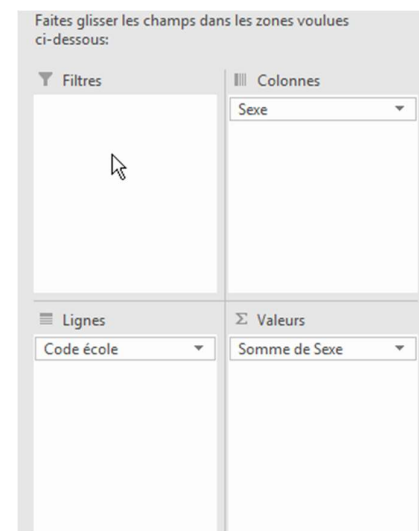
5. Cliquer sur *Tableau croisé dynamique* dans le ruban *Insertion*.
6. La fenêtre suivante apparaît :



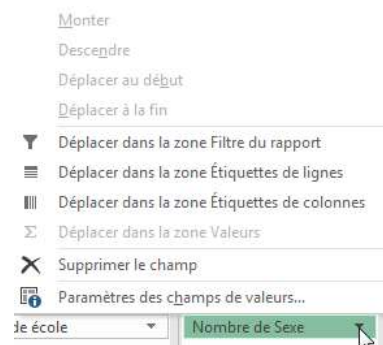
Le champ Tableau/Plage : désigne la plage de cellules sélectionnée à l'étape 2.

Dans la moitié inférieure de la fenêtre, un emplacement peut être choisi pour le tableau croisé dynamique. Dans cet exemple, j'ai cliqué sur la cellule O6 de la première feuille *Demo 1*. Cliquer sur OK.

7. La fenêtre de droite apparaît.
Faire glisser les noms de variable dans le tableau des paramètres du Tableau croisé dynamique comme illustré à droite.
Cliquer sur le triangle à côté de *Somme de sexe* et cliquer sur *Paramètres des champs de valeurs...*



La fenêtre suivante apparaît. Dans *Résumer le champ de valeur par*, mettre en surbrillance *Compte* et cliquer sur OK.



Le tableau croisé dynamique doit à présent indiquer le nombre de garçons et de filles de chaque école (1, 2, 3) ainsi que le nombre total d'élèves.

ID école →	Nombre de Sexe	Étiquettes de colonnes		Total général
	Étiquettes de lignes	0	1 (vide)	
1		10	10	20
2		10	10	20
3		11	8	19
Total général		31	28	59

Copier les trois premières colonnes dans le Classeur de nettoyage des données.

8. Pour déterminer le nombre total d'élèves et le % des élèves filles dans le Classeur de nettoyage des données, utiliser des formules :

B	C	D	E	F
Sexe				
Nombre de garçons	Nombre de filles	Nombre de sexes manquants	Nombre total d'élèves	% de filles
10	10	0	=SOMME(B7:D7)	=C7/E7

Maintenant, à vous

Ouvrir le fichier *data3.xlsx* de la dernière session et vérifier la variable Sexe. Transférer vos résultats dans le Classeur de nettoyage des données et enregistrer. Enregistrer les données nettoyées sous *data4.xlsx*.

Démonstration Vérification âge

9. Ouvrir le fichier Session_4.xlsx et aller sur la feuille *Demo 2*.<

10. Mettre en surbrillance les colonnes Âge et Code école et créer un tableau croisé dynamique avec les paramètres indiqués à droite.

Vous devez obtenir le tableau ci-dessous

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

Filtres	Colonnes
	Âge
Lignes	Valeurs
Code école	Nombre de Âge

Count of Age Row Labels	Column Labels													Grand Total
Labels	0	7	8	9	10	11	12	13	14	15	99	12345	(blank)	
1					3	5	4	4		2	1			19
2		1		2	5	3	2	2	2	3				20
3	1		1		4	4	1	1	3	3		1		19
Grand Total	1	1	1	2	12	12	7	7	5	8	1	1		58

Nombre de Âge	Étiquettes de colonnes														
Étiquettes de lignes	0	7	8	9	10	11	12	13	14	15	99 (vide)	Total général			
1					3	5	4	4		2	1			19	
2			1	2	5	3	2	2	2	3				20	
3			1	1	4	4	1	2	3	3				19	
Total général	1	1	1	2	12	12	7	8	5	8	1			58	

11. L'observation contenant l'âge 12345 semble être une donnée factice. Vérifier les autres valeurs de cette observation et déterminer si vous souhaitez retirer la totalité de cette observation.

Les âges 0 et 99 pourraient être factices également ou une erreur de frappe. Vérifier les autres valeurs de cette observation et déterminer si vous souhaitez supprimer les valeurs âge et les marquer comme valeurs manquantes ou si vous souhaitez supprimer la totalité de l'observation.

12. Copier les parties respectives du tableau croisé dynamique dans le Classeur de nettoyage des données. Si nécessaire, ajuster la plage d'âge dans le classeur par rapport à la plage d'âge des données.

Nombre total d'élèves = somme(< la totalité de la plage d'âge échantillonnée incluant les valeurs manquantes >)

Nombre d'élèves dans le protocole = somme(< la plage d'âge du protocole >)

% d'élèves dans le protocole = < nombre d'élèves dans le protocole > / < nombre total d'élèves >

Maintenant, à vous

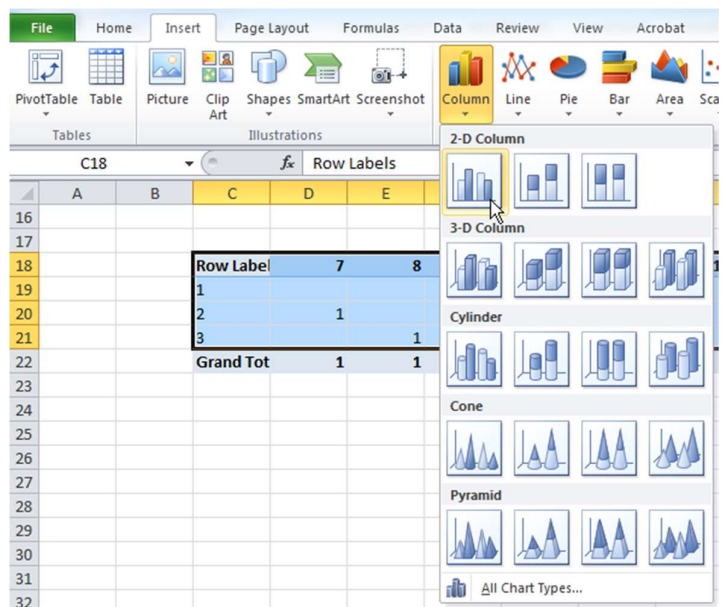
Poursuivre votre travail sur le fichier *data4.xlsx* en vérifiant la variable Âge. Transférer vos résultats dans le Classeur de nettoyage des données. À la fin de votre travail, enregistrer vos données sous 'données 4' et enregistrer le classeur de nettoyage des données.

Exercice facultatif

Pour vérifier les âges, on peut également créer une visualisation de la répartition des valeurs d'âge, par ex., un histogramme. Pour cela, vous avez besoin d'un tableau croisé dynamique qui incluent les âges comme noms de ligne

13. Créer un tableau croisé dynamique similaire à celui-ci.

Count of Age				
Column Labels				
Row Labels	1	2	3	Grand Total
7		1		1
8			1	1
9		2		2
10	3	4	4	11
11	5	3	4	12
12	4	2	1	7
13	4	2	1	7
14		2	3	5
15		2	3	5
(blank)				
20		1		1
Grand Total	18	20	17	55



14. Mettre en surbrillance la totalité du tableau croisé dynamique. Activer le ruban Insertion et cliquer sur Colonne. Sélectionner le premier *Histogramme 2-D* :

Vous devez obtenir un histogramme similaire à l'illustration de droite. Cet histogramme indique le nombre d'élèves pour chaque âge. Pour chaque école, une couleur différente est utilisée. Cliquer sur le triangle dans le bouton Code école à droite et sélectionner une école. Faire de même avec le bouton Âge en bas à gauche.

