

# Traitement des données avec Microsoft EXCEL 2016

Vincent Jalby

Faculté de Droit et des Sciences Économiques, Université de Limoges

vincent.jalby@unilim.fr

25 octobre 2016

## 1 Saisie des données

Les données collectées sont saisies dans une feuille Excel. Chaque ligne correspond à une observation (questionnaire), chaque colonne à une variable (question ou partie de question).

Observation	Q1	Q2	Q3a	Q3b
Obs 1	1	5	3	2

La première ligne contient le numéro ou l'intitulé de la question : utiliser des noms simples, courts, sans accent, sans espace.

La première colonne contient le numéro de l'observation (questionnaire) : lors de la saisie, chaque questionnaire est numéroté (physiquement) et son numéro retranscrit dans la première colonne.

## 2 Codage des questions

Le codage dépend de la nature des questions :

- **questions numériques** (age, nombre d'employés, ...) : la valeur est directement saisie ;
- **question dichotomique** (sexe, oui/non, ...) : on utilise les valeurs 0/1 (faux/vrai) ou 1/2 (homme/femme).
- **QCM à réponse unique** (situation familiale, ...) : les modalités de réponse sont numérotées ; seul le numéro de la modalité choisie est saisi.
- **QCM à réponses multiples** (moyens de transports utilisés, ...) : elles correspondent à autant de questions dichotomiques que de modalités de réponse. On codera donc ces questions par autant de colonnes que de modalités de réponse. Chaque colonne contiendra la valeur 1 ou 0 selon que l'individu a coché ou non cette modalité.
- **questions avec échelle** (Pas du tout d'accord, tout à fait d'accord, ...) : les modalités sont numérotées de 1 à 5 (dans le cas d'une échelle à 5 modalités) ou de -2 à +2. 1 ou -2 correspond à la modalité la plus négative (pas du tout d'accord), 5 ou +2 à la plus positive (tout à fait d'accord) et 3 ou 0 correspond à la modalité moyenne (ni d'accord, ni pas d'accord).

Obs	Age	Sexe	SitFam	Q4 :Voiture	Q4 :Velo	Q4 : Commun
1	25	1	1	0	1	0
2	29	2	1	1	0	1

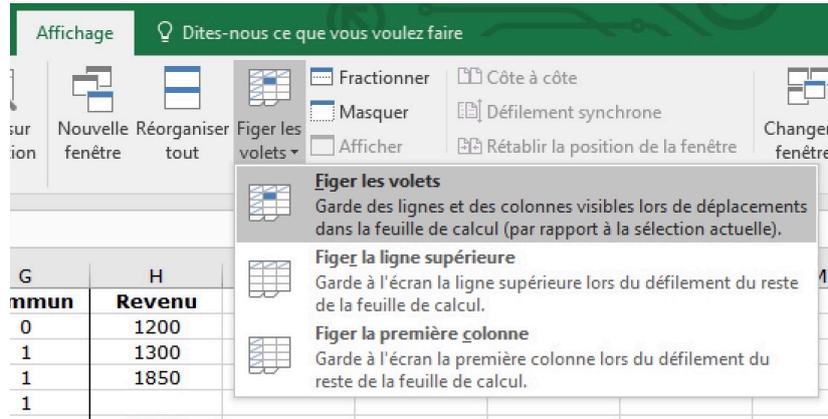
Les réponses manquantes (questions non répondues, ou incorrectement) doivent être codées, soit en laissant la cellule vide, soit par des 9/99/999. L'utilisation d'une cellule vide facilite le traitement avec les fonctions incluses dans Excel, en particulier pour les questions numériques, dichotomiques 0/1 et Likert. Pour les questions nominales et ordinale, il est préférable d'utiliser un 9/99.

Pour faciliter la saisie, il est souhaitable de rajouter deux lignes indiquant les règles de saisie. On peut aussi faire précéder la ligne des intitulés par une ligne reprenant le numéro des questions.

	A	B	C	D	E	F	G	H
1	Numéro	1	2	3	4 : Moyen de transport			5
2	Obs	Age	Sexe	SitFam	Voiture	Velo	Commun	Revenu
3	Valeur	0 à99	1 ou 2	1 à 5	0/1	0/1	0/1	0 à 9999
4	Manquante	vide	vide	vide	vide	vide	vide	vide
5	1	25	1	1	0	1	0	1200
6	2	29	2	1	1	0	1	1300
7	3	35	2	3	1	0	1	1850
8	4	19	2	2	0	0	1	
9	5		1	5				2500
10	6	45		4	1	0	0	3200
11	7	33	1	3	0	1	1	2200
12	8	22	1		1	0	0	1550

### 3 Organisation de la feuille

Afin de faciliter la saisie et la consultation des données, il est préférable de scinder la feuille en 3 : 1<sup>re</sup> colonne (numéro d'observation), 1<sup>res</sup> lignes (nom des variables, ...), données. Pour cela, il suffit de sélectionner la première cellule contenant une donnée (dans l'exemple précédent, la cellule de la 5<sup>e</sup> ligne, 2<sup>nd</sup>e colonne) puis de choisir Affichage > Fenêtre > Figurer les volets > Figurer les volets.



Une fois la saisie terminée, il faut supprimer les premières lignes (numéro question, valeur, manquante) pour ne conserver que la ligne des intitulés (et les données dessous). Pour ne pas perdre ces informations, il est préférable de dupliquer la feuille auparavant !

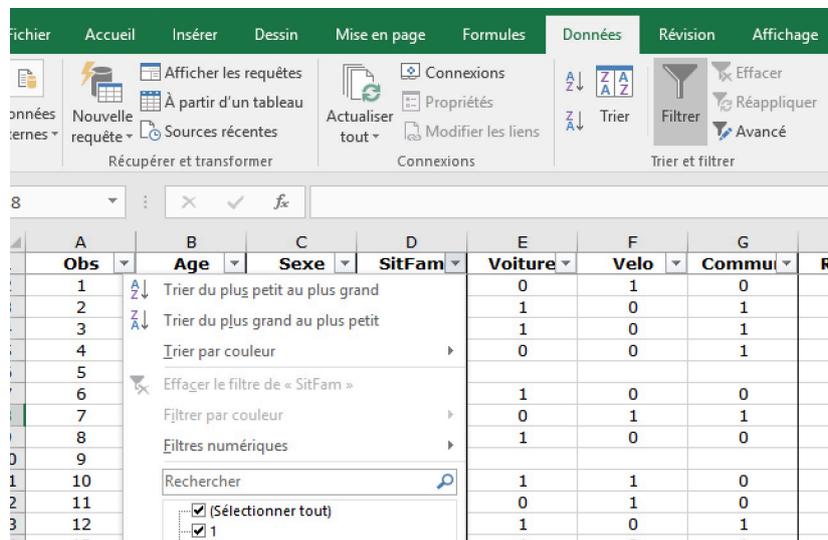
On est alors en présence d'une feuille du type

	A	B	C	D	E	F	G	H
1	Obs	Age	Sexe	SitFam	Voiture	Velo	Commun	Revenu
2	1	25	1	1	0	1	0	1200
3	2	29	2	1	1	0	1	1300
4	3	35	2	3	1	0	1	1850
5	4	19	2	2	0	0	1	
6	5		1	5				2500
7	6	45		4	1	0	0	3200
8	7	33	1	3	0	1	1	2200
9	8	22	1		1	0	0	1550
10	9		2	3				1300

L'utilisation de filtres automatiques est souvent pratique pour trier/sélectionner des données. Pour cela, il suffit de sélectionner la première ligne (nom des variables) puis de choisir Données > Trier et filtrer > Filtrer.

Les menus apparaissant à côté du nom de chaque colonne (sur la première ligne) permettent de trier l'ensemble du tableau de données par ordre croissant ou décroissant de valeurs de la variable de la colonne choisie.

Il est aussi possible de « filtrer » sur une ou plusieurs valeurs, voire une plage de valeurs, de la variable les lignes du tableau. Les autres lignes sont juste masquées.



## 4 Fonctions, Utilitaire d'analyse, Tableau croisé dynamique

Il est possible de faire un traitement des données complet en utilisant les fonctions d'EXCEL : NB, NBVAL, NBVIDE, NB.SI, MOYENNE, ECARTYPE, MIN, MAX.

Toutefois, EXCEL inclut des outils avancés permettant d'automatiser ce traitement :

- **Utilitaire d'analyse** : ensemble d'outils permettant d'étudier une ou plusieurs variables *quantitative* ;
- **Tableau croisé dynamique** : outil permettant d'étudier une ou plusieurs variables *qualitatives* mais aussi quantitative.

L'**utilitaire d'analyse** fait partie des *macros complémentaires* d'EXCEL qui ne sont pas activées par défaut. Avant la première utilisation, il faut donc l'activer :

1. Dans l'onglet Fichier, sélectionner l'item Options.
2. Dans la fenêtre qui s'ouvre, choisir Compléments puis cliquer sur le bouton Atteindre en bas de la fenêtre.
3. Dans la nouvelle fenêtre « Compléments », cocher Analysis ToolPak. puis valider (OK).

Tous les traitements seront fait dans une ou plusieurs feuilles (onglets) séparées. Dans la suite, on supposera que la feuille des données est nommée donnees et la feuille de résultat est nommée resultats.

## 5 Traitements univariés

### 5.1 Variables quantitatives

Elles correspondent aux questions numériques (revenu, âge, ...). Certains traitements s'appliquent aussi aux variables qualitatives ordinales (en particulier issues de questions avec échelle de type Likert) et aux questions dichotomiques codées en 0/1 (variables qualitatives nominales).

#### 5.1.1 Résumés statistiques

Ils correspondent à la moyenne, minimum, maximum, écart-type, somme, etc. On peut les calculer en une seule opération en utilisant l'**utilitaire d'analyse** :

1. sélectionner Données > Analyse > Utilitaire d'analyse, puis Statistiques descriptives ;
2. pour la plage d'entrée, sélectionner (dans la feuille donnees) la ou les colonnes correspondant aux variables quantitatives à étudier (il suffit pour cela de sélectionner la colonne complète, en cliquant sur la lettre de la colonne) ;
3. cocher « Intitulées en première ligne » et « Rapport détaillé » ;
4. décocher « Niveau de confiance pour la moyenne » ;
5. cocher « Plage de sortie » et sélectionner une cellule vide de la feuille resultats (les résultats seront affichés à partir de cette cellule, en dessous et à droite). ou cocher « Insérer une nouvelle feuille » pour qu'une nouvelle feuille soit créée.

	A	B	C	D	E	F	G	H
1	Obs	Age	Sexe	SitFam	Voiture	Velo	Commu	Reven
2	1	25	1	1	0	1	0	1200
3	2	29	2					
4	3	35	2					
5	4	19	2					
6	5		1					
7	6	45						
8	7	33	1					
9	8	22	1					
10	9		2					
11	10	38						
12	11	25	1					
13	12	29	2					
14	13		2					
15	14	19	2					
16	15	42	1					
17	16	45						
18	17	33	1					
19	18	22	1					
20	19	21	2					
21	20	38	2					
22	21	46						
23	22	32	1					

### 5.1.2 Résumés statistiques : méthode alternative

Il est aussi possible d'utiliser un tableau croisé dynamique pour générer les résumés statistiques d'une variable quantitative :

1. sélectionner les données dans l'onglet correspondant (il est suffit de sélectionner la colonne correspondante en cliquant sur l'entête de colonne) ;
2. choisir Insérer > Tableaux > Tableau croisé dynamique et cliquer sur OK ;
3. faire glisser la variable à étudier de la liste des champs dans la zone **Valeurs** puis cliquer sur le menu local à coté du nom de la variable et sélectionner **Paramètres des champs de valeurs**. Sélectionner alors le résumé souhaité (par exemple Moyenne) dans la liste des **synthèse des valeurs par** ;
4. recommencer l'opération précédente pour chaque résumé statistique souhaité ;
5. faire glisser l'étiquette **Valeurs** de la case **Colonnes** vers la case **Lignes** pour afficher les différents résumés en colonne (et non en ligne).

The screenshot shows an Excel spreadsheet with a PivotTable for 'Age'. The PivotTable fields are: Lignes: Valeurs, Colonnes: (empty). The 'Valeurs' field is set to 'Moyenne de Age2'. A dialog box 'Paramètres des champs de valeurs' is open, showing 'Moyenne' selected under 'Résumer le champ de valeur par'. A 'Champs de tableau croisé...' task pane is also visible on the right.

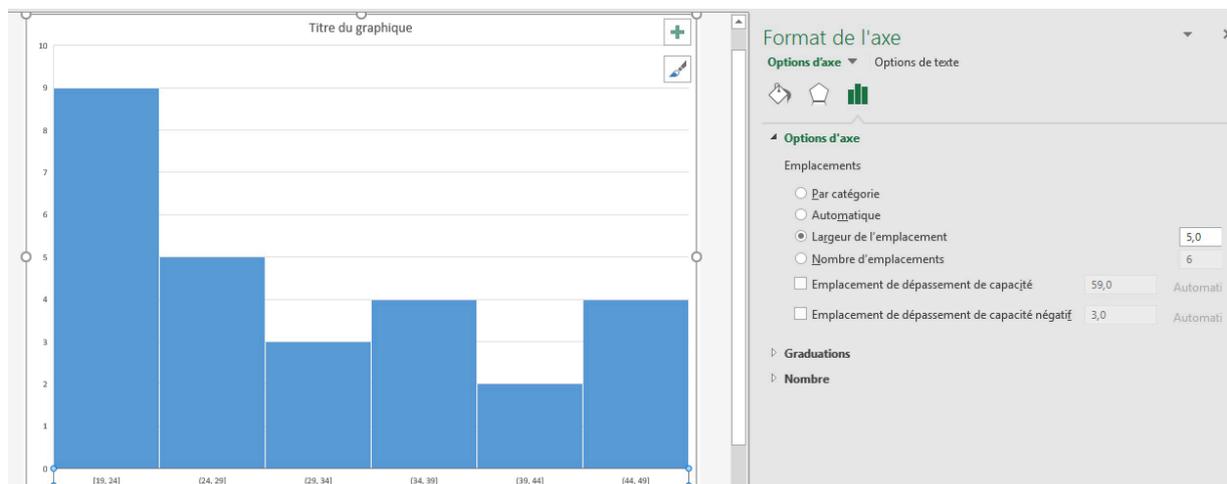
### 5.1.3 Représentation graphique : histogramme

Depuis la version 2016, il est possible de créer facilement un véritable histogramme d'une variable quantitative (continue).

The screenshot shows the 'Insérer un graphique' dialog box in Excel. The 'Graphiques recommandés' tab is active, and 'Histogramme' is selected. The background shows a PivotTable with 'Obs' and 'Age' columns.

Pour obtenir un histogramme, il faut :

1. sélectionner les données dans l'onglet correspondant (il suffit de cliquer sur l'entête de colonne de la variable concernée)
2. choisir Insérer > Graphiques > Graphiques recommandés > Tous les graphiques > Histogramme (attention de bien choisir le second « Histogramme », en bas de la liste ;
3. le graphique est créé sur l'onglet des données ; pour le déplacer, utiliser le bouton Outils de graphique > Création > Déplacer le graphique ;
4. il sera certainement nécessaire de modifier les intervalles de valeurs de chaque bâton (la largeur des plages) ; pour cela faire un clic droit sur l'axe horizontal et sélectionner Mise en forme de l'axe... Cocher alors Largeur de l'emplacement (= largeur des intervalles) et indiquer une valeur adaptée dans la case.



Il est aussi possible d'utiliser l'**utilitaire d'analyse** pour générer un histogramme. La procédure est plus lourde mais permet d'utiliser des intervalles de valeurs de longueurs différentes. Le résultat (graphique) est moins satisfaisant. La procédure est expliquée dans la version « 2010 » de ce document.

#### 5.1.4 Représentation graphique : boîte à moustaches

La version 2016 d'EXCEL permet aussi de créer facilement une boîte à moustaches. On procède comme pour l'histogramme mais en sélectionnant le graphique **zone et valeur**.

## 5.2 Variables qualitatives

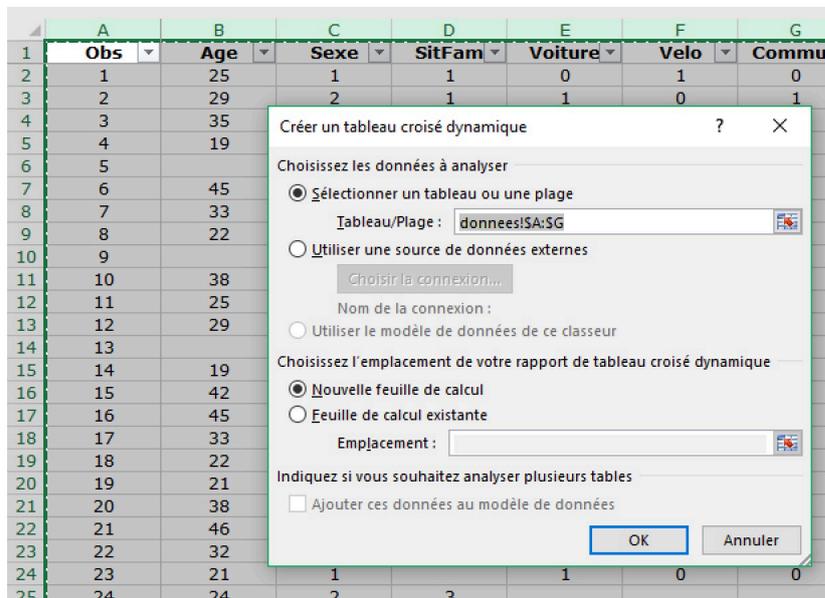
Elles correspondent aux QCM. Elles sont soit ordinales (les modalités sont classées) soit nominales (les modalités ne peuvent pas être classées), en particulier dichotomiques. Certains traitements s'appliquent aussi aux variables quantitatives discrètes (nombre d'enfants, ...).

### 5.2.1 Tri à plat

Outre les résumés statistiques (*mode* pour les variables nominales, *minimum*, *maximum*, *médiane* pour les variables ordinales, *moyenne* et *écart-type* pour les variables issues d'échelle de Likert) obtenus via les « Statistiques descriptives » de l'utilitaire d'analyse (*c.f.* section 5.1.1), la présentation de tris à plat permet de décrire complètement une variable qualitative (ou quantitative discrète).

Pour cela, on utilise le « rapport de tableau croisé dynamique » :

1. sélectionner les données dans l'onglet correspondant (il est suffit de sélectionner les colonnes correspondantes en cliquant sur les entêtes de colonne) ;
2. choisir Insérer > Tableaux > Tableau croisé dynamique et cliquer sur OK ;



3. faire glisser le nom de la variable à étudier de la liste des champs vers la case **Lignes** ;
4. faire glisser à présent la variable **Obs** (ou la variable étudiée) dans la case **Valeurs** : le tri à plat s'effectue automatiquement !

**Modalités** : numéro à remplacer par intitulé.

**Menu local** : permet de choisir les modalités à afficher.

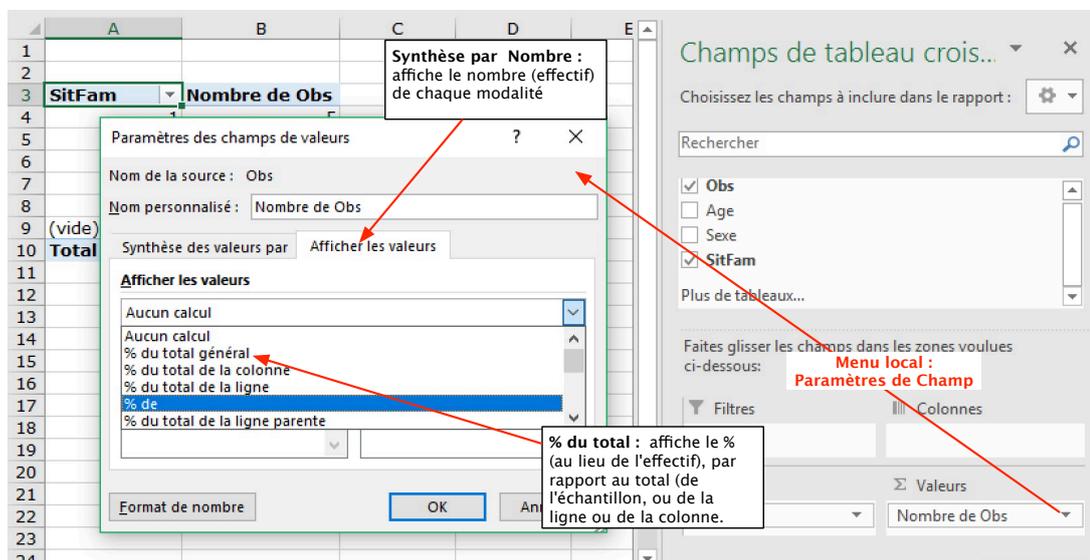
**Liste des variables** : correspond aux colonnes sélectionnées. La variable age (quantitative) ne doit pas être utilisée.

**Tableau dynamique** : Son contenu se modifie dynamiquement.

**Faire glisser ici** (pointant vers la zone Valeurs)

**Faire glisser ici** (pointant vers la zone Lignes)

5. changer ensuite dans la première colonne les numéros des modalités par leur nom ;
6. pour masquer la modalité « vide » (= donnée manquante), il suffit de cliquer dans le menu local de la colonne 1 et de décocher « vide ».
7. pour afficher les pourcentages, sélectionner **Paramètres des champs de valeurs** dans le menu local du nom de la variable de la zone **Valeurs** ; dans la nouvelle fenêtre, cliquer alors sur Afficher les valeurs, puis dans le menu Afficher les valeurs sélectionner % du total ;



### 5.2.2 Diagramme en bâtons ou en secteurs

Il suffit de cliquer sur le bouton « Graphique croisé dynamique » de la barre d'outils (cliquer d'abord sur l'onglet « Outils de tableau croisé dynamique »). Mais attention, toute modification du graphique sera répercutée sur le tableau et inversement.

### 5.3 Variables qualitatives issues d'une question à réponses multiples

Ces variables, codées en 0/1, nécessitent un traitement simultané. On peut soit utiliser les fonctions d'Excel pour effectuer les calculs manuellement, soit utiliser le « rapport de tableau croisé dynamique » de manière un peu détournée : on procède comme précédemment, mais

- au point 1, on sélectionne l'ensemble des colonnes/variables 0/1 correspondantes aux réponses à la question.
- au points 3 et 4, faire glisser toutes les variables (une par une) vers la case **Valeurs**.
- cliquer ensuite sur le menu de chaque variable et dans la zone Paramètres de champ sélectionner Somme au lieu de Nombre.
- On obtient ainsi les effectifs de chacune des réponses.

## 6 Traitements bivariés

### 6.1 Variables quantitatives

Pour étudier le lien entre deux variables quantitatives, il est possible de calculer leur coefficient de corrélation linéaire et/ou de les représenter graphiquement.

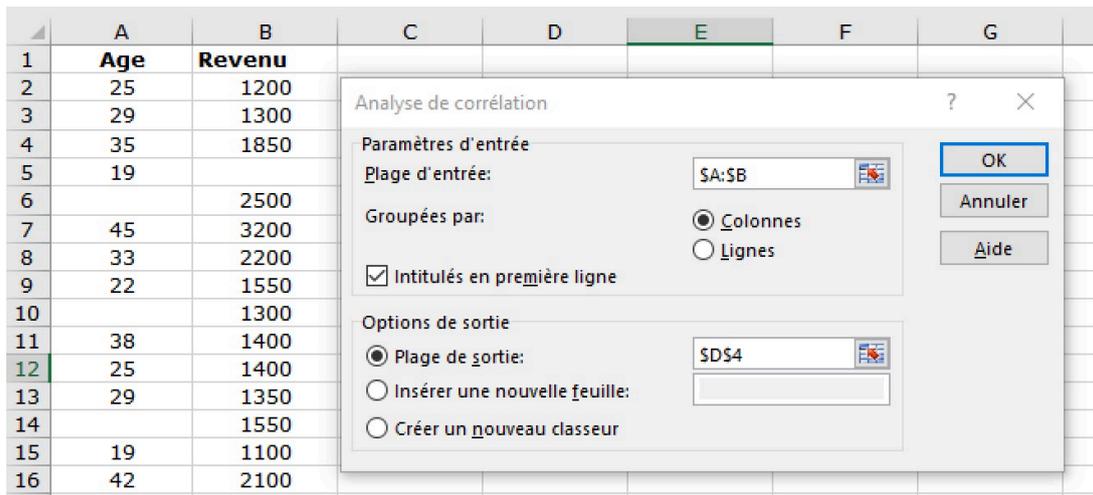
Lorsque les variables quantitatives à étudier ne se trouvent pas dans des colonnes adjacentes, il est pratique de recopier ces colonnes dans une nouvelle feuille.

#### 6.1.1 Coefficient de corrélation linéaire

Pour calculer le coefficient de corrélation linéaire, on peut soit utiliser la fonction `COEFFICIENT.CORRELATION`, soit utiliser l'utilitaire d'analyse :

1. sélectionner Données > Analyse > Utilitaire d'analyse, puis Analyse de corrélation;
2. pour la plage d'entrée, sélectionner (dans la feuille données) les colonnes correspondant aux variables quantitatives à étudier; il suffit pour cela de sélectionner les colonnes complètes, en majuscule-cliquant sur la lettre de chaque colonne;
3. cocher « Intitulées en première ligne »;
4. cocher « Plage de sortie » et sélectionner une cellule vide de la feuille résultats (les résultats seront affichés à partir de cette cellule, en dessous et à droite).

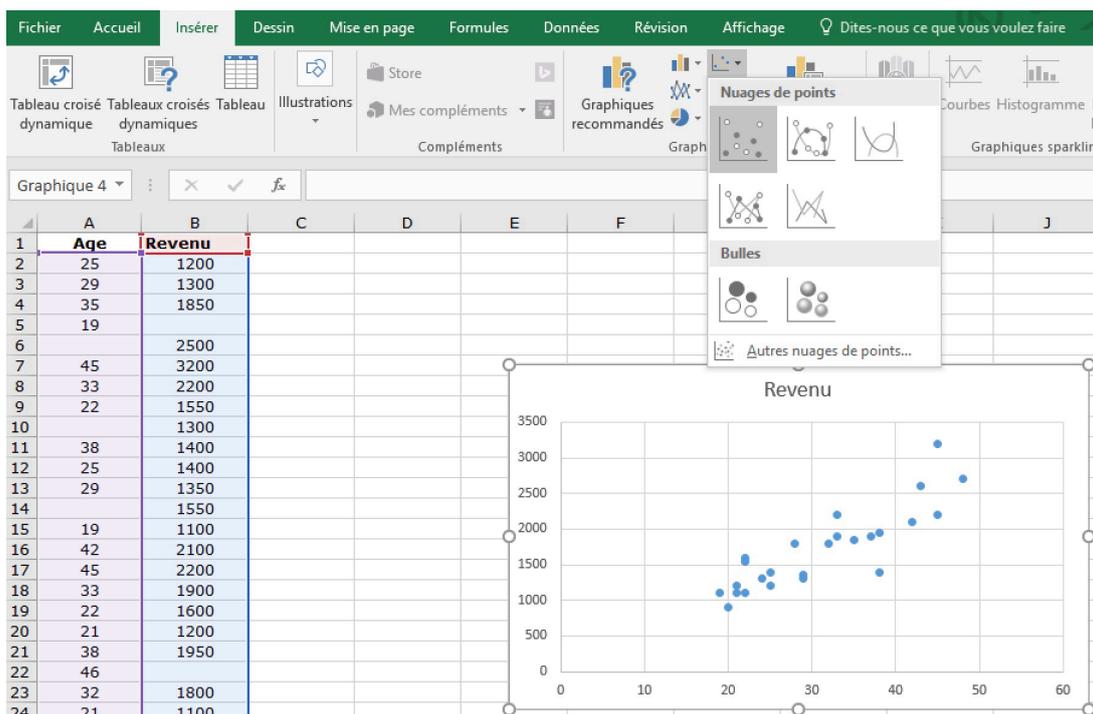
La matrice obtenue affiche pour chaque couple de variables le coefficient de corrélation linéaire correspondant.



### 6.1.2 Diagramme de dispersion

Pour représenter graphiquement deux variables quantitatives, on utilise un diagramme de dispersion (ou un *nuage de points*) :

1. sélectionner les 2 colonnes concernées ;
2. choisir Insérer > Graphique > Nuages de points en prenant le premier modèle.
3. un clic droit sur les différents éléments du graphique (axes, points, titre) permet de les modifier.



## 6.2 Variables qualitatives

Pour obtenir un tri croisé et le graphique associé de deux variables qualitatives, on utilise le « rapport de tableau croisé dynamique » :

1. sélectionner les données dans l'onglet correspondant (il est suffit de sélectionner les colonnes correspondantes en cliquant sur les entêtes de colonne). Si les colonnes ne sont pas adjacentes, sélectionner aussi les colonnes séparant les deux variables ;
2. choisir Insérer > Tableaux > Tableau croisé dynamique et cliquer sur OK;

3. dans le tableau, faire glisser le nom de la première variable de la **liste des champs** vers la case (**Ligne**) ;
4. faire glisser le nom de la seconde variable de la **liste des champs** vers la case (Colonne) ;
5. faire glisser la variable **Obs** (ou une des deux variables étudiées) de la **liste des champs** vers la case (Valeurs) : le tri croisé s'effectue automatiquement !
6. pour afficher les pourcentages en ligne ou en colonn, sélectionner **Paramètres des champs de valeurs** dans le menu local du nom de la variable de la zone **Valeurs** ; cliquer alors sur Afficher les valeurs, puis dans le menu Afficher les valeurs sélectionner % par ligne ou % par colonne.

	A	B	C	D	E	F
1						
2						
3	<b>Nombre de Obs</b>	<b>Sexe</b>	<b>1</b>	<b>2</b>	<b>Total</b>	
4	<b>SitFam</b>					
5		1	3	2	5	
6		2		5	5	
7		3	3	6	9	
8		4		1	1	
9		5	2		2	
10	<b>Total</b>	<b>8</b>	<b>14</b>	<b>22</b>		
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

Le graphique (de préférence des effectifs) s'obtient en cliquant sur « Graphique croisé dynamique » de la barre d'outils.