

Traitement des données avec Microsoft EXCEL 2010

Vincent Jalby
Faculté de Droit et des Sciences Économiques, Université de Limoges
 vincent.jalby@unilim.fr

Septembre 2013

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

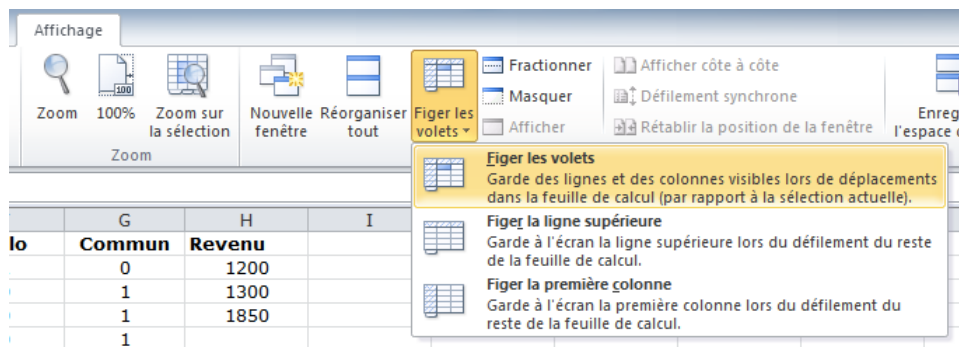
Les réponses manquantes (questions non répondues, ou incorrectement) doivent être codées, soit en laissant la cellule vide, soit par des 9/99/999. L'utilisation d'une cellule vide facilite le traitement avec les fonctions incluses dans Excel (en particulier pour les questions numériques). C'est la méthode à préférer. L'utilisation de 9 est à choisir lorsqu'on prévoit d'exporter les données vers SPSS.

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^{re} colonne (numéro d'observation), 1^{res} 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^e ligne, 2nde 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 !

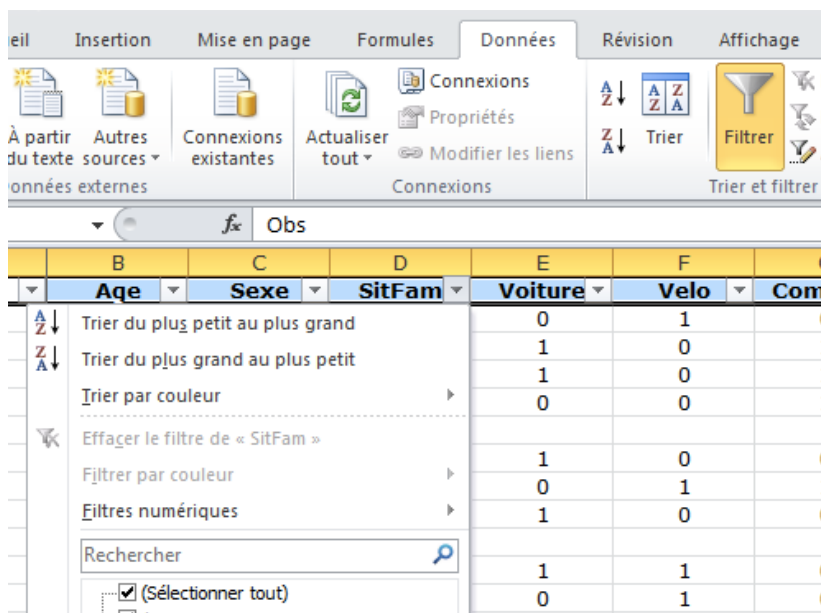
On est alors en présence d'une feuille du type

	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, SOMME.

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 « Macro complémentaires », 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

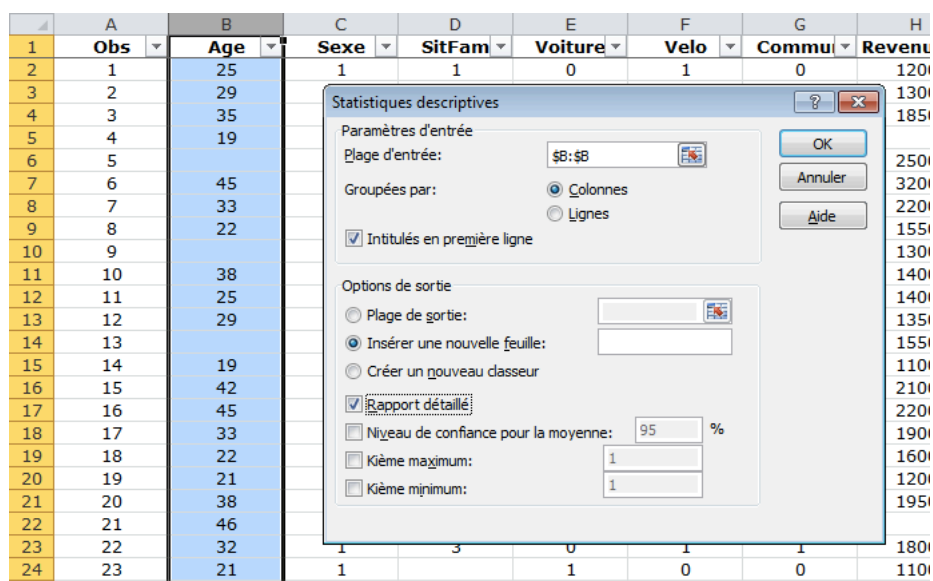
5.1 Variables quantitatives

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 » ;
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.



5.1.2 Représentation graphique : histogramme

Pour obtenir l'histogramme d'une variable quantitative (continue), il est (quasiment) indispensable d'utiliser l'utilitaire d'analyse. Pour cela, il faut auparavant définir les bornes des intervalles (classes) de valeurs servant à la discrétisation.

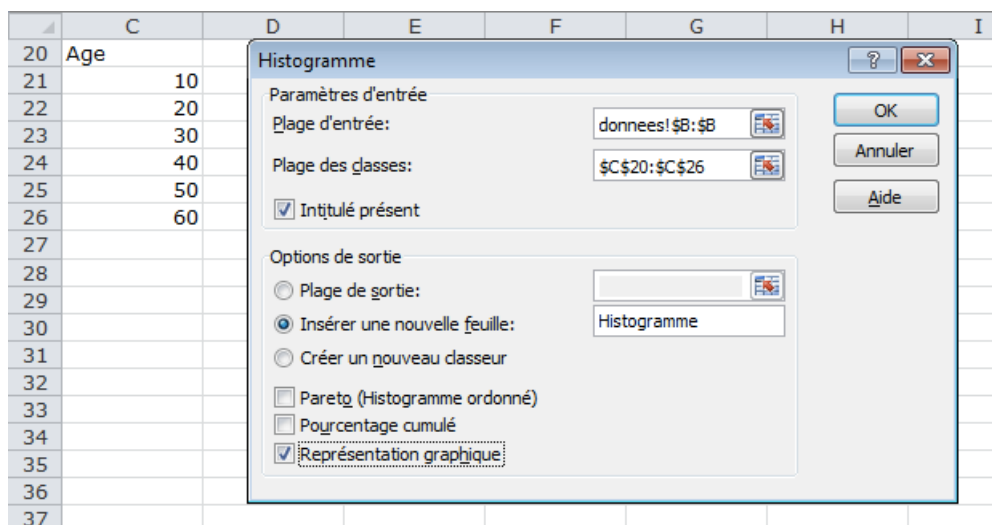
Cette opération doit être faite manuellement (en s'aidant des résultats données par Statistiques descriptives : minimum, maximum, etc). Dans une colonne de la feuille résultats, on tape l'intitulé (« Classes ») suivi de la borne supérieure de chaque classe. L'exemple ci-contre correspond à la saisie nécessaire pour obtenir les classes [0, 10],]10, 20] et]20, 30].

Classes
10
20
30

Remarque. *Il est préférable de choisir des intervalles de longueur régulière (de longueur 10 dans l'exemple précédent).*

Pour obtenir un histogramme, il faut ensuite :

1. sélectionner Données > Analyse > Utilitaire d'analyse, puis Histogramme ;
2. pour la plage d'entrée, sélectionner (dans la feuille données) la colonne correspondant à la variable quantitative à représenter (il suffit pour cela de sélectionner la colonne complète, en cliquant sur la lettre de la colonne) ;
3. de la même façon, sélectionner (dans la feuille résultats les cellules correspondants aux classes définies précédemment (y compris la cellule d'intitulé) ;
4. cocher « Intitulées en première ligne » et « Représentation graphique » ;
5. 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).
6. finalement, ne pas oublier de cocher « Représentation graphique » !



5.2 Variables qualitatives

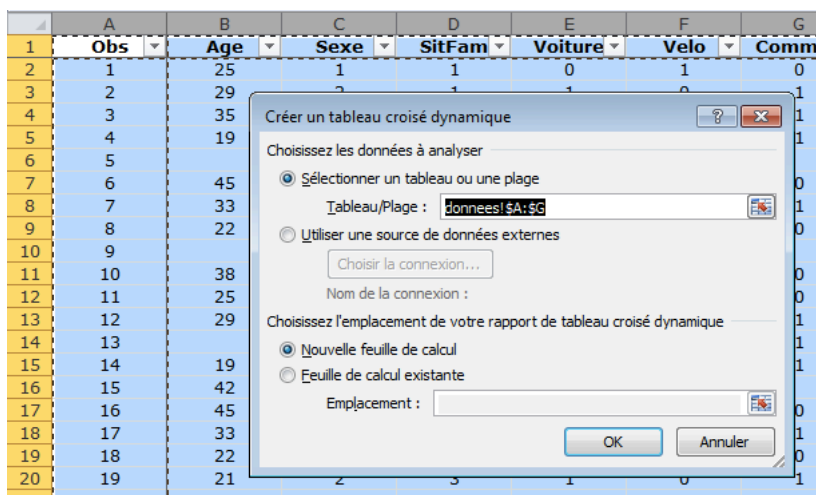
Elles correspondent aux QCM. Elles sont soit ordinales (les modalités sont classées) soit nominales (les modalités ne peuvent pas être classées), en particulier dichotomiques. Certains traitements s'appliquent aussi aux variables quantitatives discrètes (nombre d'enfants, ...).

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 discrète continue).

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

1. sélectionner Insertion > Tableaux > Tableau croisé dynamique;
2. dans le dialogue, définir la plage de données en sélectionnant les colonnes correspondantes dans la feuille données et sélectionner l'emplacement du tableau qui sera créé (« Nouvelle feuille de calcul » pour créer une nouvelle feuille contenant le tableau);



3. faire glisser le nom de la variable de la liste des champs vers la colonne de gauche du tableau (Champs de ligne);
4. faire glisser à nouveau le nom de la variable de la liste des champs vers la zone principale du tableau (Données) : le tri à plat s'effectue automatiquement!

Tableau dynamique : Chaque zone peut recevoir une ou plusieurs variables.

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

Ruban d'outils : permet de définir certaines options ainsi que l'apparence du tableau. Le bouton « Graphique » permet d'obtenir un graphique dynamique.

5. changer ensuite dans la première colonne les numéros des modalités par leur nom;
6. pour masquer la modalité « vide » (= donnée manquante¹), il suffit de cliquer dans le menu local de la colonne 1 et de décocher « vide ».
7. pour afficher les pourcentages, faire un clic droit dans la partie principale du tableau (colonne de droite), puis sélectionner Paramètres de champ; cliquer alors sur Afficher les valeurs, puis dans le menu Afficher les valeurs sélectionner % du total;

1. Cette modalité n'affiche aucune valeur. En effet, les valeurs manquantes « vides » ne sont pas prises en compte. Si nécessaire, pour remédier à cela, il faut sélectionner au point 3, outre la colonne correspondant à la variable étudiée, la première colonne correspondant au numéro d'observation. C'est cette variable qu'il faudra faire glisser (au point 6) vers la zone principale du tableau.

Déposer champs de filtre de rappo

Menu local : permet de choisir les modalités à afficher.

Clic droit : paramètres de Champ

Modalités : numéro à remplacer par intitulé.

Synthèse par Nombre : affiche le nombre (effectif) de chaque modalité

% du total : affiche le % (au lieu de l'effectif), par rapport au total (de l'échantillon, ou de la ligne ou de la colonne.

Paramètres des champs de valeurs

Nom de la source : Obs

Nom personnalisé : Nombre de Obs

Synthèse des valeurs par Afficher les valeurs

Afficher les valeurs

Aucun calcul

Aucun calcul

% du total général

% du total de la colonne

% du total de la ligne

% de

% du total de la ligne parente

Venture

Velo

Format de nombre

OK

Annuler

5.2.2 Diagramme en bâtons

A partir du tableau croisé dynamique : 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.

Directement : il suffit de sélectionner la zone de données (les deux colonnes contenant, pour la première le nom des modalités et pour la seconde les effectifs ou pourcentages correspondants, y compris la ligne d'intitulé, mais hors ligne total), puis de choisir Insertion > Graphique.

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 2, on sélectionne l'ensemble des colonnes/variables 0/1 correspondantes aux réponses à la question.
- au points 3 et 4, ne pas faire glisser de variable dans la première colonne, mais faire glisser toutes les variables (une par une) vers la zone principale du tableau.
- faire ensuite un clic droit sur chaque ligne puis sélectionner Paramètres de champ et 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.

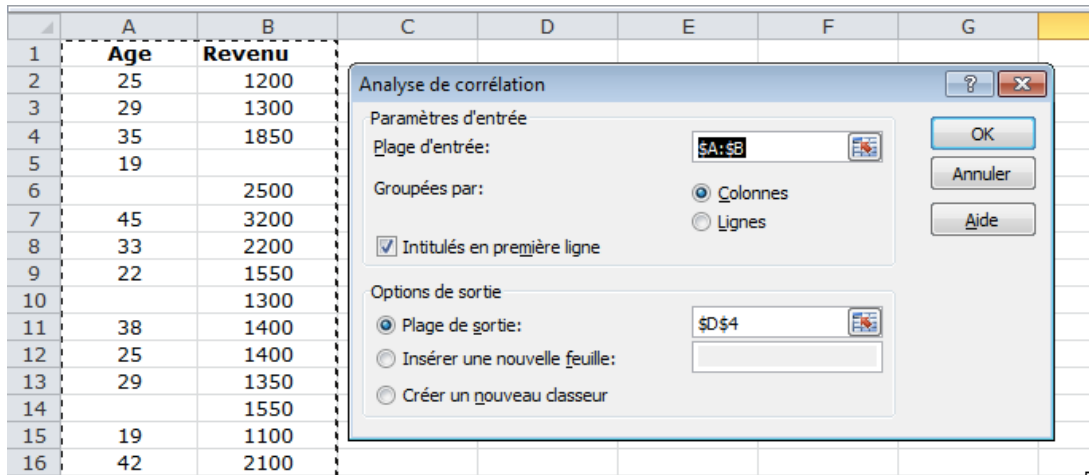
6.1.1 Coefficient de corrélation linéaire

Pour calculer le coefficient de corrélation linéaire, il est nécessaire d'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 `resultats` (les résultats seront affichés à partir de cette cellule, en dessous et à droite).

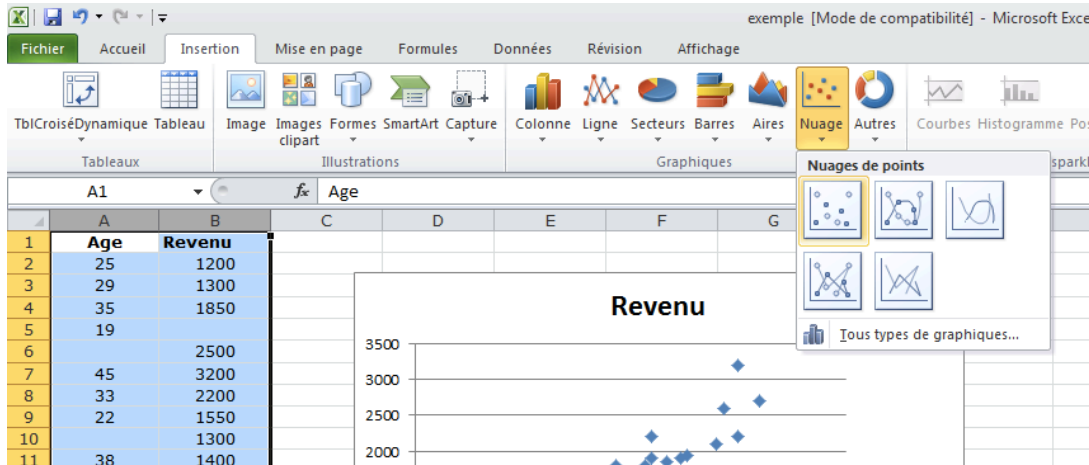
La matrice obtenue affiche pour chaque couple de variables le coefficient de corrélation linéaire correspondant.



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 Insertion > 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. dans le dialogue, définir la plage de données en sélectionnant les colonnes correspondantes dans la feuille `donnees` et sélectionner l'emplacement du tableau qui sera créé. Si les colonnes ne sont pas adjacentes, sélectionner aussi les colonnes séparant les deux variables ;
2. dans le tableau, faire glisser le nom de la première variable de la liste des champs vers la colonne de gauche du tableau (Champs de ligne) ;
3. faire glisser le nom de la seconde variable de la liste des champs vers la première ligne du tableau (Champs de colonne) ;

4. faire glisser à nouveau le nom d'une des deux variables² de la liste des champs vers la zone principale du tableau (Données) : le tri croisé s'effectue automatiquement !
5. changer ensuite dans la première colonne et la première ligne 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/ligne 1 et de décocher « vide ».
7. pour afficher les pourcentages en ligne ou en colonne, faire un clic droit dans la partie principale du tableau, puis sélectionner Paramètres de champ ; cliquer alors sur Afficher les valeurs, puis dans le menu Afficher les valeurs sélectionner % par ligne ou % par colonne.

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

2. ou mieux : faite glisser la variable « numéro d'observation » ; cela permettra d'avoir le nombre de valeurs manquantes.