

Traitement de données avec tableur appliqué à l'Economie et la Gestion.



1.0

AUTEURE : MARTINE CADOT, AGRÉGÉE DE MATHÉMATIQUES
À L'UNIVERSITÉ HENRI POINCARÉ (NANCY 1)

EXPERT ÉCONOMISTE : DENIS ABECASSIS, PROFESSEUR
D'ECONOMIE À L'UNIVERSITÉ DE NANTERRE (PARIS X)

COPYRIGHT : PÔLE DE RECHERCHE ET D'ENSEIGNEMENT
SUPÉRIEUR DE L'UNIVERSITÉ DE LORRAINE

Février 2010





Table des matières

Table des matières	3
I - Leçon 1 : Le tableur et ses échanges avec son environnement	9
A. La place du tableur dans son environnement.....	9
B. Les formats de fichiers.....	10
C. Importation de données.....	13
1. Tableau de données importé d'un butineur par un « copier-coller ».....	14
2. Données importées dans le tableur depuis un fichier *.txt.....	16
D. Exportation de données.....	19
E. L'écriture des nombres selon la langue.....	19
F. Exercices appliqués à l'économie avec correction.....	20
1. Problème 1 : Calcul d'un indice des prix.....	20
2. Problème 2 : Calcul du produit intérieur brut, de la valeur ajoutée et des importations par habitant (Import de données dans un modèle macroéconomique).....	22
G. QCM de la leçon 1.....	22
II - Leçon 2 : Les cellules de la feuille de calcul	23
A. Le classeur : un ensemble structuré de cellules.....	23
1. L'organisation par défaut des cellules.....	23
2. L'accès aux feuilles de calcul.....	24
3. Exercice : création d'un classeur contenant 2 feuilles de données à partir de 2 fichiers en format *.txt.....	26
B. Quelques « raccourcis clavier » utiles des classeurs.....	28
C. Adressage.....	29
1. Désignation d'une cellule.....	29
2. Adressage absolu/relatif.....	30
3. Exercice : Adressage.....	38
4. Exercice 2.1.....	38
5. Cascades de mises à jour de cellules.....	40
D. Formats de cellules.....	42
1. Mise en forme de la cellule.....	44
2. Format de nombre.....	45
3. Format de date.....	46
4. Exercice 2.2 : Trouver le jour de la semaine correspondant à une date donnée.....	47
5. Exercice 2.3 : Trouver le jour de la semaine du 1er mai et du 1er janvier.....	47



<i>pendant 10 ans</i>	47
E. Saisie, édition	47
1. <i>Saisie simple du contenu d'une cellule</i>	47
2. <i>Alternative pour la saisie du contenu d'une cellule contenant des références</i>	48
3. <i>Adressage en dehors de la feuille active (autre feuille, autre classeur)</i>	50
4. <i>Saisie de la référence à une plage de cellules</i>	50
F. Reproduction de cellules	53
1. <i>Le copier-coller</i>	53
2. <i>Exercice 2.4</i>	53
3. <i>La copie par « étirage »</i>	54
4. <i>Les autres recopies</i>	55
G. Exercices appliqués à l'économie avec correction	57
1. <i>Problème 3 : Calcul de remboursements de crédits</i>	57
2. <i>Problème 4 : Calcul de coûts</i>	58
3. <i>Problème 5 : Calcul d'un indice chaînes des prix (suite de l'exercice 1 de la leçon 1)</i>	59
4. <i>Problème 6 : Modèle macroéconomique</i>	61
H. QCM de la leçon 2	61

III - Leçon 3 : Formules élémentaires, conditionnelles, fonctions, calcul matriciel **65**

A. Qu'est-ce qu'une formule ? Comment l'écrit-on ?	65
1. <i>Définitions succinctes des divers types de formules</i>	65
2. <i>Questions et réponses</i>	66
3. <i>Édition d'une formule</i>	66
B. Formule élémentaire	67
1. <i>Opérateurs arithmétiques</i>	67
2. <i>Opérateurs de comparaison</i>	69
3. <i>Opérateurs de texte</i>	71
4. <i>Opérateurs de référence</i>	71
5. <i>Combinaison de plusieurs opérateurs dans une même expression</i>	72
C. Les fonctions	74
D. Formule conditionnelle	77
1. <i>La fonction SI</i>	77
2. <i>Exercice 1</i>	78
3. <i>La fonction NB.SI</i>	79
4. <i>Exercice 2</i>	80
5. <i>La fonction SOMME.SI</i>	80
6. <i>Exercice 3</i>	81
7. <i>Exercice 4</i>	81
E. Calcul matriciel	81
1. <i>Définitions et manipulations</i>	82
2. <i>Opérateurs et fonctions matricielles</i>	83
3. <i>Exercice TD</i>	86
F. Exercices appliqués à l'économie avec correction	87
1. <i>Problème 7 : Entreprise, deux usines, effet de structure</i>	87
2. <i>Problème 8 : Règles conditionnelles</i>	88
3. <i>Problème 9 : Calcul des prix de revient à partir d'Inputs</i>	90
4. <i>Problème 10 : Fichier d'une enquête sur le premier salaire</i>	91
G. QCM de la leçon 3	92

IV - Leçon 4 : Les outils supplémentaires de bureautique **99**



A. Exercice TD.....	99
B. Les tris, les filtres.....	100
1. Les tris.....	100
2. Exercice TD 1.....	102
3. Les filtres.....	103
C. Les sous-totaux et tableaux croisés dynamiques.....	104
1. Les sous-totaux.....	104
2. Le tableau croisé dynamique.....	106
3. Exercice TD 2.....	107
4. Exercices.....	116
D. La recherche de valeurs cibles, le solveur.....	116
1. La recherche de valeurs cibles.....	116
2. Exercice TD 3.....	117
3. Exercice 3.....	118
4. Le solveur.....	118
5. Exercice TD 3.2.....	120
E. Exercices appliqués à l'économie avec correction.....	121
1. Problème 11 : Maximisation du profit.....	121
2. Problème 12 : Données chronologiques.....	122
3. Problème 13 : Enquête premier salaire.....	124

V - Leçon 5 : Les diverses représentations graphiques des données **125**

A. Introduction.....	125
B. Diagrammes de statistiques descriptives.....	126
1. Introduction.....	126
2. Une série de valeurs numériques, répartition en groupes selon une variable qualitative.....	126
3. Exercice1.....	130
4. Une série de valeurs numériques, répartition en groupes selon deux variables qualitatives.....	130
5. Exercice2.....	134
6. Deux variables quantitatives.....	135
7. Exercice3.....	136
8. Exercice4.....	136
9. Plus de variables quantitatives et qualitatives.....	136
C. Une modélisation des données « à la main » utilisant le côté dynamique des graphiques.....	137
D. Exercice TD : Ajustement graphique de lois de probabilités à des données.....	138
1. Les données sont continues.....	138
2. Les données sont discrètes.....	147
E. Tendances pour modéliser les données.....	150
1. Les moyennes mobiles.....	150
2. La régression linéaire.....	152
3. Exercice TD.....	155
F. Exercices TD pour la leçon 5.....	163
1. Modèle 1 : Le salaire en fonction de l'ancienneté.....	163
2. Modèle 2 : La régression linéaire à variables qualitatives.....	167
3. Modèle 3 : La régression linéaire multiple à deux variables mixtes.....	168
4. Modèle 4 : La régression linéaire multiple avec 2 variables qualitatives et une quantitative.....	169
5. Modèle 5 : La régression linéaire multiple avec interaction entre 2 variables.....	169



<u>G. Exercices appliqués à l'économie avec correction.....</u>	<u>170</u>
<u>1. Problème 4 : Calcul de coûts.....</u>	<u>170</u>
<u>2. Problème 14 : Offre et demande : équilibre du marché.....</u>	<u>172</u>
<u>3. Problème 15 : Données chronologiques.....</u>	<u>173</u>
<u>4. Problème 16 : Enquête premier salaire.....</u>	<u>175</u>
Solution des exercices de TD	177
Solution des Quiz	219
Glossaire	225

Objectifs



La ressource proposée présente l'utilisation d'un tableur (deux tableurs sont considérés : Microsoft Excel et OpenOffice Calc) pour traiter des données ou mettre en œuvre des techniques classiques de résolution de problèmes issues des domaines de l'Economie et de la Gestion. L'accent est mis sur le conditionnement des éléments du problème et des données numériques ou textuelles associées, afin d'assurer un traitement et une visualisation adaptées à l'aide du tableur. De plus, l'aspect informatique est mis en exergue à travers des procédures typiques de manipulation de fichiers, d'import/export de données, etc.



Leçon 1 : Le tableur et ses échanges avec son environnement

La place du tableur dans son environnement.	9
Les formats de fichiers	10
Importation de données	13
Exportation de données	19
L'écriture des nombres selon la langue	19
Exercices appliqués à l'économie avec correction	20
QCM de la leçon 1	22

A. La place du tableur dans son environnement.

Actuellement le tableur est un des éléments d'une « suite bureautique » : le tableur OpenCalc fait partie d'OpenOffice (Fig. 1.1), le tableur Excel de MS-Office (Fig 1.2). Le choix d'une suite bureautique par l'utilisateur d'un ordinateur dépend entre autres choses du système d'exploitation de l'ordinateur, Windows, Linux, MacIntosh étant les plus courants. Les deux sortes de tableurs, très voisines, auxquelles nous nous référerons ici, sont MS-Office disponibles sous Windows et MacIntosh, et OpenOffice sous Windows, Mac, Linux. Les versions utilisées de MS-Office et OpenOffice ne sont pas toujours les dernières. Elles varient selon les personnes, particuliers ou entreprises, les mises à jour n'étant pas toujours faites au fur et à mesure. Heureusement, les modifications d'une version à l'autre ne remettent pas en cause les principes généraux que nous étudions dans ce cours, qui restent communs à toutes les versions de tableur existant depuis plus de dix ans. De plus, les versions successives d'un même tableur assurent une compatibilité ascendante (on peut lire des fichiers écrits dans une version précédente à l'aide de la dernière version, mais l'inverse n'est possible que si on respecte certaines conditions, que nous verrons).



Figure 1.1 : Version 2.4 d'OpenOffice et son tableur

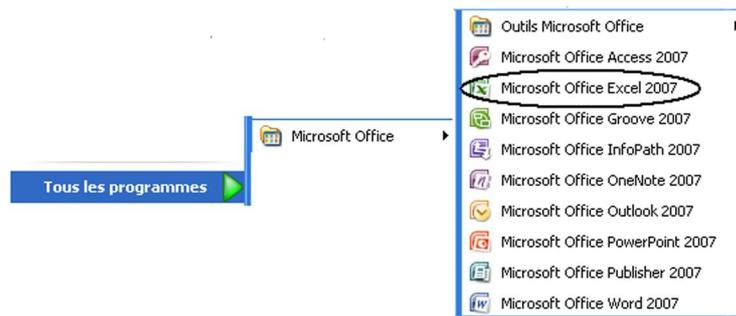


Figure 1.2 : Version 2007 de MS-Office et son tableau

L'installation du tableur sur un ordinateur se fait habituellement en même temps que les autres éléments de la suite, plus ou moins nombreux selon la version, contenant au minimum le traitement de textes (fig. 1.1 : OpenWriter pour OpenOffice, fig. 1.2 : Word pour MS-Office). On dispose alors d'une certaine compatibilité au sein d'une même suite bureautique, par exemple on peut faire des « copier-coller* » de tableaux, de dessins ou graphiques, du tableur vers le traitement de textes ou inversement. L'opération de « copier-coller » permet ainsi d'échanger des données entre deux applications (ici tableur et traitement de textes) ouvertes simultanément sur un même ordinateur sans se poser de questions sur la façon dont elles sont codées. Mais en dehors de cette manipulation très particulière, une connaissance minimale du codage des données est nécessaire à l'utilisateur averti du tableur, elle concerne les formats de fichiers.

B. Les formats de fichiers

Quand on saisit des nombres dans un tableur, du texte dans un traitement de textes, les données sont situées en mémoire centrale, et si l'ordinateur s'éteint, ou l'application se ferme, les informations sont perdues. Si on désire garder les informations pour les utiliser à nouveau, les compléter, les transformer, on fait une sauvegarde sous forme d'un fichier. Quand on sauvegarde son travail pour la première fois, en utilisant la commande « Enregistrer » du menu « fichier » (voir Fig. 1.3 la suite de sous-menus à dérouler qu'on représentera par « Fichier>Enregistrer »), ou en cliquant sur l'icône de la disquette, une fenêtre apparaît dans laquelle on peut saisir le nom du fichier (en général on remplace « classeur1 » par un nom plus indicatif du contenu du fichier). On peut également choisir un autre type de fichier que celui proposé par défaut (*.xls).

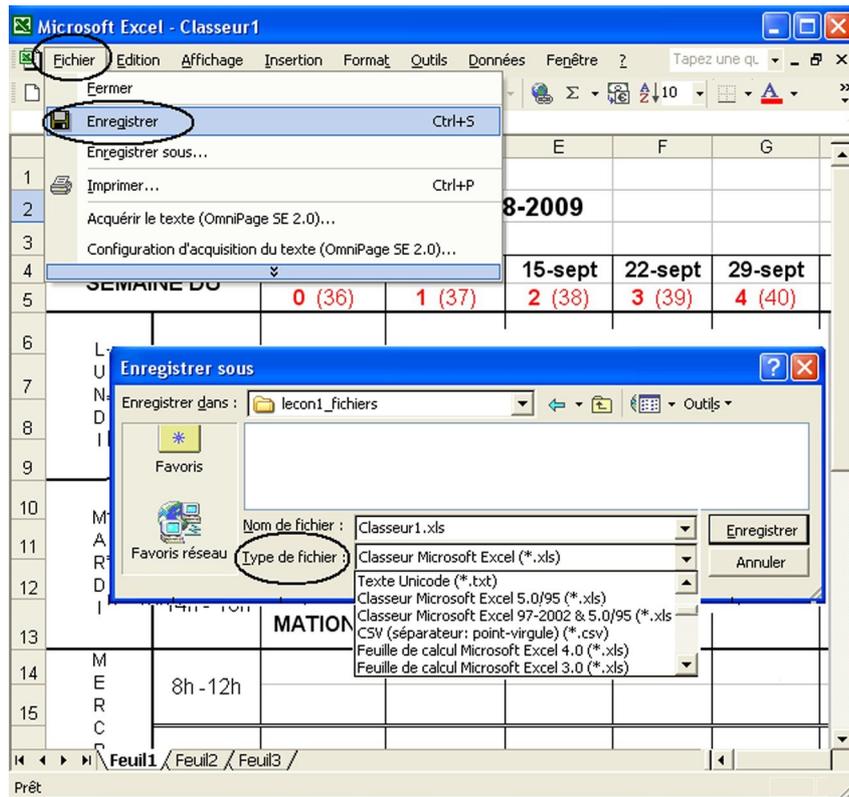


Figure 1.3 : Première sauvegarde sous Excel 2002 d'un emploi du temps en cours de réalisation

Lors des sauvegardes suivantes, la fenêtre permettant les modifications de nom de fichier et de type de fichier ne s'ouvrira que si on choisit la commande « Enregistrer sous ». Dans la figure 1.3, on aperçoit différents types de fichiers proposés par Excel 2002. Chaque type est repéré par son extension*, habituellement formée de 3 caractères, et le nom complet du fichier comporte généralement un intitulé suivi d'un point et de l'extension, par exemple EmploiTemps.xls. Le choix du type de fichier détermine la quantité d'informations ajoutées aux données saisies par l'utilisateur. Les formats « texte » (d'extension .txt, indiqué par la notation *.txt, où le symbole * est un « Joker », qui peut être remplacé par un nombre quelconque de caractères) sont ceux qui en contiennent le moins, voire pas du tout, les données pouvant alors être copiées telles quelles dans diverses applications. Par exemple, la figure 1.4 montre l'effet d'une sauvegarde de l'emploi du temps en cours de réalisation dans le tableau selon la figure 1.3. Seules les informations textuelles sont restées.

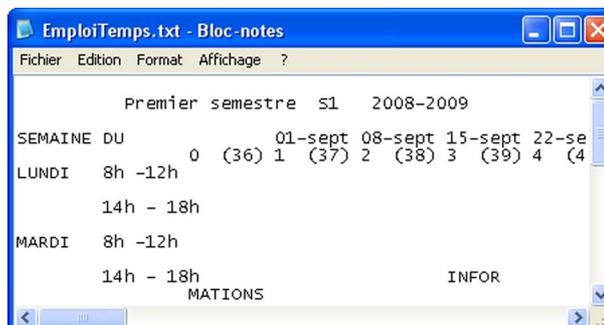


Figure 1.3

A l'autre extrême, les formats par défaut de l'application (ici *.xls) sont ceux qui

contiennent le plus d'informations possibles afin de pouvoir restaurer à l'identique l'état de l'application avant la sauvegarde. Parmi ces informations ajoutées citons la police de caractères, leur taille, leur couleur, leur orientation, mais également toutes les informations sur les graphiques ou les images présentes, les formules, les fonctions ajoutées, le type d'impression souhaité, la date, ainsi que parfois, le numéro de version de l'application, et des informations plus personnelles comme le numéro de licence, le nom de l'ordinateur, de son propriétaire. Une fois les données sauvegardées selon le format dédié à l'application, un double-clic* sur le fichier a pour effet de lancer à nouveau l'application et de récupérer son état avant sauvegarde. Il est également possible d'obtenir le même effet en lançant l'application (atteignable par le chemin visible sur les figures 1.1 qu'on représentera par Démarrer>Tous les programmes>MicroSoft Office> MicroSoft Office 2002) et en « ouvrant » le fichier (en suivant les menus et sous-menus. Le codage des informations supplémentaires est déterminé par l'application, ce qui rend difficile l'ouverture du fichier par une autre application.

Dans la figure 1.5, on voit les informations du fichier *EmploiTemps.xls* transcrites par un simple éditeur de textes, le « bloc-notes » de Windows (atteignable par le chemin Démarrer>Tous les programmes>Accessoires>Bloc-notes). Au lieu d'interpréter les informations lues, il les a recopiées à sa façon, le résultat étant inexploitable. Ainsi on voit apparaître les formats des nombres sous Excel, comme « -#,##0 », inutilisables par le bloc-notes qui n'a pas vocation de différencier textes et nombres, et la liste des polices de caractères du texte d'origine, comme « Arial », que le bloc-notes ne peut utiliser car il est limité à l'emploi d'une seule police de caractères dans un même texte.

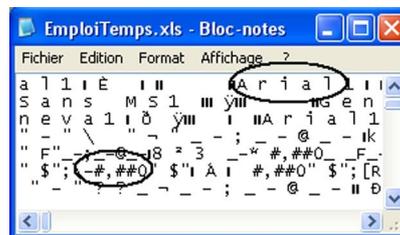


Figure 1.5 : Ouverture avec le « bloc-notes » du fichier en format « xls »

Pour conclure, avant de sauvegarder un travail fait avec un tableur, et plus largement avec une application, il faut réfléchir à sa diffusion. S'il est destiné à de nombreuses personnes, il y a peu de chances qu'elles disposent de la même version de l'application, voire de la même application. Notamment un travail utilisant les nouveautés de la dernière version ne sera totalement récupéré que s'il est lu depuis un ordinateur sur lequel cette dernière version est installée. Par exemple la version 2007 de Microsoft Excel produit des fichiers *.xlsx. L'utilisateur souhaitant qu'ils soient ouverts par les possesseurs des versions précédentes, doit veiller à sauvegarder son travail en choisissant un des types de fichiers *.xls.

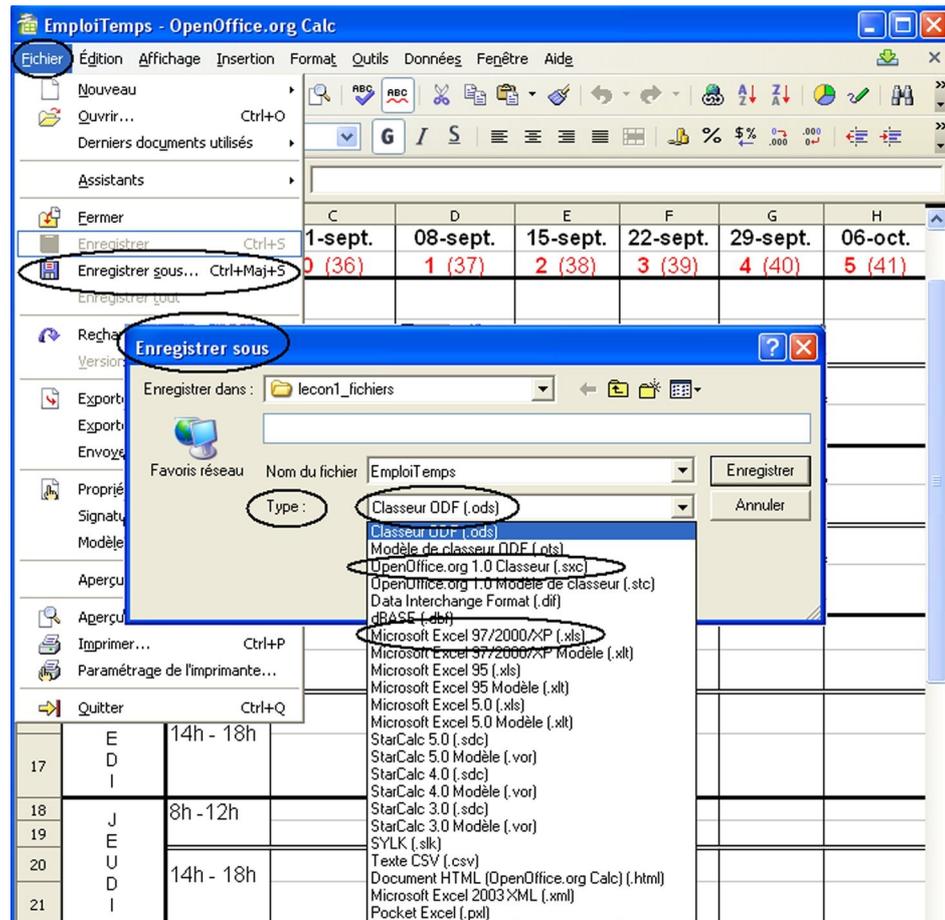


Figure 1.6 : Sauvegarde sous OpenOffice 2.4

En ce qui concerne le transfert de fichiers entre des tableurs de deux suites bureautiques différentes, il est possible à condition de sauvegarder sous un format approprié choisi dans le menu déroulant de « type de fichier ». Par exemple (Fig 1.6), un travail fait sous OpenOffice peut être sauvegardé en format *.xls au lieu du format par défaut (*.sxc pour OpenOffice1.0, *.ods pour OpenOffice2.4), ce qui permettra de l'ouvrir sous Excel en récupérant le maximum d'informations.

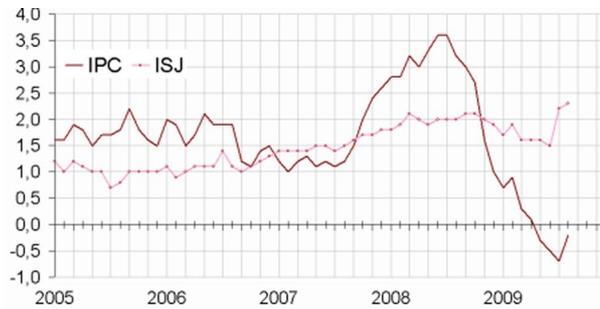
C. Importation de données

On a vu qu'il était possible de travailler à partir d'éléments apparaissant dans une autre fenêtre (par un copier-coller), ou présents dans des fichiers en utilisant le sous-menu « ouvrir » du menu « fichier ». Quand la fenêtre est celle produite par une autre application, par exemple des données visibles dans le « butineur¹ », comme Internet-Explorer, Mozilla, ou quand le fichier est produit par une autre application, on parle alors « d'importation » de données.

Un exemple de données récupérées depuis le site de l'INSEE en allant sur le site <http://www.indices.insee.fr/>¹, en tapant « indice des prix » dans la fenêtre de recherche, on arrive sur une page concernant les indices de prix à la consommation (IPC), on clique alors sur l'hyperlien permettant d'accéder à une page particulière,

1 - <http://www.indices.insee.fr/>

ici l'indice des prix d'août 2009 (Le chemin à suivre pour y accéder est indiqué en haut de la page : [Accueil](#)² > [Thèmes](#)³ > [Conjoncture](#)⁴ > [Indicateurs...](#)⁵ > [Indice des ...](#)⁶ > Les prix à la consommation augmentent de 0,5 % en août 2009).



Source(s) : Insee, Indice des prix à la consommation

:: Pour en savoir plus

- [Données complémentaires \(xls, 45 Ko\)](#)
- [Indices conjoncturels de regroupements : août 2009 \(pdf, 26 Ko\)](#)

:: Liens

- [Questions/réponses](#)
- [Méthodologie détaillée](#)
- [Consulter toutes les séries](#)
- [Pouvoir d'achat de l'euro et du franc](#)
- [Comment revaloriser une pension alimentaire](#)
- [Simuler un indice des prix personnalisé](#)

Variations des indices de prix détaillés

base 100 : année 1998

Regroupements conjoncturels	Pondérations 2009	Variations (en %) au cours		
		Indice Août 2009	du dernier mois	des 12 derniers mois
a) Ensemble des ménages France				
ENSEMBLE (00 E)	10000	119,66	0,5	-0,2
Ensemble CVS (00 C)	10000	119,51	0,5	-0,2
ALIMENTATION (4000 E)	1654	122,19	-0,7	-0,7
Produits frais (4001 E)	205	110,17	-7,5	-9,6

La figure 1.7 est un extrait de la page trouvée (à ce moment-là). Si on désire récupérer des données, on peut procéder de deux façons :

- sélectionner une partie du tableau en bas de la figure et la recopier par un « copier-coller* » vers le tableur.
- récupérer un fichier contenant les données en cliquant sur l'icône correspondante (en haut à droite).

1. Tableau de données importé d'un butineur par un « copier-coller »

On voit les résultats produits par un « copier-coller » de ce tableau dans les tableurs Excel (Fig. 1.8) et OpenOffice 2.4 (Fig. 1.9). Selon la compatibilité entre le tableur et le butineur, le tableau de données devient une ligne de texte écrite dans une cellule, les « séparateurs » de colonnes étant transformés en un simple espace, ou un tableau ressemblant à celui d'origine, sur plusieurs colonnes.

2 - <http://www.insee.fr/fr/>

3 - <http://www.insee.fr/fr/themes/>

4 - <http://www.insee.fr/fr/themes/theme.asp?theme=17>

5 - http://www.insee.fr/fr/themes/theme.asp?theme=17&sous_theme=1

6 - <http://www.insee.fr/fr/themes/indicateur.asp?id=29>

Regroupements conjoncturels	Pondérations 2009	Index Août 2009	Variations (en %) au cours du dernier mois	des 12 derniers mois	
a) Ensemble des ménages France					
ENSEMBLE (00 E)	10000	119,66	0,5	-0,2	#VALEUR!
Ensemble CVS (00 C)	10000	119,51	0,5	-0,2	#VALEUR!
ALIMENTATION (4000 E)	1654	122,19	-0,7	-0,7	#VALEUR!
Produits frais (4001 E)	205	110,17	-7,5	-9,6	#VALEUR!
Alimentation hors produits frais (4002 E)	1449	123,84	0,2	0,6	1,2
TABAC (4004 E)	172	101,62	0	0,6	1,2

Figure 1.8 : Recopie dans Excel 2002 du tableau présent dans la page Web de l'Insee

Regroupements conjoncturels	Pondérations 2009	Index Août 2009	Variations (en %) au cours du dernier mois	des 12 derniers mois	Ajust d'un calcul
a) Ensemble des ménages France					
ENSEMBLE (00 E)	10000	119,66	0,5	-0,2	0
Ensemble CVS (00 C)	10000	119,51	0,5	-0,2	0
ALIMENTATION (4000 E)	1654	122,19	-0,7	-0,7	0
Produits frais (4001 E)	205	110,17	-7,5	-9,6	0
Alimentation hors produits frais (4002 E)	1449	123,84	0,2	0,6	1,2

Figure 1.9 : Recopie dans OpenOffice 2.4 du tableau présent dans la page Web de l'Insee

A ce problème venant des diverses interprétations des séparateurs s'ajoute le problème lié aux diverses écritures des nombres. Alors que sur la page Web (Fig. 1.7) il ne fait pas de doute qu'il y a des nombres négatifs et positifs dans le tableau, les tableurs (Fig. 1.8 et Fig. 1.9) n'ont pas reconnu comme il convient les nombres négatifs. L'utilisateur vigilant peut être alerté par la position des nombres dans la colonne : les nombres positifs sont alignés à droite, comme le sont par défaut les nombres dans un calculateur, alors que les nombres négatifs sont alignés à gauche, comme les textes ordinaires. On acquiert la certitude de ce défaut de reconnaissance en ajoutant la colonne à droite en rouge, dans laquelle chaque nombre de la dernière colonne du tableau d'origine (colonne F pour Excel, E pour OpenOffice) a été multiplié par 2. En cas de nombre positif, le résultat est bien le double (ligne 10), mais quand le nombre est négatif, le résultat n'est pas le double, mais 0 (Fig 1.9) ou un code d'erreur (Fig 1.8), comme le résultat qu'on obtiendrait en essayant de multiplier un texte par 2. Si on efface le nombre incriminé et qu'on le ressaisit depuis le clavier, on réalise que son signe '-' est plus petit que le signe '-' initial. Les deux symboles ne sont pas les mêmes : le code ASCII du signe '-' des nombres est 45 alors que celui présent dans le tableau d'origine est 150 (Fig.

1.10).

Correspondance entre les codes ASCII et les caractères (fonction 'car' d'Excel et son inverse 'code')													
1	21	41	61	81	101	121	141	161	181	201	221	241	
2	22	42	62	82	102	122	142	162	182	202	222	242	
3	23	43	63	83	103	123	143	163	183	203	223	243	
4	24	44	64	84	104	124	144	164	184	204	224	244	
5	25	45	65	85	105	125	145	165	185	205	225	245	
6	26	46	66	86	106	126	146	166	186	206	226	246	
7	27	47	67	87	107	127	147	167	187	207	227	247	
8	28	48	68	88	108	128	148	168	188	208	228	248	
9	29	49	69	89	109	129	149	169	189	209	229	249	
10	30	50	70	90	110	130	150	170	190	210	230	250	
11	31	51	71	91	111	131	151	171	191	211	231	251	
12	32	52	72	92	112	132	152	172	192	212	232	252	
13	33	53	73	93	113	133	153	173	193	213	233	253	
14	34	54	74	94	114	134	154	174	194	214	234	254	
15	35	55	75	95	115	135	155	175	195	215	235	255	
16	36	56	76	96	116	136	156	176	196	216	236		
17	37	57	77	97	117	137	157	177	197	217	237		
18	38	58	78	98	118	138	158	178	198	218	238		
19	39	59	79	99	119	139	159	179	199	219	239		
20	40	60	80	100	120	140	160	180	200	220	240		

On constate dans la figure 1.10 que des codes différents peuvent produire des caractères très proches. En cas de problème avec une recopie de formule numérique vers le tableur, il faut trouver le code de chacun des caractères la constituant et si l'un deux est au-dessus de 128, le remplacer par le plus ressemblant de code inférieur à 128 devrait résoudre le problème. En effet, le codage des caractères entre 128 et 256 est utile pour le texte car il permet de prendre en compte des particularités régionales comme les voyelles accentuées françaises par exemple, mais les formules mathématiques n'en ont généralement pas car elles sont plus universelles.

Les données de l'Insee sont également disponibles sous le format *.pdf , qui est lu par Adobe Reader (en haut à droite de la figure 1.7). Pour pouvoir les importer dans le tableur, on les a déjà copiées-collées vers le bloc-notes, puis sauvegardées sous le format *.txt (voir Figure 1.11). Les tabulations sont devenues des espaces, mais l'assistant d'importation permet de les restituer. La partie suivante est consacrée à son utilisation.

2. Données importées dans le tableur depuis un fichier *.txt

L'import de données se lance par la séquence de sous-menus Fichier>Ouvrir dans Excel, et par Insertion>Feuille dans Open-Office. Puis il convient de choisir le fichier et son type (Fig 1.12). L'assistant d'importation se lance quand on choisit un type de fichier texte. Il propose alors de délimiter les colonnes de deux façons : soit en leur donnant une largeur fixe (la même pour chaque ligne, mais qui peut différer selon les colonnes) en déplaçant à la souris des lignes verticales (Fig 1.13), soit en choisissant le (ou les) séparateurs de colonnes (Fig 1.14). On choisit la deuxième option, puis l'assistant invite l'utilisateur à choisir la position future de la feuille issue de ce fichier avant de réaliser l'import.

Leçon 1 :

The image shows a PDF document titled "Indices_conjoncturels_regroupements_aout_2009.pdf" in Adobe Reader. The document content is a table of monthly price indices for France, base 100 in 1998, for August 2009. The table includes columns for designations, weights, and indices for the months of 2008 and 2009, along with percentage variations. A Notepad window titled "Insee_prix_consommation_aout_2009.txt" is open over the PDF, showing the same table content that has been copied from the PDF.

Désignations des regroupements	Pondérations 2009	Indices pour le mois de					Variations (en %) au cours		
		Août 2008	Mai 2009	Juin 2009	Juil. 2009	Août 2009	du dernier mois	des 3 derniers mois	des 12 derniers mois
00 E ENSEMBLE	10000	119,88	119,43	119,58	119,05	119,66	+0,5	+0,2	-0,2
01 E PRODUITS ALIMENTAIRES ET BOISSONS NON ALCOOLISÉES	1499	123,88	125,36	124,57	123,55	122,53	-0,8	-2,3	-1,1
011 E PRODUITS ALIMENTAIRES	1371	124,76	126,30	125,42	124,27	123,05	-1,0	-2,6	-1,4
0111 E Pain et céréales	209	123,83	124,20	124,03	124,12	124,37	+0,2	+0,1	+0,4
0112 E Viande	396	129,41	130,03	129,90	130,06	130,41	+0,3	+0,3	+0,8
0113 E Poissons et crustacés	120	124,96	123,02	123,05	123,54	124,77	+1,0	+1,4	-0,2
0114 E Lait, fromage et oeufs	227	123,84	122,12	121,51	121,17	120,88	-0,2	-1,0	-2,4
0115 E Huiles et graisses	38	131,77	127,93	126,83	126,65	126,29	-0,3	-1,3	-4,2

Figure 1.11 : Copier-coller d'un fichier *.pdf vers un fichier *.txt.

The image shows the OpenOffice Calc application interface. The 'Insertion' menu is open, and the 'Feuille à partir d'un fichier...' option is selected. A file selection dialog is displayed, showing a list of files in the 'lecon1_fichiers' folder. The file 'Insee_prix_consommation.txt' is selected. The file type is set to 'Texte CSV (*.csv;*.txt;*.xls)'. The 'Insérer' button is highlighted.

Figure 1.12 Choix du fichier et de son type à importer dans le tableur OpenOffice

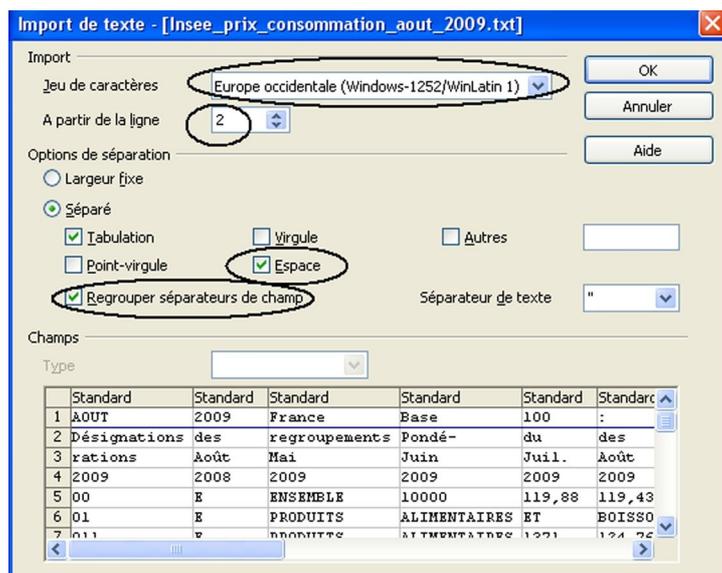


Figure 1.13 : Choix des paramètres de l'import.

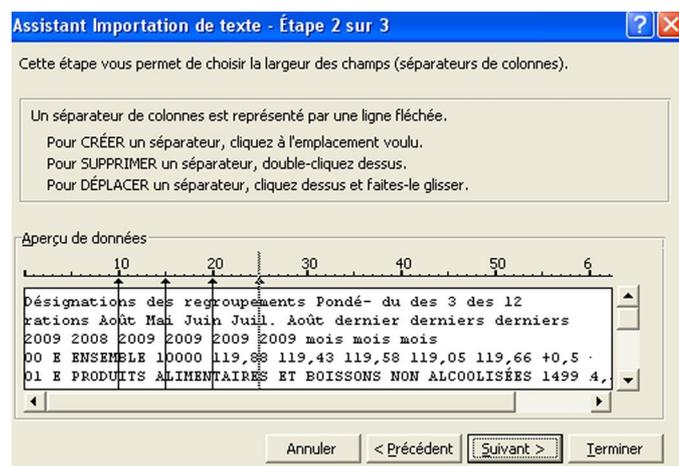


Figure 1.14 : Séparation des colonnes par des traits verticaux en cas de choix de « largeur fixe »



Figure 1.15 : Choix de la position de la feuille contenant les données importées

Le résultat de l'importation n'est pas parfait. Le jeu de caractères a été reconnu correctement (les voyelles accentuées n'ont pas été transformées en d'autres caractères), la première ligne, en dehors des colonnes, a été retirée (on a commencé en ligne 2), les mots et les nombres ont été restitués fidèlement, mais

leurs positions dans les colonnes ne sont pas celles attendues. Dans la figure 1.16, on a entouré tous les éléments successifs d'une colonne du fichier *.pdf de la figure 1.11. Ce résultat surprenant s'explique par le fait que l'espace délimitant les colonnes est de même nature que celui séparant les mots. Et dans la première colonne, le nombre de mots par lignes était variable. Par exemple dans la ligne 7 du tableur, en ligne 7, « Pain et céréales » se retrouve sur 3 colonnes (C, D et E) alors que « Viande » est sur une seule colonne (C). Ainsi le nombre 209 qui suit « Pain et céréales » se retrouve en colonne F alors que le nombre 396 qui suit « Viande » est en colonne D. Pour éviter cela, il aurait fallu importer séparément la colonne d'intitulés et les colonnes de nombres.

	A	B	C	D	E	F	G	H	I
1	Désignations des regroupements		Pondé	du	des		3 des		1.
2	rations	Août	Mai	Jun	Juil.	Août	mois	mois	mois
3		2009	2008	2009	2009	2009	2009		
4	U E		ENSEMBLE	10000	119,88	119,43	119,58	119,05	119,61
5	1 E		PRODUITS ALIMENTAIRES ET			BOISSONS NON	ALCOOLISÉES		149,
6	11 E		PRODUITS ALIMENTAIRES			1371	124,76	126,3	125,42
7	111 E		Pain	et	céréales	209	123,83	124,2	124,0
8	112 E		Viande			396	129,41	130,03	129,9
9	113 E		Poissons	et	crustacés		120	124,96	123,02

Figure 1.16 : Le résultat de l'importation n'est pas parfait : les positions dans les colonnes ne sont pas respectées.

D. Exportation de données

L'exportation des données présentes dans le tableur se fait par un copier-coller vers la fenêtre d'une autre application, éditeur ou traitement de textes, ou par sauvegarde dans un fichier. Comme pour l'importation, il convient de veiller au format des données, et de choisir avec soin le symbole qui remplacera le séparateur de colonnes. Dans le cas d'un copier-coller, c'est la tabulation qui séparera les contenus des colonnes, et dans le cas d'une sauvegarde, la tabulation ou d'autres séparateurs sont proposés à l'utilisateur, comme l'espace ou le point-virgule dans le format *.csv, de façon plus ou moins explicite (voir figures 1.3 et 1.6). Le choix du séparateur doit se faire en fonction de l'utilisation ultérieure des données. Si on veut pouvoir les importer par la suite dans le tableur, et qu'elles contiennent du texte, il est préférable d'utiliser les tabulations comme séparateurs pour éviter de se retrouver dans le cas de la figure 1.16. Si les données ne contiennent que des nombres, le point-virgule peut aussi être utilisé comme séparateur, et l'espace s'il n'est pas utilisé comme séparateur de milliers.

E. L'écriture des nombres selon la langue

On peut basculer d'une écriture à l'autre selon les besoins. Ce changement ne se fait pas dans l'application, mais dans le système d'exploitation. Dans Windows comme dans Linux, il se fait à travers les « options régionales ». On y accède depuis Windows par le menu « Démarrer » (voir figure 1.20). Nous verrons dans le chapitre suivant que le format des nombres et des dates peut être modifié dans le

tableur, sans remettre en cause le symbole décimal (en France, c'est une virgule, et un point aux Etats-Unis ou au Royaume-Uni), le séparateur de milliers (un espace en France, une virgule pour les Anglo-Saxons), et le symbole monétaire (€ pour la France).

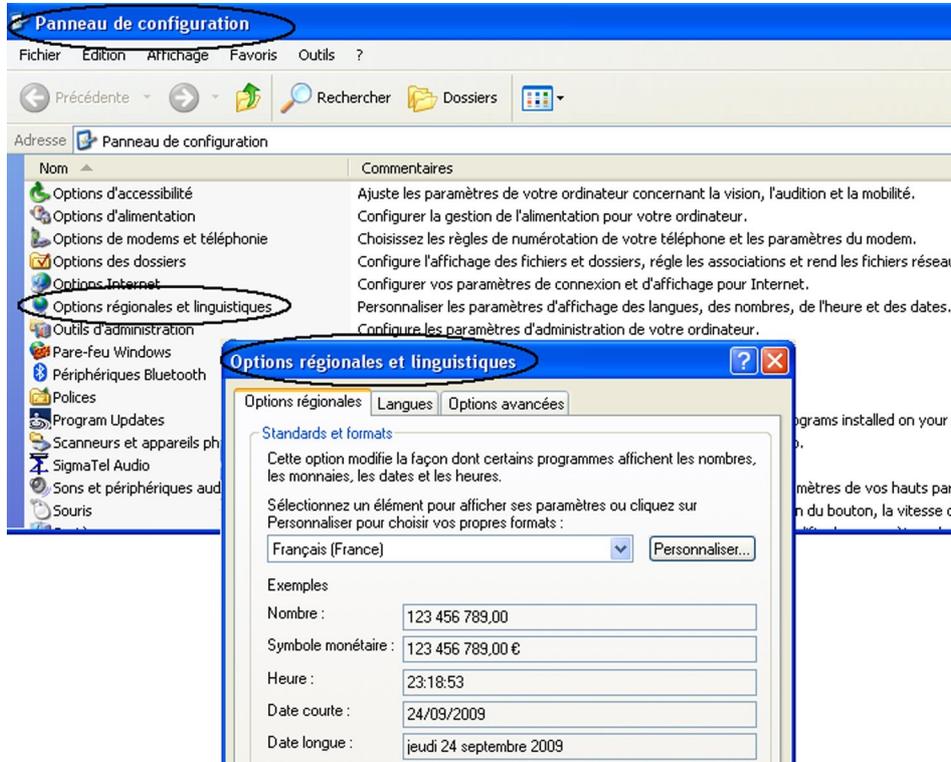


Figure 1.20 : l'écriture des nombres dans le système d'exploitation de l'ordinateur

F. Exercices appliqués à l'économie avec correction

1. Problème 1 : Calcul d'un indice des prix

Durée : 2 heures 30

Le calcul d'un indice des prix est réalisé sur des milliers de produits (des dizaines de fromages, des centaines de légumes (parmi lesquels une dizaine de tomates différentes), des dizaines de viandes, ... Pour simplifier, nous traiterons ici la situation d'une ménagère, madame Dupont, qui achète chaque semaine au marché 5 produits, toujours les mêmes, pour sa famille :

1. Viande
2. Légume
3. Pain
4. Fromage
5. Poisson

Seuls varient les prix et les quantités. La semaine 0, la ménagère remplit son panier avec les produits suivants :

Produit	Quantité	Prix
Viande	2 kg	12 € par kilo
Légume	8 kg	3 € par kilo
Pain	12 baguettes	0,90 € la baguette
Fromage	4 boîtes de 250g	2,50 € la boîte
Poisson	1 kg	10 € le kg

Tableau 1 : Le panier de Mme. Dupont pour la semaine 0

Question 1

[Solution n°1 p 155]

- 1) Quel est le montant de la dépense de madame Dupont la semaine 0 ?
 Quelques semaines plus tard (semaine t) :

Produit	Quantité	Prix
Viande	3 kg	15 € par kilo
Légume	9 kg	2,5 € par kilo
Pain	10 baguettes	1 € la baguette
Fromage	4 boîtes de 250 g	2 € la boîte
Poisson	2 kg	9 € le kg

Tableau 2 : Le panier de Mme. Dupont pour la semaine t

Question 2

[Solution n°2 p 155]

- 2) Quel est le montant de la dépense de madame Dupont la semaine t ?
 Quel est le pourcentage d'augmentation ?

Question 3

[Solution n°3 p 155]

- 3) Une voisine de madame Dupont achète exactement les mêmes quantités que madame Dupont, sans viande mais consomme 2 kg de poisson de plus, sa famille étant végétarienne, que pensez-vous de ses dépenses, pour la semaine 0 et pour la semaine t ?
 Quel est le pourcentage d'augmentation ?

Question 4

[Solution n°4 p 155]

- 4) Par commodité on décide de construire un indice des prix pour le panier de consommation de madame Dupont. Cet indice des prix ne doit pas être affecté par les différences dans les quantités achetées. Comment construire cet indice ?

Question 5

[Solution n°5 p 155]

- 5) On choisira de construire l'indice en tenant compte des quantités de l'année 0 (indice de Laspeyres).
 Calculez cet indice pour le panier de Madame Dupont.

Question 6

[Solution n°6 p 155]

- 6) Suite aux critiques, Madame Dupont considère que cet indice ne reflète pas la réalité, on calculera l'indice selon les quantités de l'année t (indice de Paasche).

Calculez l'indice.

Question 7

[Solution n°7 p 155]

7) Calculez les mêmes indices pour le panier de la voisine, qu'en pensez-vous ?

2. Problème 2 : Calcul du produit intérieur brut, de la valeur ajoutée et des importations par habitant (Import de données dans un modèle macroéconomique)

Durée : 1 heure

Question 1

[Solution n°8 p 155]

1) Dans les comptes de l'INSEE, trouvez la population des pays de la population européenne et importez les données correspondantes.

Question 2

[Solution n°9 p 156]

2) Toujours dans les comptes de l'INSEE, recherchez le décompte du produit intérieur brut* selon les trois approches (production, demande et revenu).

Question 3

[Solution n°10 p 156]

3) Calculez le PIB par habitant en 2008.

Question 4

[Solution n°11 p 156]

4) Calculez la valeur ajoutée par habitant

G. QCM de la leçon 1

Exercice 1

[Solution n°1 p 195]

On désire sauvegarder des données depuis Excel afin de les traiter avec un logiciel fonctionnant sous Dos, et incompatible avec Excel. Parmi les extensions suivantes, correspondant aux formats classiques, laquelle choisissez-vous?

- .xsl
- .txt
- .bmp
- aucun de ceux-là



Leçon 2 : Les cellules de la feuille de calcul

Le classeur : un ensemble structuré de cellules	23
Quelques « raccourcis clavier » utiles des classeurs	28
Adressage	29
Formats de cellules	42
Saisie, édition	47
Reproduction de cellules	53
Exercices appliqués à l'économie avec correction	57
QCM de la leçon 2	61

A. Le classeur : un ensemble structuré de cellules

1. L'organisation par défaut des cellules

Dès le lancement du tableur (en suivant un raccourci ou le chemin montré en Fig. 1.1 ou 1.2), l'utilisateur a directement accès à un espace de stockage de données appelé « classeur », nommé par défaut « Classeur1 » (Fig.2.1) ou « Sans nom1 » (Fig.2.4). Comme on le voit dans la figure 2.1, la saisie d'un nombre, ou d'une formule, peut s'effectuer dès cet instant dans la « cellule » sélectionnée, ici A1, de la « feuille de calcul » sélectionnée (Feuil1). Chaque feuille de calcul contient un grand nombre de lignes et de colonnes avec une cellule à l'intersection de chaque ligne et colonne. Le nombre de lignes et de colonnes, donc de cellules par feuille et le nombre de feuilles disponibles augmentent à chaque nouvelle version du tableur.

Des millions de cellules dans les versions actuelles des tableurs : dans Excel-2007, une feuille de calcul contient 1 048 576 (2^{20}) lignes et 16 384 (2^{14}) colonnes numérotées de A à XFD au lieu de 65 536 lignes (2^{16}) et 256 colonnes (2^8) numérotées de A à IV pour Excel-97 à Excel-2003 et pour le tableur d'OpenOffice 2.4.

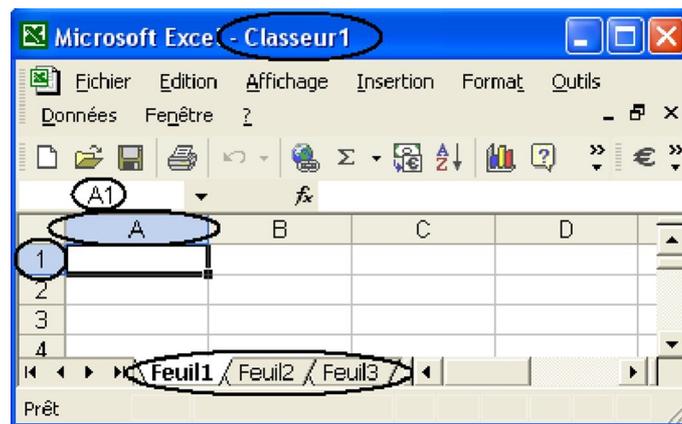


Figure 2.1 : L'état du tableur Excel-2003 au lancement

2. L'accès aux feuilles de calcul

Le nombre de feuilles proposées dans la figure 2.1 est de 3. Pour Excel, les paramètres d'affichage de la première fenêtre peuvent être ajustés selon les besoins de l'utilisateur après avoir enfoncé l'onglet « Général » de la fenêtre « Options » (Fig. 2.2) accessible depuis le menu « Outils » (Fig. 2.1) soit Outils>Options>Général. Le nombre maximal de feuilles par nouveau classeur est de 255 pour cette version d'Excel.

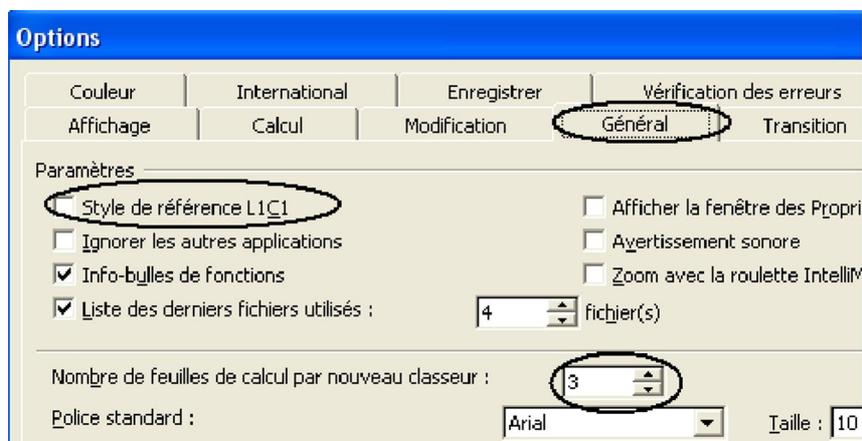


Figure 2.2 : Comment changer le nombre de feuilles par défaut selon Excel-2003

Dans OpenOffice 2.4 également, la commande Options du menu Outils permet de paramétrer de façon détaillée le tableur. Toutefois, on ne retrouve pas exactement les mêmes possibilités d'un tableur à l'autre, notamment le choix du nombre de feuilles du nouveau classeur et du style de référence ne peut se faire de la même façon que dans la figure 2.2. Notons que la limitation à 3 feuilles de calcul au départ n'est pas très gênante dans la mesure où l'insertion ultérieure de feuilles supplémentaires est prévue (voir Fig 2.3) dans tous les tableurs.

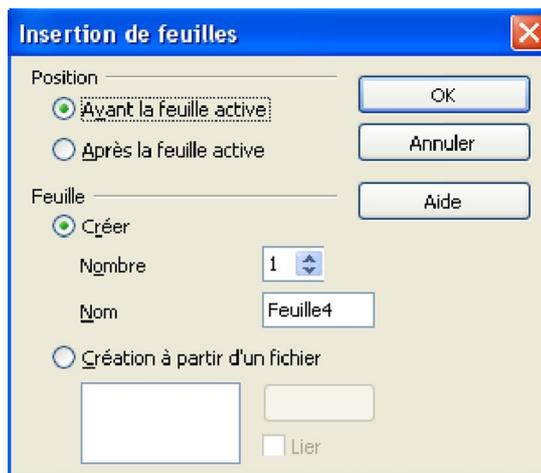


Figure 2.3 : L'insertion de feuilles supplémentaires sous OpenOffice 2.4

Les feuilles de calcul peuvent être déplacées, supprimées, renommées en utilisant le menu contextuel qui s'affiche par un clic droit sur leur onglet (Fig. 2.4).

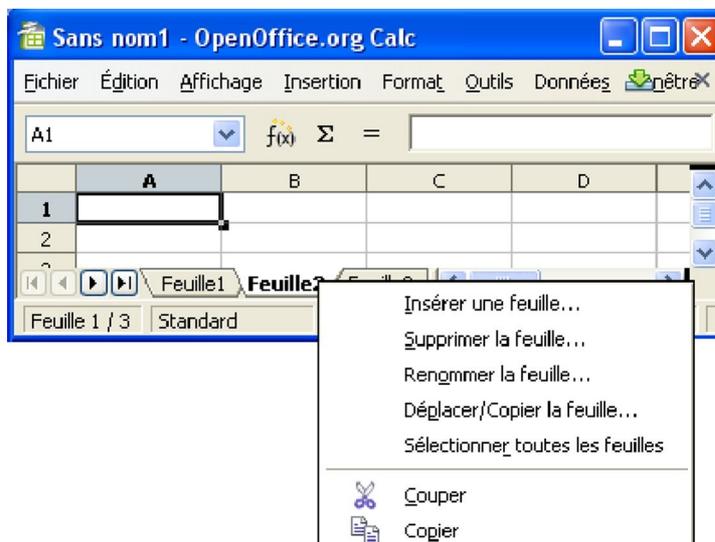


Figure 2.4 : L'insertion de feuilles supplémentaires sous OpenOffice 2.4

La sélection d'une feuille de calcul s'obtient par un clic gauche sur l'onglet correspondant. Il est impossible d'avoir une vue complète de la feuille de calcul dans une même fenêtre, mais on peut en voir plusieurs zones éloignées en utilisant le menu Fenêtre>Fractionner, ou Fenêtre>Figer les volets sous Excel, et Fenêtre>Scinder ou Fenêtre>Fixer sous OpenOffice, qui donne 2 ou 4 aperçus de la feuille selon que cette commande est appelée après avoir sélectionné une cellule du bord de la feuille ou de l'intérieur de la feuille (Voir Figure 2.5). La différence entre ces deux commandes est que dans le premier cas, les vues couvrent la totalité de la feuille, chaque cellule étant reproduite autant de fois qu'il y a de parties, alors que dans le second, elles forment une partition de celle-ci, une cellule ne pouvant être que dans une seule des parties. Pour revenir à une seule vue, il suffit de choisir la commande adaptée du menu Fenêtre.

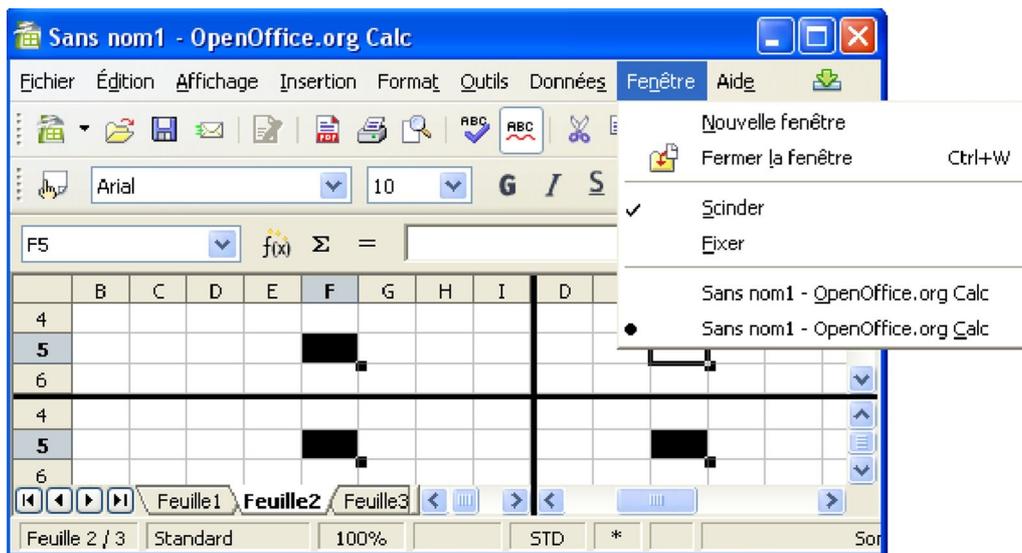


Figure 2.5 : 4 vues sur la feuille de calcul dans une même fenêtre

Pour afficher plus de cellules dans une même fenêtre, voici deux possibilités : 1) on peut faire un zoom jusqu'à 10% pour parcourir très vite la feuille avec les ascenseurs en bas et sur le côté, si on ne sait plus à quel endroit est un graphique, par exemple. 2) on peut diminuer la largeur des colonnes en déplaçant la souris clic droit enfoncé à partir de la séparation entre deux intitulés de colonnes. Toutes les colonnes sélectionnées verront leur largeur modifiée.

3. Exercice : création d'un classeur contenant 2 feuilles de données à partir de 2 fichiers en format *.txt

Attention :

Pour réaliser facilement cet exercice, il faut avoir assimilé les notions d'importation de la leçon précédente.

Il consiste à mettre en forme dans le tableur les données de fichiers récupérés sur des bases de données de l'UCI Repository.

Question

1) Récupérer sur UCI repository (<http://archive.ics.uci.edu/ml/datasets.html>) les données « Adult » sous la forme de deux fichiers adult.data et adult.names (en cas de difficulté, ils ont été recopiés sur le serveur). Les sauvegarder sur votre disque dur. Observez leur contenu en utilisant un éditeur de textes comme WordPad ou NotePad++ (Fig 2.7). Il y a 32561 lignes écrites dans le premier et 110 dans le second.

Conseil :

Si vous ne disposez pas de NotePad++ sur votre machine, il serait bien de le télécharger. Il dispose de nombreuses fonctionnalités intéressantes, comme celle bien utile d'afficher les caractères « non affichables » qui pourraient perturber le fonctionnement du tableur. On voit ici l'affichage de « LF » qui indique la fin de ligne dans les systèmes de type Unix. Il a des fonctions de recherche très perfectionnées (utilisant les expressions régulières)

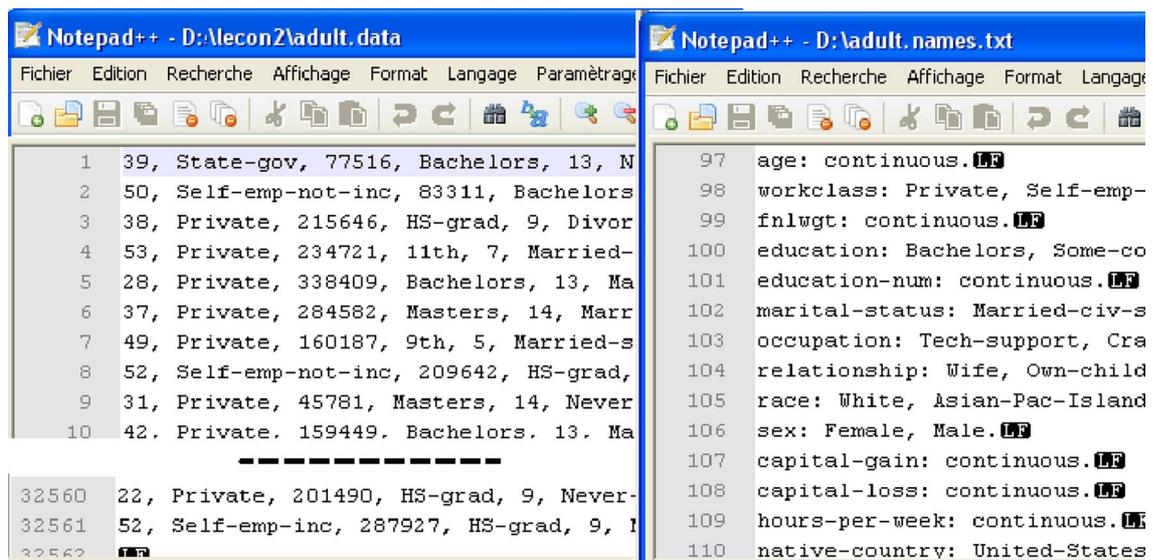


Figure 2.7 : Les deux fichiers « adult » lus avec l'éditeur NotePad++

2) Importez le premier dans le tableur (choisissez bien le séparateur de colonnes). Vous devez obtenir 15 colonnes et 32561 lignes de données qui contiennent des textes et des nombres.

3) Importez le second dans le tableur afin de n'avoir que les noms des variables et la liste de leurs modalités.

Aide :

4) Sauvegardez le fichier contenant ces deux feuilles renommées respectivement « data » et « names » au format tableur par défaut (adult.xls ou adult.ods ou ...).

5) Sélectionner les 14 cellules de la feuille names (en les balayant à la souris, clic-gauche enfoncé), les copier, puis sélectionner la cellule A16 et les coller en les transposant (par Edition>Collage-spécial). Puis sélectionner la ligne 16 qui contient maintenant les intitulés de colonnes et l'insérer en première ligne de la feuille data. Vérifiez bien que vous n'avez pas « écrasé » la première ligne de données. Le résultat doit être celui de la figure 2.12.

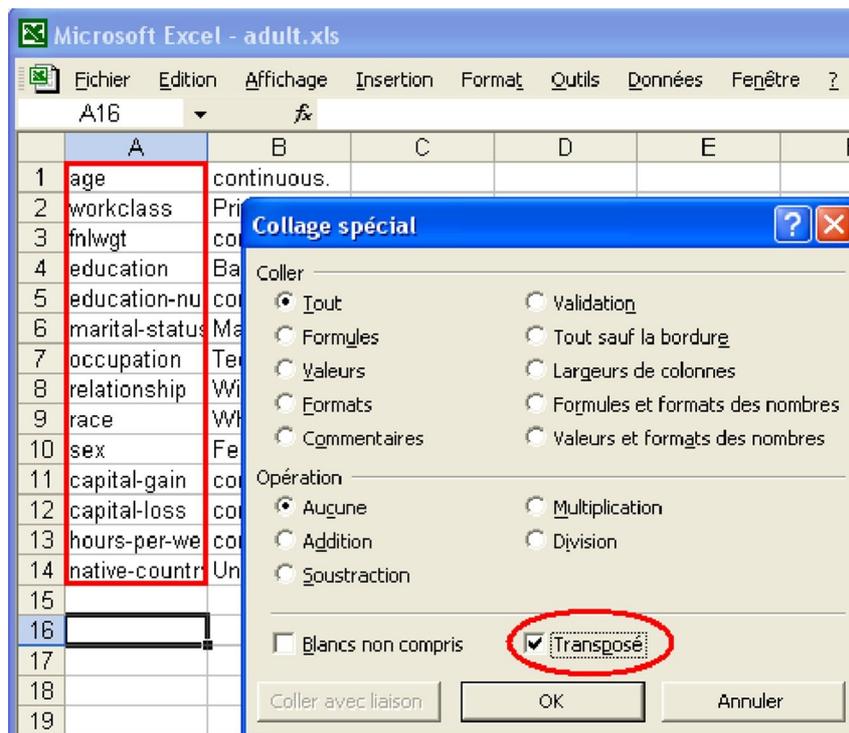


Figure 2.10 : Collage avec transposition sous Excel 2003

B. Quelques « raccourcis clavier » utiles des classeurs



Complément

La circulation dans une feuille de données peut se faire en utilisant les touches de déplacement du clavier. Si le nombre de cellules contenant de l'information est grand, elle est facilitée par les « raccourcis clavier ».

En général, ils sont obtenus en enfonçant d'abord une ou deux touches ne produisant pas d'affichage, Ctrl, Majuscule (notée Maj, ou Shift) ou Alt, puis pendant que celle-ci (ou celles-ci) est enfoncée, en tapant sur une touche de caractère ou sur une touche de fonction (ici F4). Tous sont communs à Excel et OpenOffice, sauf le dernier.

(Ctrl)(←) Passe à la colonne de gauche du bloc actif ou au bloc précédent (ou à la colonne A en l'absence de bloc).; Même principe pour **(Ctrl)(→)** ; **(Ctrl)(↑)** ; **(Ctrl)(↓)** ;

(Ctrl)(flèche début) Positionne le curseur dans la cellule A1 ;

(Ctrl)(Fin) Positionne le curseur à la fin du dernier bloc (placé en bas à droite) de la feuille de calcul active.

Si la touche **(Maj)** est enfoncée avec la touche **(Ctrl)** pour ces déplacements, la zone rectangulaire de cellules comprises entre la cellule de départ et celle d'arrivée est sélectionnée.

(Ctrl)(*) Sélectionne le bloc entier dont la cellule active fait partie. ;

(Ctrl)(c) Copie dans le « tampon » (ou Buffer) le contenu de la zone sélectionnée

(Ctrl)(x) Colle dans la zone sélectionnée le contenu du tampon

(Ctrl)(v) Copie dans le tampon le contenu de la zone sélectionnée et efface la zone

sélectionnée

(Maj)(F4) pour OpenOffice ou (F4) pour Excel : Bascule le statut d'une référence de relative à absolue en passant par les quatre statuts possibles (A1, \$A\$1, \$A1, A\$1).

Le « copier-coller » utilise successivement **(Ctrl)(c)** et **(Ctrl)(v)**, et le « couper-coller » **(Ctrl)(x)** et **(Ctrl)(v)**.

Pour traiter les informations présentes dans les cellules du classeur sans déformation, il faut d'abord les localiser précisément, ce qui est le rôle de l'adressage, les traduire correctement grâce au format, et être capable de les saisir et/ou reproduire correctement. Nous allons voir ces 3 points dans les parties suivantes.

C. Adressage

1. Désignation d'une cellule

Dans l'affichage classique, les colonnes de la feuille de calcul sont désignées par des lettres, les lignes par des chiffres et les cellules par l'association des deux. Ainsi dans la figure 2.13, on a sélectionné la cellule C5 de la feuille « feuille1 ». On voit s'afficher son adresse à gauche, la formule qu'elle contient à droite. La formule contient une adresse de cellule : l'adresse de la cellule « B\$3 ». Le résultat (5,60) du calcul obtenu en remplaçant l'adresse de la cellule (B\$3) par son contenu (2,3) dans la formule ($=2*B\$3+1$) apparaît dans la cellule elle-même.

Dans les premiers tableurs, les colonnes étaient désignées par des nombres, comme les lignes. Les adresses des cellules s'écrivaient avec les deux numéros précédés chacun de la lettre « L » ou « C » convenable : l'adresse actuelle C5, s'écrivait alors L5C3 pour désigner la cellule de la ligne 5 et de la colonne 3. Bien que ces anciennes notations ne soient plus trop utilisées, la logique d'adressage des tableurs est basée dessus et peut paraître obscure avec les notations actuelles. C'est pourquoi nous détaillerons les adressages et leurs effets dans ces deux types d'écriture.

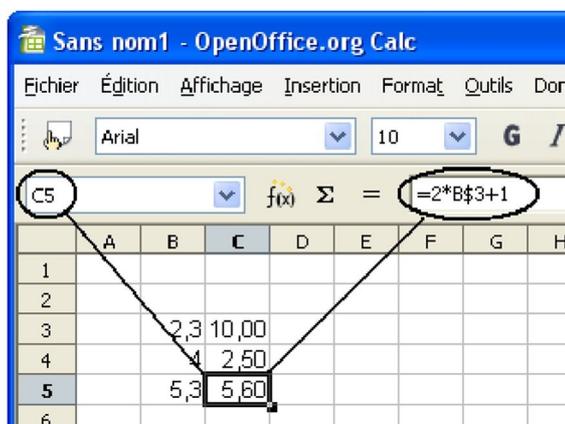


Figure 2.13 : Une cellule contenant une formule avec une référence à une autre formule

2. Adressage absolu/relatif



Fondamental

Le principe général: ce qui fait la différence entre « l'absolu » et le « relatif » est le choix de la cellule de référence : c'est la cellule A1 dans le cas d'adressage absolu, et la cellule « active » (sélectionnée) dans le cas d'adressage relatif. Ainsi lors des déplacements d'une cellule à une autre, la cellule désignée par une adresse relative se déplace d'autant, alors que la cellule désignée par une adresse absolue ne bouge pas. Par défaut, la saisie d'une adresse de cellule produit toujours une adresse relative, que l'utilisateur doit corriger lui-même s'il souhaite la transformer en adresse absolue.



Attention

Bien que le principe d'adressage absolu/relatif soit simple, sa mise en œuvre n'est pas toujours facile à appréhender. En effet l'adresse d'une cellule est formée de deux parties : la désignation de la colonne et celle de la ligne. Ces deux parties peuvent être indépendamment absolues ou relatives, ce qui donne quatre adresses possibles pour une même cellule. Dans les figures 2.14 à 2.16 de l'exemple, nous écrivons dans la cellule C5 la même formule avec les 4 formes de l'adresse de la cellule B3, et nous faisons un « copier-coller » de la cellule C5 vers la cellule D6. Nous voyons que les 4 formes donnent 4 résultats différents, d'où l'importance pour l'utilisateur écrivant une formule de se poser systématiquement la question du choix d'une de ces 4 formes.

Nous détaillons maintenant l'écriture des quatre formes de l'adresse d'une cellule sur un exemple afin que cette notion d'adressage absolu/relatif soit bien assimilée.



Exemple

Considérons la formule « =2*B\$3+1 » de la cellule C5 de la figure 2.13. Dans cette formule la colonne B est relative à la colonne C (donc Colonne-1, qui s'écrit C(-1) dans l'ancienne écriture), et la ligne 3 est absolue car précédée d'un symbole \$ (donc L3) , ce qui s'écrit L3C(-1) dans l'ancienne écriture.

Voici les 4 cas de références à la cellule B3 à partir de la cellule C5 écrits dans les deux écritures, actuelle et ancienne :

B\$3 ou L3C(-1) : la ligne est toujours la ligne 3 (L3), la colonne est la colonne précédente (C(-1))

B3 ou L(-2)C(-1) : la ligne est située 2 lignes au dessus (L(-2)), la colonne est une colonne avant (C(-1))

\$B\$3 ou L3C2 : la ligne est toujours la ligne 3 (L3), la colonne est toujours la colonne 2 (C2)

\$B3 ou L(-2)C2 : la ligne est située 2 lignes au dessus (L(-2)), la colonne est toujours la colonne 2 (C2)

Ils sont représentés dans la figure 2.14 avec une flèche rouge pour chaque déplacement relatif : horizontale d'une case vers la gauche quand on recule d'une colonne, verticale de 2 cases vers le haut quand on remonte de 2 lignes.

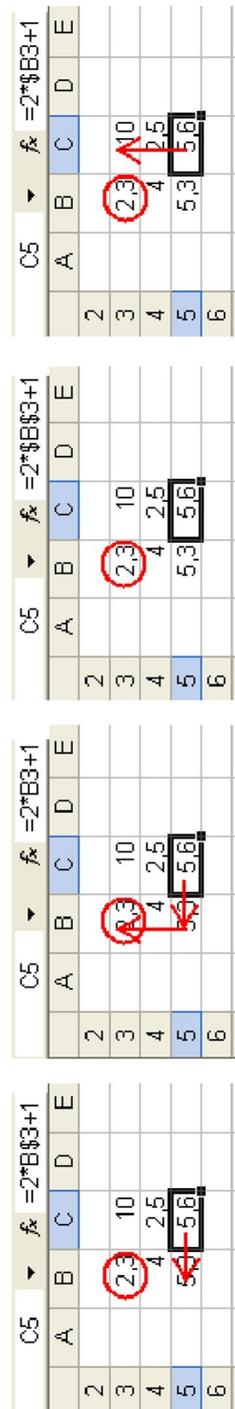


Image 1 : 2.14 : Les 4 adressages possibles de la cellule B3 depuis la cellule C5

Ces formules sont maintenant recopiées par un copier-coller de C5 vers D6, soit un déplacement de +1 ligne et +1 colonne. Les formules selon l'ancienne écriture n'ont pas changé (Fig 2.15) mais les cellules ainsi désignées par ces formules se sont déplacées chaque fois que l'adressage était relatif, les flèches rouges ayant suivi la cellule active dans son déplacement. Dans la figure 2.16, on voit que les formules selon l'écriture actuelle ont changé en même temps.

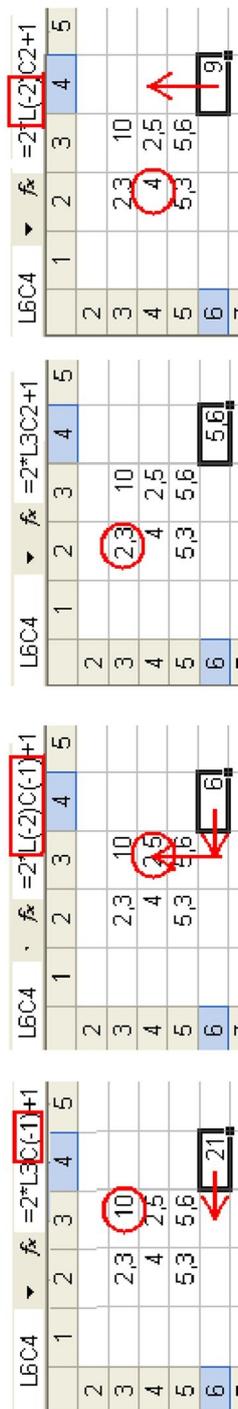


Figure 2.15 : L'effet de la recopie de la cellule C5 vers la cellule D6 selon l'ancienne écriture des références

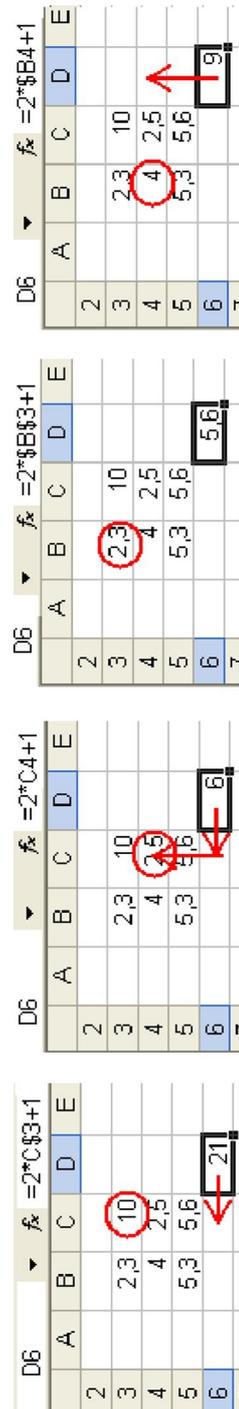


Figure 2.16 : L'effet de la recopie de la cellule C5 vers la cellule D6 selon l'écriture actuelle des références

Quand on recopie une formule **x** lignes plus loin, le seul changement possible porte sur les références relatives des lignes (relative : pas fixée par un \$) qui sont augmentées de **x**, les références absolues aux lignes et les références relatives/absolues aux colonnes demeurant inchangées. Même effet quand on copie une formule **y** colonnes plus loin, les références relatives aux colonnes sont augmentées de **y**.



Attention

Quand la cellule à laquelle on se réfère ne contient pas de nombre, qu'elle soit vide ou qu'elle contienne du texte, elle est remplacée par 0 lors de son utilisation dans un calcul algébrique (ici « $=2*D\$3+1$ » donne $2*0+ 1$).



Fondamental

Si une cellule contient une formule avec des adresses d'autres cellules, la valeur qui s'affiche dans cette cellule peut changer sans que la formule ne change. En effet, toute modification des contenus des autres cellules présentes dans sa formule déclenchera son re-calcul. Par exemple, dans la figure 2.13, si on tape une autre valeur dans la cellule B3, la cellule C5 sera recalculée. Et les cellules C5 à C9 de l'exercice contiennent la même formule dépendant de B3, ce qui fait qu'un changement de valeur de B3 entraînera le changement des valeurs de toutes ces cellules.

3. Exercice : Adressage

Les 6 nombres figurant dans les cellules de B3 à C5 de la figure 2.17 ont été saisis directement au clavier. Dans la cellule C5, qui contient le nombre -3, on saisit la formule « $=2*B\$3+1$ » (pour saisir cette formule, on sélectionne la cellule C5, on tape la formule, puis on enfonce la touche « Entrée »). Le nombre -3 est alors remplacé par la valeur 5,6 obtenue en remplaçant B\$3 par 2,3 (voir à gauche de la figure 2.14).

	A	B	C	D	E
1					
2					
3		2,3	10		
4		4	2,5		
5		5,3	-3		
6					
7					
8					
9					

Figure 2.17 : Écriture de 6 nombres dans le tableur

Puis on sélectionne à nouveau la cellule C5, et on la copie (clic droit de souris, puis copier). On sélectionne ensuite les cellules de C5 à E9, et on colle (clic droit de souris, puis coller). Les 15 cellules se remplissent alors des valeurs issues de la recopie de la formule « $=2*B\$3+1$ ».

Question

[Solution n°12 p 156]

Trouver les formules et les nombres figurant dans ces cellules.

4. Exercice 2.1

Question

[Solution n°13 p 156]

Les valeurs des cellules sont celles de la figure 2.17. On tape dans la cellule B5 la formule « $=\$B\$3+C3+\$C4$ ». Trouver les formules et les valeurs obtenues dans les cellules grisées après un copier-coller depuis la cellule B5.

	A	B	C	D	E
1					
2					
3		2,3	10		
4		4	2,5		
5		5,3	-3		
6					
7					
8					
9					

Figure 2.17 : Écriture de 6 nombres dans le tableau

5. Cascades de mises à jour de cellules

Les formules contenues dans les cellules peuvent se référer à des cellules ne contenant que des valeurs ou à des cellules contenant des formules. Dans ce dernier cas, la mise à jour d'une cellule déclenchera la mise à jour d'une autre et ainsi de suite.



Exemple

Calcul de la somme obtenue en plaçant $S=100$ euros pendant une période $P=5$ ans à un taux d'intérêt annuel de $t=5\%$.

Ce placement peut se modéliser de façon itérative par les deux équations suivantes :

$$I_T = t S_{T-1}$$

$$S_T = S_{T-1} + I_T$$

Où I_T représente les intérêts acquis lors de la $T^{\text{ème}}$ année, et S_T la somme dont on dispose à la fin de l'année T .

	A	B	C	D	E	F	G
1							
2		S	100		Année	Intérêt	Capital
3		t	0,05		0	0	100
4					1	5	105
5					2	5,25	110,25
6					3	5,51	115,76
7					4	5,79	121,55
8					5	6,08	127,63

Figure 2.19 : L'effet de la recopie de la cellule C5 vers les cellules C5 à E9

Dans les cellules B2 à C3 de la figure 2.19 figurent les « paramètres du modèle », ici la somme initiale et le taux d'intérêt annuel. Les résultats des calculs année par année sont dans les colonnes E, F, G. La première ligne (de E3 à G3) contient ce qui s'appelle en informatique « l'initialisation ». Il y a ensuite autant de lignes que « d'itérations », c'est-à-dire une par année. Dans la ligne 3, seule la cellule G3 contient une formule qui est « =C2 ». Inutile ici de mettre des \$, l'initialisation ne donnant pas lieu à une recopie. La ligne 4 contient les formules suivantes :

$$E4 : = E3+1$$

$$F4 : = C3*G3$$

$$G4 : = G3+F4$$

Pour pouvoir recopier automatiquement ces cellules vers le bas, il convient de

mettre des \$ seulement là où il convient. Ici seul C3 est modifié en C\$3. Une fois cette modification effectuée, on fait un copier-coller des cellules de E4 à G4 vers le bas. On peut voir dans la figure 2.20 que 11 cellules dépendent en cascade de C2. Si on tape 120 dans la cellule C2, toutes ces cellules sont mises à jour en cascade.

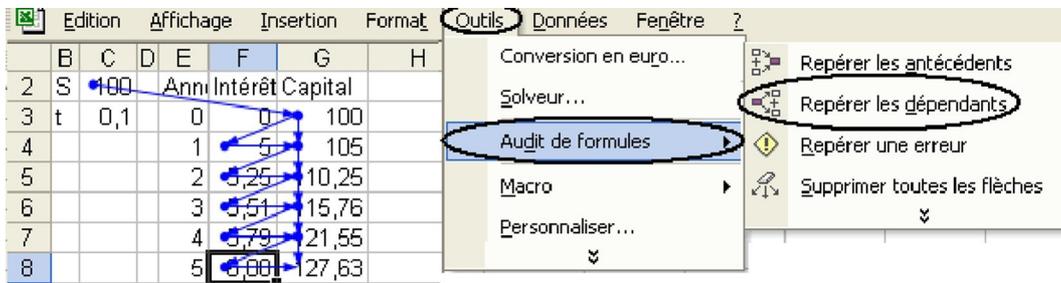


Figure 2.20 : Mise au jour de la succession des dépendances de la cellule C2 grâce à l'audit de formules

Remarque



Cette façon de présenter les résultats n'est bien sûr pas la seule : on peut commencer les calculs à partir de l'année 1 et non comme ici, de l'année 0.

Conseil



Un tableur peut contenir, en théorie, autant de formules que de cellules. Dans la pratique, il faut éviter d'en avoir un très grand nombre dans un même classeur. En effet, la taille d'un classeur augmente en même temps que le nombre de ses cellules utilisées, et la saisie du contenu d'une nouvelle cellule prend de plus en plus de temps, même quand aucune mise à jour n'est nécessaire. Les risques de « plantage » augmentent également avec la taille du classeur. Cette taille « limite » dépend à la fois de la version du tableur et des possibilités de l'ordinateur. Dans tous les cas, pensez à faire des sauvegardes régulières tout au long de votre travail.

Attention



L'écriture de formules dans de nombreuses cellules peut donner lieu à une erreur appelée « références circulaires » sous Excel (voir fig. 2.22 en haut à gauche), et notée Err :522 sous OpenOffice (voir Fig. 2.21), quand une cellule dépend d'elle-même par l'intermédiaire d'autres cellules, alors qu'on attendait une dépendance « en cascade ». Il convient de corriger s'il s'agit d'une erreur. Toutefois, les références circulaires peuvent faire l'objet d'un choix de l'utilisateur pour certaines applications particulières, comme la recherche de la valeur de convergence d'une suite. Dans ce cas les paramètres peuvent être ajustés dans Outils>Option>Calcul après avoir coché « Itérations ».

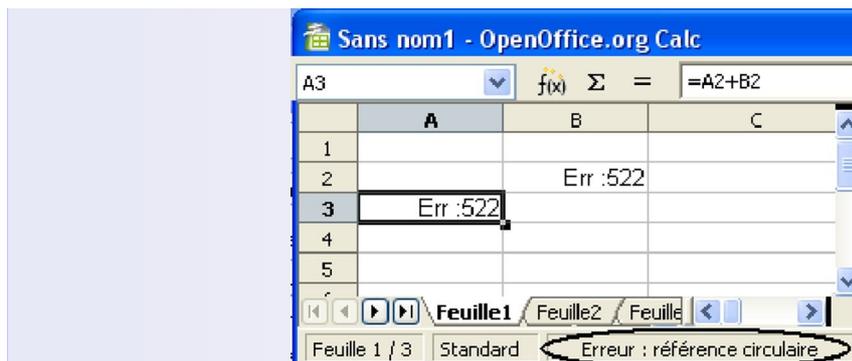


Figure 2.21 : Référence circulaire B2 (=A2+B3) et A3 (=A2+B3) sous OpenOffice

D. Formats de cellules

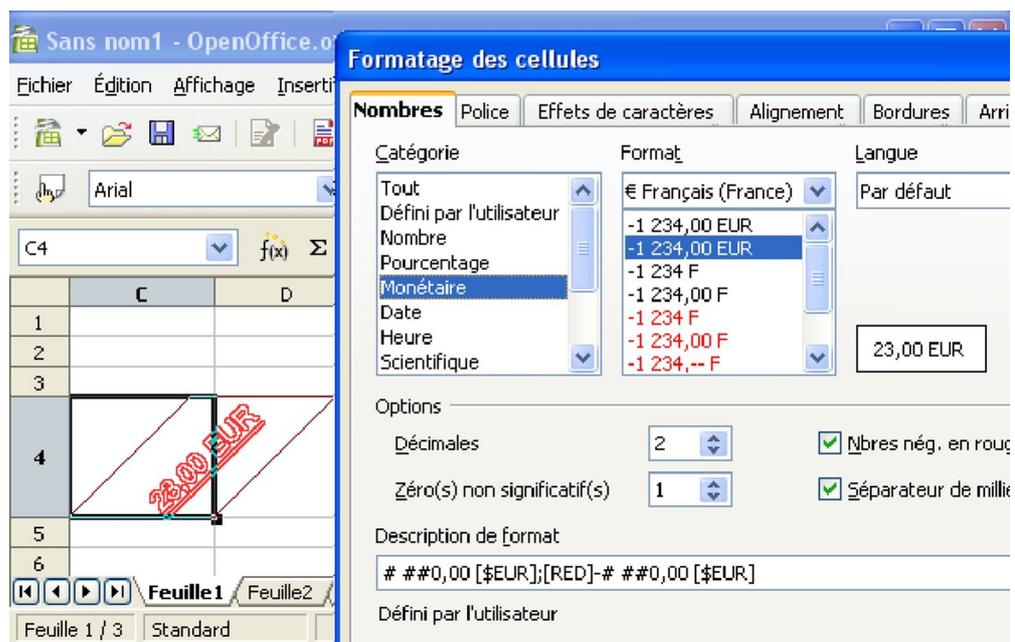


Figure 2.22 : Référence circulaire des cellules A3 et B2, et format de la cellule C4 sous Excel

Une cellule contient de nombreux éléments, les trois principaux étant 1) sa formule, s'il en contient une, 2) sa valeur, le plus souvent un nombre, qui peut être tapé directement ou calculé à partir de la formule et 3) son format. Ce dernier comprend une mise en forme de la cellule (sens de l'écriture, couleur, bordure, etc.) indépendante de son contenu, qu'il soit de type numérique ou textuel et une mise en forme spécifique des informations de type numérique (format des nombres) qui peut changer leur interprétation (pourcentage, date).

1. Mise en forme de la cellule

Dans la figure 2.22, le format de la cellule C4 a été modifié en changeant « l'alignement » (écriture du nombre du bas vers le haut), « la police » (changement de couleur, de police de caractère, soulignement), « de bordure » (de couleur prune, en tiretés). Le nombre est resté inchangé (pas de virgule, pas de %...).

L'écriture du bas vers le haut se pratique surtout pour les intitulés de gros tableaux de nombres. Si on désire afficher le plus possible de colonnes de nombres dans un format A4, il peut être intéressant de réduire la largeur de chaque colonne. La largeur des intitulés écrits de bas en haut n'est que de 1 caractère. S'ils sont trop longs, il faut les découper sur plusieurs lignes. Dans les figures 2.24 et 2.25, on a utilisé plusieurs options de l'alignement sur les premières lignes de la feuille « names » du fichier « adult.xls » de l'exercice de la partie 1.3. de cette leçon. Le format de cellule peut être appliqué à un groupe de cellules après les avoir sélectionnées, comme c'est fait dans ces figures, ou à une seule cellule avant d'être recopié sur le groupe de cellules.

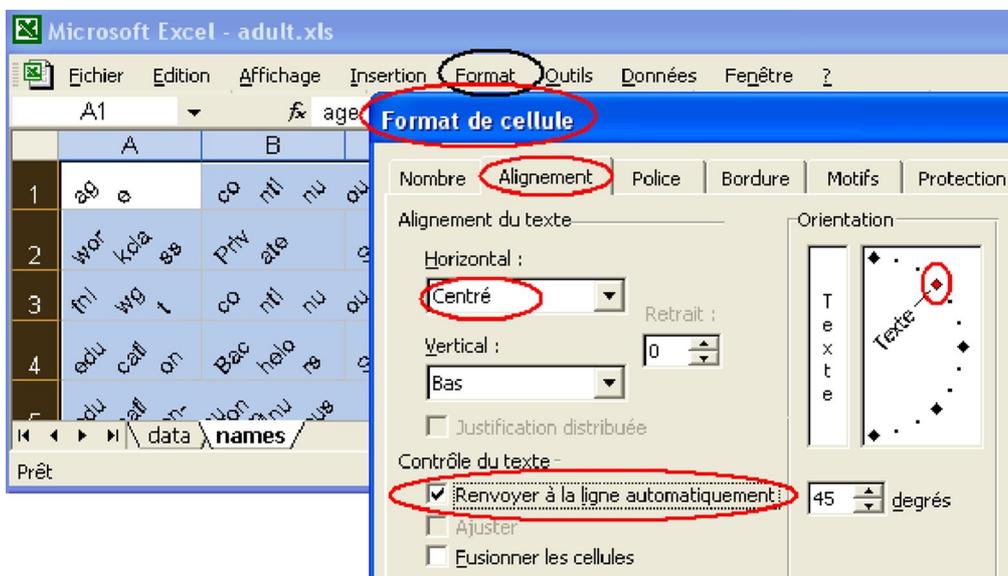


Figure 2.23 : Format cellule, option Alignement, avec Excel

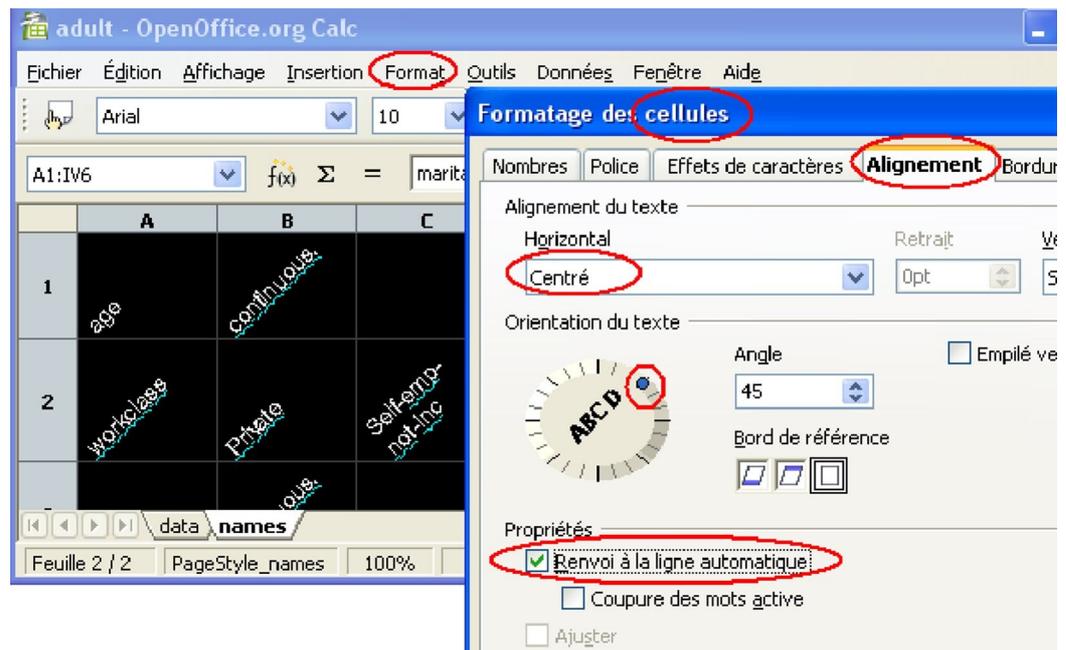


Figure 2.24 : Format cellule, option Alignement, avec OpenOffice

2. Format de nombre

Dans la figure 2.25, on a choisi un format monétaire de nombre, en euros, avec deux chiffres après la virgule, rouge si le nombre est négatif. On aurait pu aussi définir un format spécifique (« Défini par l'utilisateur » selon OpenOffice ou « Personnalisé » selon Excel), en utilisant la formulation appropriée. La formulation du format monétaire est indiquée pour Excel à droite de la figure 2.22 sous l'intitulé « type » et pour OpenOffice en bas de la figure 2.25 sous l'intitulé « Description de format ».

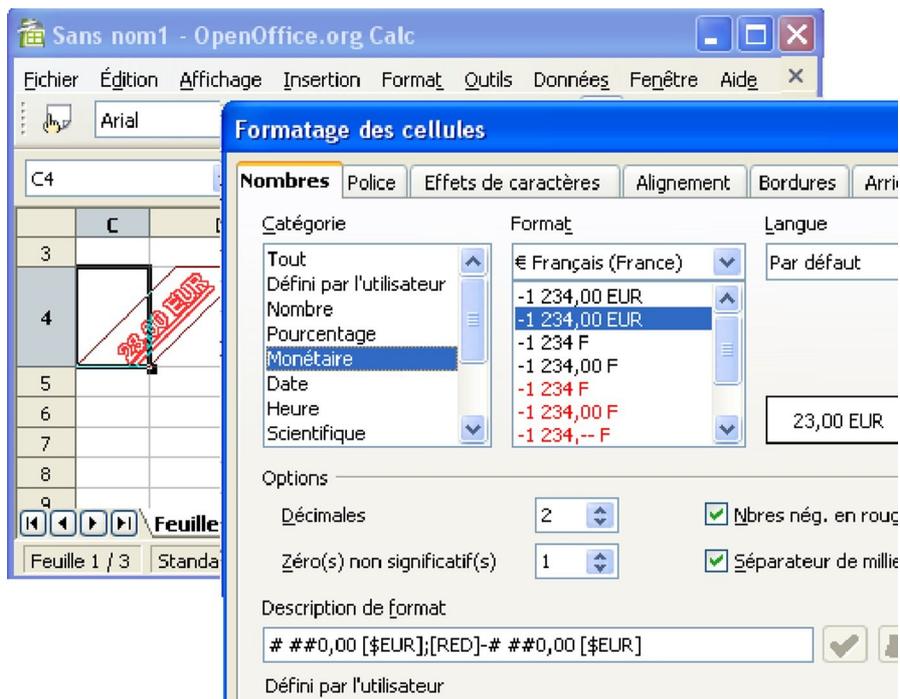


Figure 2.25 : Format de la cellule C4 sous OpenOffice



Attention

Comme on écrit des nombres, on peut de la même façon écrire des textes dans les cellules, en saisissant caractère après caractère au clavier. C'est le tableur qui détermine lors de la saisie si c'est un nombre ou un texte que l'on tape. Par défaut, des nombres sont attendus. Aussi tant que les caractères saisis au clavier semblent composer un nombre, par exemple « 23,00 », le tableur affecte le type « nombre » à cette cellule, mais dès que s'ajoute un caractère qui ne rentre pas dans la composition des nombres, la cellule bascule de type « nombre » à type « texte ». La capacité des tableurs à reconnaître « EUR » comme n'étant pas un texte mais une indication de la monnaie est variable. Dans le doute, il est préférable de taper 23, puis de choisir le format de cellule approprié, ici monétaire, et la valeur affichée dans la cellule sera bien « 23,00 EUR », tout en restant de type numérique. Inversement le tableur ne reconnaîtra jamais un nombre comme texte, aussi si l'on désire que 23 soit traité comme un texte, il faut taper une « quote » avant 23, soit '23.

3. Format de date

Pour écrire une date dans une cellule du tableur, on peut la saisir directement au clavier en utilisant le symbole « / » comme séparateur entre les 3 nombres indiquant respectivement le jour, le mois et l'année. Par exemple pour saisir la date du 12 janvier 1986, il faut taper les 8 caractères « 12/01/86 » à la suite sans inclure d'espaces entre eux. Une fois la date saisie, son affichage se fait ensuite selon le format souhaité (voir Fig 2.26) que l'on atteint par le chemin « Format > Cellules > Nombres > Date » pour OpenOffice, et par un chemin voisin pour Excel (voir Fig. 2.22). On peut aussi créer son propre format de date. En cas de problème lors de la saisie de la date, il convient de vérifier que l'installation du tableur et/ou du système d'exploitation s'est faite avec les options linguistiques correctes (voir partie 4 de la leçon 1).

La date n'est pas stockée comme elle est saisie, mais sous forme d'un nombre qui est le nombre de jours entre une date de départ fixée dans le tableur et la date saisie. Ce codage permet d'obtenir le nombre de jours entre deux dates par simple soustraction, mais autorise aussi des opérations qui n'ont aucun sens comme des additions de dates, des multiplications, etc. Pour connaître cette date initiale, il suffit de taper le nombre 0 dans une cellule et de choisir son affichage avec un format de date.

4. Exercice 2.2 : Trouver le jour de la semaine correspondant à une date donnée.

Tapez la date de naissance d'un de vos amis dans une cellule du tableur, puis choisissez le format de date indiquant le jour de la semaine (voir Fig. 2.26), vous pourrez alors lui dire quel jour de la semaine il est né.

5. Exercice 2.3 : Trouver le jour de la semaine du 1er mai et du 1er janvier pendant 10 ans.

1. Écrire les 4 intitulés de colonne de la ligne 2.
2. Remplir les colonnes B et C comme indiqué sur la figure 2.27
3. Écrire les dates dans les cellules D3 et E3.
4. Transformer les dates des deux cellules pour obtenir l'affichage de la figure 2.27.
5. Écrire la formule convenable dans la cellule D4 : il faut ajouter à la date figurant dans D3 le nombre de jours dans une année normale (365) et le nombre figurant dans C3 (soit 1 si l'année précédente était bissextile, 0 sinon).
6. Mettre les \$ nécessaires dans la formule de la cellule D4 puis faire une copie de la cellule D4 vers les cellules de D4 à E12.

E. Saisie, édition

1. Saisie simple du contenu d'une cellule

On sélectionne la cellule par un clic gauche de souris, puis on tape le nombre (suite de chiffres, avec ou sans le caractère décimal, ou selon la notation scientifique « 2,3E-10 » par exemple), ou la date (« 12/01/86 » Fig. 2.26) ou le texte (par exemple « Comment ça va ? »), ou le signe « = » suivi de la formule (« =2*B\$3+1 » Fig 2.13). Une fois le contenu saisi, on appuie sur la touche entrée pour le valider, ce qui sélectionne la cellule suivante.

Si on sélectionne à nouveau la cellule précédente, son format peut être modifié (voir partie 3. de cette leçon), mais pas son contenu. Pour pouvoir modifier le contenu, il faut faire un clic gauche sur la ligne d'édition de la formule, entourée en haut à droite de la figure 2.13. On peut alors changer certains chiffres, certaines lettres, ou des éléments de la formule avant de valider la modification en appuyant sur la touche entrée. On peut aussi importer le contenu de la cellule par un copier-coller d'une autre application vers la cellule.

2. Alternative pour la saisie du contenu d'une cellule contenant des références

On a proposé dans le paragraphe 4.1 de saisir la formule « =2*B\$3+1 » caractère par caractère au clavier. Il y a une façon plus pratique de procéder pour saisir une adresse de cellule dans une formule : après avoir tapé « =2* », , on clique sur la cellule B3 au lieu de taper les 3 caractères « B\$3 », et la formule se complète en « =2*B3 » puis on insère le \$ entre B et 3 et on termine l'écriture de la formule. On peut aussi changer l'adressage de la cellule « B3 » en « B\$3 » en utilisant le raccourci clavier **(F4)** sous Excel ou **(Maj)(F4)** sous OpenOffice autant de fois que nécessaire.

Si on souhaite modifier cette formule, on clique sur sa ligne d'édition (Fig 2.13 en haut à droite). A ce moment-là, les adresses de la formule changent de couleur, en même temps qu'apparaissent des rectangles de mêmes couleurs indiquant leurs positions dans le tableur. Il est alors possible de déplacer ces rectangles, ce qui modifie en conséquence les adresses correspondantes.

Voir définition du "glisser/déposer"*



Attention

Lors de l'écriture d'une formule, on est en mode « édition de formule » dans lequel tout clic gauche sur une cellule du tableur produit l'écriture de son adresse dans la formule, ainsi que toute utilisation des flèches de déplacement du clavier. Pour sortir du mode « édition de formule », il faut enfoncer la touche « Échappement » (ou ESC) en haut du clavier à gauche, et les corrections faites depuis la dernière validation disparaissent, ou bien valider la nouvelle formule.

3. Adressage en dehors de la feuille active (autre feuille, autre classeur)

En cas d'un grand nombre de données, on préfère souvent avoir une feuille de calcul pour chaque méthode, plus une pour les données brutes. Dans la figure 2.29, on voit que les données de la figure 2.28 ont été copiées dans la feuille 2, et la formule de la cellule H5 a été réécrite dans la cellule B2 de la feuille 1. On procède alors ainsi :

1) sélection de la cellule B2 de la feuille 1, écriture du début de la formule « =somme(», puis on change de feuille par un clic gauche sur l'onglet Feuille2, on sélectionne la plage de cellules souhaitée, on tape le symbole « ; », on sélectionne la cellule suivante, et ainsi de suite, et on termine par «). L'appui de la touche entrée, nous fait rebasculer sur la page 1. On voit que dans la formule, les adresses des cellules sont précédées de l'indication de la feuille dans laquelle elles sont.



Conseil

Lors de l'écriture d'une adresse, d'une fonction dans une formule, l'écrire en lettres minuscules. Cela permet une vérification immédiate de sa syntaxe : si elle est écrite correctement, elle s'affichera en lettres majuscules.

On peut se référer de la même façon à des cellules présentes dans un autre classeur. Cela risque toutefois de poser des problèmes de cohérence si le contenu de ce classeur, ou le classeur lui-même est effacé par la suite.

4. Saisie de la référence à une plage de cellules

Si on veut sélectionner une plage rectangulaire de cellules au sein d'une formule, il suffit de faire un clic gauche sur la cellule à l'extrémité d'une diagonale du rectangle, de taper le symbole « : » et de faire un clic gauche sur l'autre extrémité de la diagonale. Par exemple la plage D5 :F8 qui figure dans la formule de la figure 2.28 correspond aux 12 cellules du rectangle bleu.

Pour sélectionner une autre plage, il faut cliquer sur les cellules une à une en séparant les adresses par des « ; ».



Attention

Ne pas se tromper entre « : » et « ; » : les formules « =SOMME(D5:F8;C7;C5) » et « =SOMME(D5; F8;C7;C5) » donnent des résultats différents. La première est la somme d'une plage obtenue par réunion d'un rectangle de 12 cellules et de 2 cellules isolées, la seconde est la somme d'une plage obtenue par réunion de 4 cellules.

F. Reproduction de cellules

On a détaillé dans la leçon 1 les quatre étapes d'un copier-coller, et dans le paragraphe 2.2 de cette leçon les différents effets d'un copier-coller sur les formules contenant des adresses selon leur mode d'adressage (absolu/relatif). Nous allons dans cette partie remettre en ordre ce qui a été vu et donner quelques compléments.

1. Le copier-coller

Dans une feuille de calcul, on évite de saisir plusieurs fois la même formule bien sûr, mais également des formules proches. Pour cela on choisit avec soin la disposition des cellules et les types de références afin qu'un copier-coller permette de donner les résultats attendus. Par exemple si on veut faire la moyenne de 2 notes suivant certains coefficients, et de 2 autres suivant d'autres coefficients, la même formule peut convenir du moment qu'on écrit les coefficients dans les cellules appropriées.

	A	B	C	D	E	F	G	H
1		Coef1	Coef2	Coef3	Coef4			
2		1,5	0,5	2	3			
3	Numéro	Note1	Note2	Note3	Note4	Moyenne1	Moyenne2	
4	01EF10X35	12	10	15	7	11,5		10,2
5	01EG13X21	5	12	8	6			
6	01EA12X17	17	11	7	12			
7	01EF23X11	15	12	10	9			

Figure 2.30 : Une seule formule pour plusieurs moyennes

On voit dans la figure 2.30 que la cellule F4 contient la moyenne de note1 et note2

avec les coefficients coef1 et coef2. Si on fait un copier-coller de cette cellule vers la cellule H4, on obtient la moyenne de note3 et note4 avec les coefficients coef3 et coef4. Il n'y a plus qu'à déplacer cette cellule de H4 vers G4.

2. Exercice 2.4

Question 1

[Solution n°14 p 158]

	A	B	C	D	E	F	G	H
1		Coef1	Coef2	Coef3	Coef4			
2		1,5	0,5	2	3			
3	Numéro	Note1	Note2	Note3	Note4	Moyenne1	Moyenne2	
4	01EF10X35	12	10	15	7	11,5		10,2
5	01EG13X21	5	12	8	6			
6	01EA12X17	17	11	7	12			
7	01EF23X11	15	12	10	9			

Figure 2.30 : Une seule formule pour plusieurs moyennes

Mettre les \$ nécessaires dans la formule de la cellule F4 pour pouvoir recopier vers F5, et vers H4.

Question 2

Peut-on transformer la formule de la cellule F4 afin que lors de la recopie vers H4 on obtienne la moyenne des notes Note3 et Note4 avec les coefficients Coef1 et Coef2 ? Si oui, donnez la formule corrigée.

3. La copie par « étirage »

Dans la cellule B1, on a écrit le libellé Coef1. Si on sélectionne cette cellule et qu'on « étire » vers la droite, on remplit correctement les cellules suivantes de la ligne 1. Le texte « Coef1 » devient Coef2, et ainsi de suite.

Dans la figure 2.29, on a représenté diverses étapes de cet étirage. 1) On sélectionne la cellule B1, puis on place le curseur en bas à gauche comme indiqué par une ellipse rouge, il doit changer de forme et devenir une petite croix. 2) on enfonce le clic gauche de la souris, et on se déplace vers la droite. D'abord d'une cellule, ce qui donne « Coef2 », puis de 2 et on s'arrête à E1. A ce moment, on lâche le clic gauche de la souris et les cellules se remplissent des intitulés incrémentés.

Cette opération peut se faire vers la droite ou la gauche, et vers le bas ou le haut, mais successivement, pas simultanément. Elle peut se faire aussi avec des formules contenant des adresses, ce qui correspond à un copier-coller plus rapide.

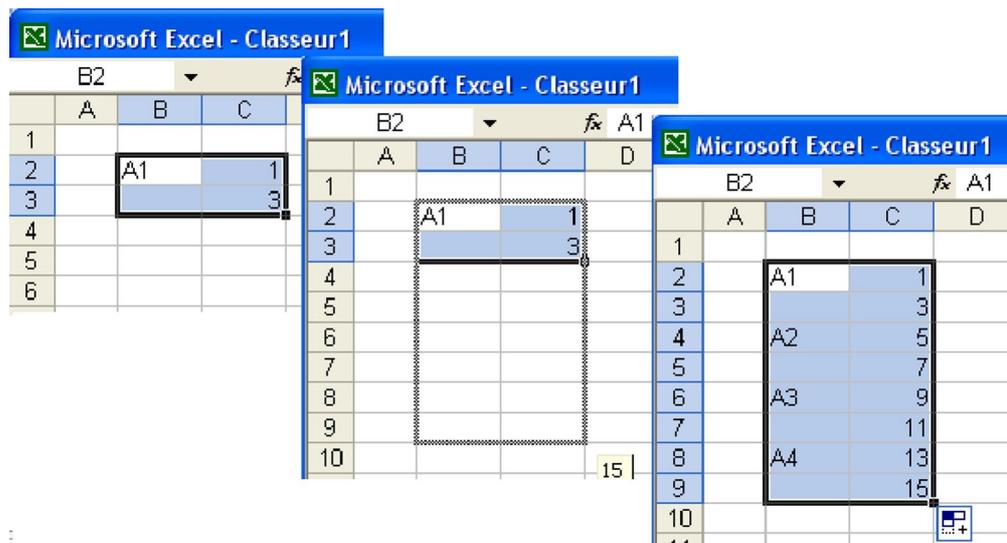


Figure 2.32 : Copie incrémentée de cellules sur 2 lignes par « étirage » vers le bas

Si on sélectionne au départ une suite de deux cellules avec des valeurs numériques dont la différence est D , en étirant, les éléments suivants de la suite sont obtenus en ajoutant D (c'est ce qu'on appelle une suite arithmétique de raison D). Dans la figure 2.32, on a ainsi sélectionné les cellules C2 et C3 contenant les nombres 1 et 3 de différence $D=2$ et on obtenu les valeurs suivantes de la suite arithmétique de valeur initiale 1 et de raison 2 qui sont 5, 7, ...

Cette façon de faire produit des suites « naturelles » qui peuvent différer selon les versions de tableur. Dans Excel 2003 comme dans OpenOffice 2.4, si on écrit « lundi » dans une cellule, on obtient par étirage les jours suivants de la semaine. On voit dans la figure 2.32 que la recopie par étirage permet également de recopier tout un groupe de cellules.

4. Les autres recopies

a) Le collage-spécial

On a vu qu'une cellule contient divers éléments, les 3 principaux étant la formule, la valeur et le format. Lors d'un copier-coller d'une cellule vers une autre, la formule, dans son ancienne forme, et le format sont recopiés. La formule est ensuite recalculée pour donner la valeur qui s'affiche dans la cellule de destination. On peut aussi souhaiter ne copier que la valeur de la cellule d'origine, et perdre ainsi la formule, ou que le format et ne recopier ainsi qu'une mise en forme des cellules sans valeurs ni contenus. Pour cela, la partie « copier » du « copier-coller » ne change pas, mais on doit remplacer la partie « coller » par un collage-spécial. On peut voir la fenêtre du collage spécial d'Excel dans la figure 2.10, et celle d'OpenOffice dans la figure 2.33. Il y a de petites différences entre les deux tableurs. Alors qu'Excel n'autorise qu'une seule rubrique parmi formules, Valeurs et Formats, ce qui force à faire deux collages, un pour les valeurs, l'autre pour les formats, OpenOffice accepte qu'on coche Nombres, Formules et Formats, ce qui oblige à décocher Formules quand on désire Nombres. Tous deux autorisent la transposition que nous avons utilisée précédemment avec Excel.

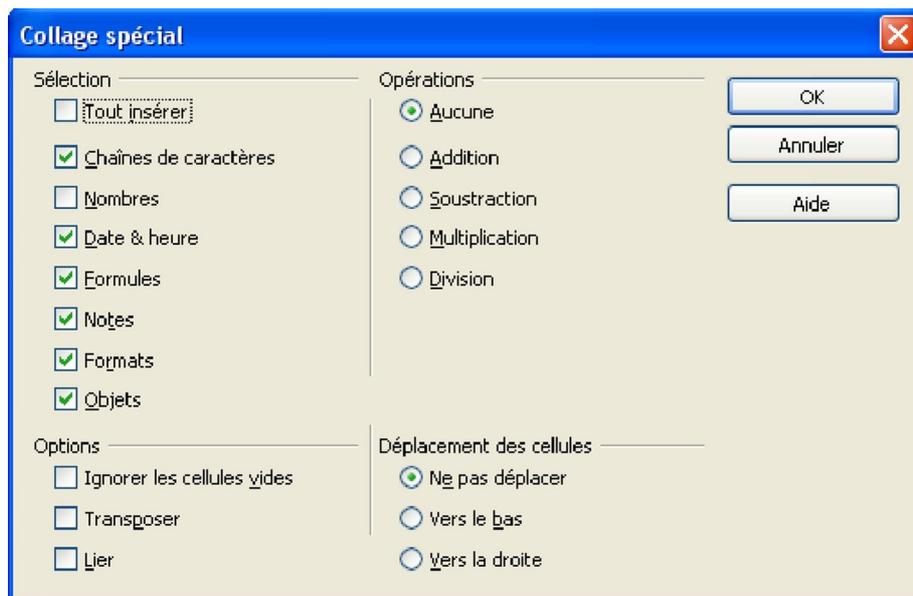


Figure 2.33 : Collage spécial d'OpenOffice 2.4

b) La copie vers une autre application

Si on fait un copier-coller d'une cellule vers une autre application (éditeur de textes par exemple), c'est la valeur que l'on copie et non la formule. Si on souhaite copier la formule, il faut passer en mode édition de formule (en haut de la feuille de calcul), sélectionner la totalité de la formule, la copier, puis sortir de ce mode par ESC et la coller dans l'application.

c) Le couper-coller

Il est utilisé pour déplacer une cellule vers une autre. La nouvelle cellule garde son format, sa formule (nouvelle forme) et sa valeur. On le réalise en 4 étapes comme le copier-coller 1) sélectionner la cellule par un clic gauche 2) Faire un clic droit et choisir « couper » dans le menu qui apparaît 3) sélectionner la cellule de destination par un clic gauche 4) Faire un clic droit et choisir « coller » dans le menu qui apparaît. On peut aussi tout simplement déplacer la cellule avec la souris, ou utiliser les raccourcis clavier au lieu des étapes 2 et 4

d) Copie d'un groupe de cellules

Au lieu de sélectionner une cellule, on en sélectionne un groupe, de préférence formant une zone rectangulaire, et une fois copié, on sélectionne une seule cellule, qui est en haut à gauche du rectangle, puis on colle.

G. Exercices appliqués à l'économie avec correction

1. Problème 3 : Calcul de remboursements de crédits

Un particulier envisage d'acheter une automobile neuve, auprès d'un

concessionnaire. Après négociation, le véhicule est proposé à 16000 €, livrable dans un mois. Le client doit décider rapidement d'un mode de financement.

Première possibilité :

Le garagiste propose une option de location longue durée avec option d'achat (leasing). Il faut verser une première annuité de 4500 € au moment du retrait du véhicule, le 17 octobre, il faut aussi verser un dépôt de garantie de 1900 € (valeur de l'option d'achat), puis verser chaque 17 octobre, pendant 5 ans, la somme de 2400 €.

Question 1

[Solution n°15 p 159]

1a) Calculez la somme totale versée.

Question 2

[Solution n°16 p 159]

1b) Le garagiste a annoncé un taux d'intérêt de 3 % par an pour la somme restant due pendant 5 ans (comme la période est supérieure à 1 an, on utilise la formule d'intérêts composés).

Qu'en pensez-vous ?

Question 3

[Solution n°17 p 159]

1c) Quel devrait être le montant du remboursement ?

Quel était le taux appliqué ?

Deuxième possibilité :

Le client s'adresse à son banquier qui lui propose un crédit de 16000 € sur cinq ans, lui permettant ainsi de conserver un matelas de liquidités (de 6400 €) qui sera placé à 4 %, alors que son crédit sera calculé sur la base d'un taux de 3,5 %. Le montant de l'annuité sera de 3700 € pendant cinq ans.

Question 4

[Solution n°18 p 159]

2a) Calculez le tableau des remboursements, qu'en pensez-vous ?

Question 5

[Solution n°19 p 159]

2b) Après avoir vérifié vos calculs, vous indiquez que le taux annoncé correspond à 3544 € par an. Le banquier vous explique que le prêt est assorti de frais et d'une assurance, le taux correspond alors à environ 5 %.

Troisième possibilité :

Vous envisagez alors d'emprunter à la banque 9600 €, d'utiliser vos 6400 €.

Question 6

[Solution n°20 p 159]

3) Calculez le montant à verser chaque année au taux de 5 %. Que décidez-vous ?

2. Problème 4 : Calcul de coûts

Cet exercice aborde des points clés de la leçon 2 et de la leçon 5.

Cet exercice porte sur le calcul de coûts pour une entreprise qui dispose d'une capacité de production quotidienne de 20 unités, c'est-à-dire qu'elle peut fabriquer chaque jour une quantité comprise entre 0 et 20 unités. Dans un premier temps on raisonnera en unités entières.

RAPPELS

a) Le coût moyen est égal au coût d'une unité produite, il est donc égal au coût total (cf. " p 48) voir rappels sur le coût total plus loin) divisé par q.

$$CM(q) = CT(q) / q$$

Le coût moyen est rarement constant, il dépend de la quantité produite.

Exemple : si le coût total est égal à : $CT(q) = 10 + (15q)^{1/6}$,

la fonction de coût moyen est égale à : $CM(q) = [10 + (15q)^{1/6}] / q$ $CM(q) = (10/q) + (15^{1/6} q^{-5/6})$

b) Le coût marginal est égal au coût de la dernière unité produite.

Remarques :

Certains auteurs choisissent la prochaine unité produite. Lorsque la fonction de coûts est continue et dérivable, on utilisera la dérivée.

$$Cm(q) = CT(q) - CT(q-1)$$

Le coût marginal dépend aussi de la quantité produite.

$Cm(q) = [10 + (15q)^{1/6}] - [10 + (15(q-1))^{1/6}]$ (**attention** : les crochets doivent être remplacés par des parenthèses dans les formules tableur et ne pas oublier de rajouter le signe de la multiplication et de l'exposant)

LE COUT TOTAL

On appelle coût total (CT) la somme des coûts de tous les facteurs de production utilisés.

$$CT = \sum p_i f_i$$

Avec : p_i = prix d'une unité de facteur i et f_i = quantité de facteur i.

Toutes les dépenses de l'entreprise sont considérées comme servant à produire et sont donc des facteurs de production.

Σ = somme pour tous les facteurs utilisés

Le coût total dépend de la quantité produite, en effet on utilisera certains facteurs de production en fonction de la quantité à produire. L'entreprise utilisera les facteurs de production de manière optimale, c'est-à-dire de manière à réduire ses coûts.

On appelle fonction de coût total la fonction qui associe le coût total minimal à une quantité produite (q) :

$$CT(q) = f(q)$$

Le coût total dépend donc de la quantité produite, mais aussi des prix des facteurs de production.

$$CT(q) = f(q, p_1, p_2, \dots)$$

Exemple, lorsque les prix de production sont fixes, on peut avoir la fonction de coût total suivante :

Coût Total (en milliers d'Euros) = $10 + (15q)^{1/6} + q^{1.5}$

Cliquez sur la loupe pour agrandir la fiche

COÛTS FIXES ET COÛTS VARIABLES

Le coût total peut aussi s'exprimer par la somme de coûts fixes et de coûts variables.

$$CT(q) = CF + CV(q)$$

La distinction entre les coûts fixes et les coûts variables est principalement liée au terme envisagé.

Les Coûts Fixes (CF) sont des coûts indépendants des quantités produites : se sont les coûts des facteurs fixes à court terme, tels que les loyers, les charges locatives, les assurances, les frais liés aux remboursements d'emprunts, certains salaires, ... Les coûts fixes ne dépendent pas des quantités produites.

Dans l'exemple les coûts fixes sont : CF = 10 000 Euros.

A long terme les coûts fixes deviennent variables : on peut déménager, modifier les contrats d'assurance, ...

Les Coûts Variables (CV) sont les coûts des facteurs variables, c'est-à-dire des facteurs de production dont les quantités varient avec la quantité produite. Les coûts variables sont donc fonction des quantités produites.

Dans l'exemple les coûts variables sont égaux à $CV(q) = (15 q)^{(1/6)} + q^{1,5}$

Soit la fonction de coût total suivante :

Coût Total (en milliers d'Euros) = $10 + (15 q)^{(1/6)} + q^{1,5}$

Question 1

[Solution n°21 p 159]

1) Calculez les coûts totaux, les coûts moyens et marginaux pour des quantités produites comprises entre 0 et 20.

Question 2

[Solution n°22 p 159]

2) Tracez le graphe correspondant, que constatez-vous ?

Question 3

[Solution n°23 p 159]

3) Comparez le coût moyen et le coût marginal.

Question 4

[Solution n°24 p 159]

4) Déterminez intuitivement la quantité de production qui maximise le profit, pour un prix de vente égal à 5,4€. Calculez le profit correspondant.

Question 5

[Solution n°25 p 159]

5) Même question avec un prix de vente égal à 5,8€

6) Même question pour un prix de vente égal à 3€.

Question 6

[Solution n°26 p 160]

7) Que pensez-vous des résultats obtenus ?

3. Problème 5 : Calcul d'un indice chaînes des prix (suite de l'exercice 1 de la leçon 1)

Question

[Solution n°27 p 160]

1) Compte tenu des évolutions de la structure de consommation, on propose de calculer un indice de Laspeyres pour chaque année, avec pour base l'année précédente.

2) A partir des calculs précédents, calculez un indice chaînes. Qu'en pensez-vous ?

	Q ₀	P ₀	Q ₁	P ₁	Q ₂	P ₂	Q ₃	P ₃	Q ₄	P ₄
Viande	2 kg	12 € par kilo	3	15,00 €	4	16,00 €	5	14,00 €	6	15,00 €
Légume	8 kg	3 € par kilo	9	2,50 €	10	3,00 €	11	3,20 €	12	3,50 €
Pain	12	0,90 €	10	1,00 €	9	1,10 €	8	1,20 €	7	1,20 €
Fromage	4 boîtes	2,50 € la boîte	4	2,00 €	3	2,50 €	2	2,30 €	2	2,60 €
Poisson	1 kg	10,00 €	2	9,00 €	3	10,00 €	4	11,00 €	5	12,00 €

Tableau 3 : Evolution

4. Problème 6 : Modèle macroéconomique

Suite de la leçon 1 : En recherchant les données dans Eurostats :

Question 1

[Solution n°28 p 160]

1) Calculez le PIB par habitant en 2008 et en Euros Courants pour chacun des 27 pays de l'union Européenne.

Question 2

[Solution n°29 p 160]

2) Calculez la part de chaque pays dans le PIB européen.

Question 3

[Solution n°30 p 160]

3) Calculez ensuite le PIB par habitant pour l'ensemble de l'Europe des 27, en faisant une moyenne pondérée des moyennes.

Question 4

[Solution n°31 p 160]

4) Vérifiez vos résultats.

Remarque :

La moyenne pondérée des moyennes est obtenue en pondérant chaque moyenne par les effectifs du pays correspondant.

H. QCM de la leçon 2

Exercice 1

[Solution n°2 p 195]

Question 1 :

	E	F	G	H
1			1	0
2	3	1	3	2
3		3	1	4
4				

Figure 2.38

Dans la cellule G3, on tape la formule =F3+G1.

Après validation, quelle est la valeur obtenue dans cette cellule :

On recopie alors cette formule vers le bas. Qu'obtient-on dans G4 :

On recopie ensuite cette formule à droite. Qu'obtient-on dans H3 :

Exercice 2

[Solution n°3 p 195]

Question 2 :

	A	B	C	D
3	x	y	z	x+y+z
4	1	-2	4	
5		-1		
6		0		
7		1		
8		2		

Figure 2.39

Dans la plage sélectionnée, on veut faire apparaître la somme $x+y+z$, où x et z sont des valeurs constantes figurant dans les cellules A4 et C4, et y une variable pouvant prendre les valeurs situées de B4 à B8. Cette formule une fois tapée et validée sera recopiée vers le bas, et dans la cellule D7, par exemple apparaîtra la valeur $x+y+z=1+1+4$, soit 6. Quelle formule doit-on taper dans la cellule D4 pour que la copie automatique permette d'obtenir les valeurs attendues :

- =A4+\$B4+C4
- =A\$4+B\$4+C\$4
- =somme(A4:C4)
- =A\$4+\$B4+\$C\$4
- aucune de celles-là

Exercice 3

[Solution n°4 p 196]

Question 3 :

Dans les cellules A2, A3, A4, on a tapé la nombre 1, dans B2,B3,B4, la nombre 2, dans C2,C3,C4, le nombre 3, et rien dans les autres cellules, qui sont vides.

Une fois tapée la colonne de 1, y-a-t-il un moyen de remplir les deux autres colonnes des nombres souhaités sans les taper tous systématiquement ?

- Vous tapez 2 dans les cellules B2, B3 et B4 puis vous sélectionnez à l'aide de la souris la plage A2 :A4 à B2 :B4 et vous tirez vers la gauche, l'ordinateur reconnaîtra la suite de nombre et inscrira le nombre 3 dans les cellules C2 à C4.
- Vous sélectionnez la plage de A2 à A4 et vous tirez vers la gauche, l'ordinateur écrira une suite de nombre 1,2,3 dans les cellules B2 à B4 et C2 à C4.
- Vous écrivez le nombre 2 dans la colonne C2, puis à l'aide de souris, vous sélectionnez la cellule C2 puis vous tirez vers le bas. Ensuite vous sélectionnez la plage A2 :A4 à B2 :B4 et vous tirez vers la gauche pour remplir les cases de C2 à C4.
- Vous rentrez toutes les données à la main.

Exercice 4

[Solution n°5 p 197]

Question 4 :

Dans la cellule B4, on tape la formule $=\$A4-B2+C\3 . Puis on la recopie automatiquement dans la plage rectangulaire des cellules B4:D5. Donner la formule figurant dans chaque cellule, ainsi que le nombre de cellules B4 :D5. Donner la formule figurant dans chaque cellule, ainsi que le nombre qui apparaît à l'affichage.

cellule | formule | nombre
 B3 | $\$A4-B2+C\3 | **2**
 B4 | $\$A5-B3+C\3 | **1**
 C3 | $\$A4-C2+D\3 | **-2**
 C4 | $\$A5-C3+D\3 | **-3**
 D3 | $\$A4-D2+E\3 | **1**
 D4 | $\$A5-D3+E\3 | **0**

cellule | formule | nombre
 B3 | $\$A4-B2+C\3 | **2**
 B4 | $\$A4-B2+C\3 | **2**
 C3 | $\$A4-B2+C\3 | **2**
 C4 | $\$A4-B2+C\3 | **2**
 D3 | $\$A4-B2+C\3 | **2**
 D4 | $\$A4-B2+C\3 | **2**

cellule formule nombre
 B3 | $\$A4-B2+C\3 | **3**
 B4 | $\$A5-B3+C\3 | **2**
 C3 | $\$A4-C2+D\3 | **-3**
 C4 | $\$A5-C3+D\3 | **5**
 D3 | $\$A4-D2+E\3 | **1**
 D4 | $\$A5-D3+E\3 | **0**

cellule formule nombre
 B3 | $\$A4-B2+C\3 | **3**
 B4 | $\$A4-B2+C\3 | **3**
 C3 | $\$A4-B2+C\3 | **3**
 C4 | $\$A4-B2+C\3 | **3**
 D3 | $\$A4-B2+C\3 | **3**
 D4 | $\$A4-B2+C\3 | **3**



Leçon 3 : Formules élémentaires, conditionnelles, fonctions, calcul matriciel

Qu'est-ce qu'une formule ? Comment l'écrit-on ?	65
Formule élémentaire	67
Les fonctions	74
Formule conditionnelle	77
Calcul matriciel	81
Exercices appliqués à l'économie avec correction	87
QCM de la leçon 3	92

A. Qu'est-ce qu'une formule ? Comment l'écrit-on ?

1. Définitions succinctes des divers types de formules

a) Formule élémentaire

Formée d'opérandes et d'opérateurs. Les opérandes sont indiqués par leur valeur ou par leur adresse (ex : 2,3 ou \$B3), les opérateurs sont tapés directement depuis le clavier (ex : +, *, ...). L'écriture de la formule suppose la compatibilité des opérandes et des opérateurs utilisés (ex : 2* « oui », 3/0 fournissent un résultat de type « erreur »). Le résultat d'une formule figurant dans une cellule peut ensuite être utilisé comme opérande d'une autre formule.

b) Fonction

De la forme **fn(arg1,arg2, ..)**, où **arg1, arg2, ...** sont des arguments définis comme précédemment les opérandes, et **fn** est une fonction indiquant le type d'opération à effectuer sur ces arguments (Log, exp, somme, ...). L'écriture de cette formule est soumise également aux contraintes de compatibilité précédentes et son résultat utilisé de la même façon.

c) Formule conditionnelle

Sa forme est identique à celle de la fonction, mais un de ses arguments est interprété comme un « critère » (par exemple « >3 ») qui est appliqué sur un autre de ses arguments, et les résultats dépendent de la façon dont l'argument vérifie le critère. Actuellement il n'y a que 3 formules conditionnelles, les deux les plus utilisées étant SI et NB.SI.

d) Formule complexe

Son écriture mêle des opérandes, des opérateurs ainsi que des fonctions et/ou conditions avec leurs arguments. On a le choix entre l'écriture d'une formule complexe dans une seule cellule, ou l'écriture dans plusieurs cellules de sa décomposition en éléments simples (formules élémentaires, fonctions, conditions) à raison d'un élément par cellule.

e) Calcul matriciel

Une partie des formules ci-dessus (+, moyenne,...) peuvent opérer sur des ensembles de nombres. Quand ces ensembles de nombres sont sous forme de tableaux rectangulaires, on les appelle « matrices » et on peut leur appliquer des formules matricielles spécifiques. Certaines ont comme résultat une matrice, qui s'écrit dans plusieurs cellules. L'utilisation du calcul matriciel faite dans ce cours n'exige pas de connaissances spécifiques en mathématiques, l'aspect pratique étant privilégié sur la théorie.

2. Questions et réponses

Question 1

[Solution n°32 p 160]

Q1 : Faut-il mettre ou non des espaces dans une formule ?

Question 2

[Solution n°33 p 160]

Q2 : Doit-on mettre en majuscules les lettres figurant dans une formule ?

Question 3

[Solution n°34 p 160]

Q3 : Un collègue qui a la même version de tableur que moi dispose de fonctions que je n'ai pas. Comment est-ce possible ?

3. Édition d'une formule

Nous avons vu précédemment comment écrire dans une cellule une formule simple (élémentaire ou avec fonction) contenant une adresse. L'édition de toutes les formules peut se faire ainsi, en tapant directement au clavier le signe « = », puis les nombres, les séparateurs et les fonctions, et pour les adresses, en les tapant ou en cliquant sur les cellules correspondantes, avant d'ajouter des \$ pour les rendre absolues le cas échéant. On peut aussi se laisser guider par l'aide à la saisie des formules qui apparaît au fur et à mesure qu'on tape les éléments de la formule au clavier (Fig.3.1 à droite, la fonction « type » est proposée), ou aller chercher

directement dans l'aide les fonctions souhaitées (Fig.3.1 à gauche, la liste des dernières fonctions utilisées, et au milieu, « l'assistant fonction », qui propose toutes celles existant dans le tableur).



Figure 3.1 : L'aide à l'écriture de formules dans OpenOffice 3.1



Conseil

Si la formule que vous écrivez vous semble fautive, n'appuyez pas sur la touche « Échappement », car tout ce que vous avez écrit disparaîtrait sans laisser de trace. Validez-la par l'enfoncement de la touche « Entrée », puis sélectionnez à nouveau la cellule pour la corriger. Si le tableur refuse votre validation et qu'il vous faut l'effacer, sélectionnez-la auparavant depuis la zone d'édition à l'aide de la souris, et faites un copier-coller vers un éditeur de textes ouvert par ailleurs avant d'appuyer sur la touche « Échappement » pour sortir du mode édition. Vous pouvez alors tester plusieurs modifications de la formule par des copier-coller de l'éditeur de textes vers la cellule du tableur.

B. Formule élémentaire

Les opérateurs sont utilisés dans une expression de la forme suivante:

Opérande_1 Opérateur opérande_2

Opérande_1 et Opérande_2 sont généralement de même type (nombre, texte, condition) et l'évaluation de cette expression fournit un résultat du type attendu si l'opérateur est compatible avec les opérandes, sinon le résultat est une erreur. Les opérateurs peuvent être arithmétiques, comparatifs, textuels ou relatifs à des références. Nous décrivons d'abord l'usage isolé de chacun d'entre eux, puis leur association au sein d'une même formule.

1. Opérateurs arithmétiques

Les opérandes sont des nombres et le résultat est un nombre, sauf en cas d'erreur.

+ addition, - soustraction, * multiplication, / division, ^ exponentiation

Parmi ces opérateurs, celui de la soustraction, qui est binaire, partage son symbole avec l'opérateur unaire alors que les deux opérateurs ne s'interprètent pas exactement de la même façon. (« -B\$3 » est l'opposé de « B\$3 »). Un autre opérateur unaire « % » existe qui permet de prendre le « pourcentage » d'un nombre. Par exemple =2,4567% tapé dans une cellule produit l'affichage de 2,46% dans cette cellule, qui représente le nombre 0,024567.

Question :

Les booléens « VRAI » et « FAUX » sont-ils des textes ou des nombres ?

Réponse :

Les booléens sont d'un type différent, ni des nombres, ni des textes. Toutefois quand ils figurent dans une formule avec des opérateurs arithmétiques, ils sont transformés en nombres : FAUX devient 0 et VRAI devient 1 (voir cellules B9 et B10 des figures 4.2 et 4.3). Inversement quand un nombre figure dans des expressions booléennes, s'il est nul il devient le booléen FAUX, sinon le booléen VRAI.

Remarque



La fonction « TYPE » permet de connaître le type du contenu d'une cellule, à travers un code. Sous OpenOffice, la correspondance entre les types et les codes est la suivante :

Contenu de la cellule	Nombre	Texte	Booléen	Formule	Erreur	Matrice
Valeur rendue par type()	1	2	4	8	16	64

Tableau 4 : Tableau 1 : correspondance entre les types et les codes

La seule différence entre Excel 2003 et OpenOffice 3.1 porte sur le type « formule » : alors que la présence d'une formule dans la cellule produit le code 8 sous OpenOffice, elle produit un autre code sous Excel 2003, qui correspond à la valeur obtenue avec la formule, et non à la formule elle-même, comme on peut le voir dans les figure 3.2 et 3.3. Dans ces deux figures, la colonne C contient les résultats obtenus par l'application de la fonction type() sur la colonne B. A droite de ces figures, on a les formules, à gauche les valeurs obtenues, et au centre la façon de basculer d'un mode d'affichage à l'autre à l'aide du menu « options ».

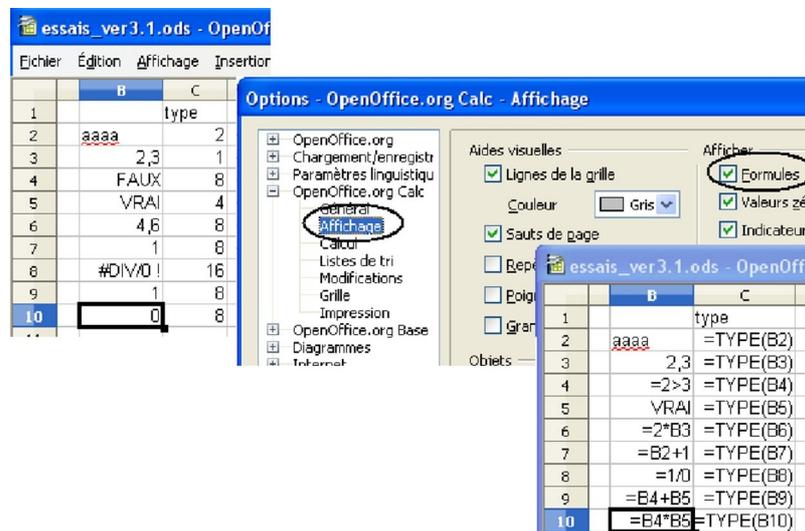


Figure 3.2 : L'option d'affichage des formules dans les cellules avec OpenOffice 3.1

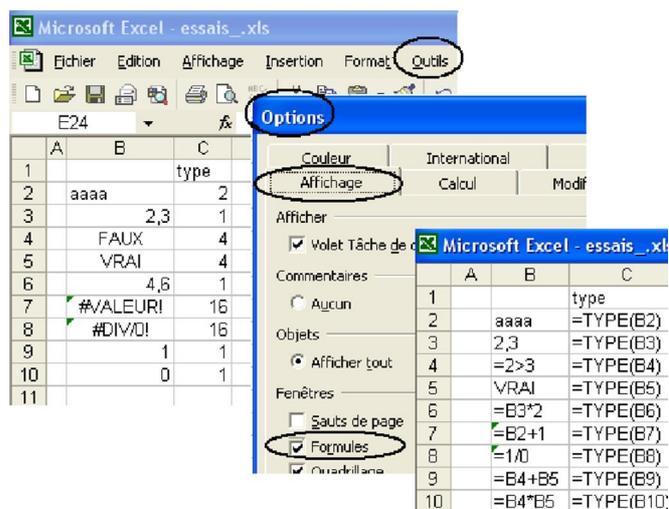


Figure 3.3 : L'option d'affichage des formules dans les cellules avec Excel 2003



Remarque

Les calculs sont organisés de façon à minimiser le nombre d'erreurs, par une conversion implicite de type. Ainsi un booléen peut devenir un nombre (VRAI devient 1 et FAUX devient 0) ou inversement (un nombre différent de 0 devient VRAI, et 0 devient FAUX). Ce qui fait que « = FAUX +VRAI » est équivalent à « =0+1 » et produit le nombre 1 et non une erreur (voir cellules B9 et B10, figures 3.2 et 3.3). Par contre « 'aaaa'+1 » produit une erreur sous Excel, du fait de la différence de type entre le texte 'aaaa' et le nombre 1, alors que le texte est transformé en le nombre 0 sous OpenOffice (voir cellule B7, figures 3.2 et 3.3), pour pouvoir s'ajouter au nombre 1.

2. Opérateurs de comparaison

Les opérandes sont des nombres, des booléens ou des textes et le résultat est booléen.

= égalité, > supérieur à, >= supérieur ou égal, < inférieur, <= inférieur ou égal, <> inégalité

Si les opérandes sont tous deux des nombres, c'est l'ordre des nombres réels qui permet de trouver le résultat, s'ils sont booléens VRAI>FAUX.

Si ce sont deux textes, l'ordre de référence est l'ordre lexicographique. Tant que les textes sont composés de lettres de la même casse (minuscules ou majuscules), cet ordre est le même que l'ordre alphabétique. Les chiffres présents dans les textes sont considérés comme des lettres de rang inférieur, dans l'ordre habituel de 0 à 9. La comparaison de lettres minuscules et majuscules ne donne pas le même résultat selon les versions de tableur. En Excel 2003, leurs valeurs sont les mêmes ("Ab"="aB") alors qu'elles peuvent différer en OpenOffice (en OpenOffice 3.1, "Ab">"aB"), comme on peut le voir dans la cellule B6 des figures 3.4 et 3.5. Par exemple, voici l'ordre de 5 textes selon les deux types de tableur (la dernière chaîne est composée de 4 caractères. On ne voit pas le dernier car c'est un espace).

"2A" < "aa" ; "aa" < "Aa2" ; "Aa2" < "aAb" ; "aAb" < "aAb " (voir Fig. 3.4 et 3.5 les cellules de B7 à B10)

De plus tout nombre est inférieur à toute chaîne de caractères (B5), et pour Excel toute chaîne est inférieure à tout booléen, ce qui n'est pas le cas dans OpenOffice

puisque les booléens sont du même type que les nombres (voir Fig. 3.4 et 3.5

	B	C	D	E
2	ch(i)<ch(i-1)	ch	nbcar(ch)	type(ch)
3		FAUX	1	4
4	VRAI	2345	4	1
5	VRAI	2a	2	2
6	VRAI	2A	2	2
7	VRAI	aa	2	2
8	VRAI	Aa2	3	2
9	VRAI	aAb	3	2
10	VRAI	aAb	4	2

Figure 3.4 : L'ordre selon l'opérateur « < » avec OpenOffice 3.1

	B	C	D	E
2	ch(i)<ch(i-1)	ch	nbcar(ch)	type(ch)
3		FAUX	4	4
4	FAUX	2345	4	1
5	VRAI	2a	2	2
6	FAUX	2A	2	2
7	VRAI	aa	2	2
8	VRAI	Aa2	3	2
9	VRAI	aAb	3	2
10	VRAI	aAb	4	2

Figures 3.4 et 3.5 : L'ordre selon l'opérateur « < » avec OpenOffice 3.1 (haut) et Excel 2003 (bas)

Remarque



Dans les figures 3.4 et 3.5, chaque cellule de la colonne D contient la fonction NBCAR() qui donne le nombre de caractères de la valeurs écrite dans la cellule correspondante de la colonne C.

Conseil



Pour rendre l'utilisation d'un travail réalisé avec le tableur plus facile, et transposable d'une version de tableur à une autre, il est mieux de n'utiliser les opérateurs de comparaison qu'avec deux valeurs de même type, et si ce sont des chaînes de caractères, il faut veiller à utiliser la casse de la même façon.

Méthode



Il arrive que le correcteur orthographique du tableur intervienne pour corriger automatiquement l'écriture de chaînes de caractères du genre « aB » pour la rendre compatible avec les normes en usage dans la langue française. Dans ce cas, deux solutions s'offrent à l'utilisateur : soit désactiver ou modifier les corrections automatiques en allant dans le menu « options », soit taper d'abord le caractère apostrophe puis la chaîne de caractères souhaitée.

3. Opérateurs de texte

Le seul opérateur de texte est la « concaténation » (symbole &). Ses opérandes sont des textes ou des nombres, et le résultat est un texte.

= « Bonjour » & « Monsieur » produit « BonjourMonsieur »

=23&341&42 produit « 2334142 »



Remarque

Quand on concatène des chaînes de caractères, il faut penser à ajouter des espaces pour rendre le résultat plus lisible.

4. Opérateurs de référence

Les opérandes sont les références de deux cellules et le résultat est un ensemble de cellules. Il y a deux opérateurs celui de plage « : » et celui d'union « ; ». Par exemple,

- B5 : F7 correspond aux 15 cellules de la plage rectangulaire dont une des deux diagonales a pour extrémités ces deux cellules.
- B5 ; F7 est simplement la réunion de ces deux cellules.

On a représenté dans la figure 3.6 une utilisation de ces deux types d'opérateurs pour le calcul d'une somme. Nous verrons plus loin que l'opérateur « : » est très utile pour écrire les formules conditionnelles et matricielles.

	B	C	D	E	F
2					
3	=SOMME(B5:C6;D6:F8;B7)				
4					
5	2	-1			
6	2	3	4	8	
7	1		5		
8					

Figure 3.6 : Utilisation des opérateurs de référence « : » et « ; » dans la formule de la cellule B3



Rappel

On a vu dans le paragraphe 4.4 de la leçon 2 comment éditer des formules contenant des plages rectangulaires.

5. Combinaison de plusieurs opérateurs dans une même expression

a) Plusieurs opérateurs arithmétiques

Il est courant d'écrire plus d'un opérateur arithmétique au sein d'une même expression. Quand c'est le même opérateur qui est utilisé plusieurs fois, le calcul se fait par étapes successives de la gauche vers la droite : pour calculer par exemple 2-3-4, on calcule d'abord 2-3, qui fait -1, puis -1-4, qui fait -5. Quand il y a des opérateurs arithmétiques différents, il faut commencer par celui qui a l'ordre de priorité le plus élevé. Cet ordre diffère selon les langages. Pour la quasi-totalité de ceux-ci, la multiplication et la division ont un ordre de priorité plus élevé que l'addition et la soustraction et l'exponentiation a un ordre plus élevé que la multiplication et la division. Dans la plupart des calculatrices, le moins unaire a la même priorité que la soustraction, alors que dans les deux versions de tableurs que nous avons examinées (OpenOffice 3.1 et Excel 2003), il a une priorité plus élevée que l'exponentiation, et donc que la multiplication, la division, l'addition et la soustraction. Si on désire imposer un ordre particulier, on le fait en ajoutant des parenthèses à l'endroit approprié.

Calculons par exemple l'expression arithmétique $-2^4+3*4/2-10$ comme le tableur le ferait : Pour indiquer l'ordre des opérations, on met d'abord des parenthèses par ordre de priorité décroissante.

$$\begin{aligned}
 &= -2^4+3*4/2-10 \\
 &= (-2)^4+3*4/2-10 \text{ (- unaire)} \\
 &= ((-2)^4)+3*4/2-10 \text{ (exponentiation)} \\
 &= ((-2)^4)+((3*4)/2)-10 \text{ (* et /)} \\
 &= (((-2)^4)+((3*4)/2))-10 \text{ (+ et -)}
 \end{aligned}$$

Puis on effectue les calculs en allant des parenthèses les plus internes aux plus externes.

$$\begin{aligned}
 &= (16 +((3*4)/2))-10 \\
 &= (16 +(12 /2))-10 \\
 &= (16 + 6)-10 \\
 &= 22 -10 \\
 &= 12
 \end{aligned}$$

On obtient ainsi 12 alors qu'on aurait obtenu -20 en effectuant ce calcul avec une calculatrice scientifique ordinaire. On peut voir dans la figure 3.7 une autre façon de représenter l'ordre des opérations. Cette notation appelée « notation polonaise », nationalité de son inventeur Jan Łukasiewicz, permet une écriture « préfixée des opérations, utilisée dans certains langages informatiques (Lisp) et calculatrices : l'opérateur est avant ses opérands et non entre eux comme dans la notation classique appelée aussi « infixée ».

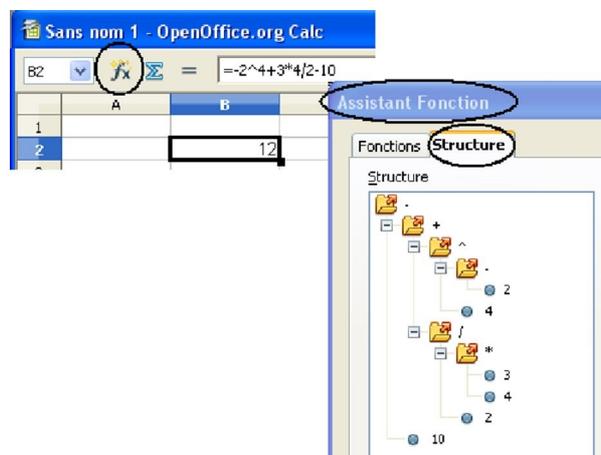


Figure 3.7 : L'ordre de calcul de l'expression indiqué par sa « structure » arborescente sous OpenOffice 3.1

b) Plusieurs opérateurs de types différents

Le calcul d'une expression comportant des opérateurs de divers types se fait sur le même principe que celle comportant plusieurs opérateurs arithmétiques : en suivant l'ordre de priorité s'il n'y a pas de parenthèses, et en tenant compte d'abord des parenthèses s'il y en a. Toutefois mélanger des opérateurs de quatre catégories différentes définies précédemment a rarement du sens. Dans le tableau 3.1 figure l'ordre de priorité des opérateurs d'Excel.

Opérateur	Description
: (deux-points) et ; (point-virgule)	Opérateurs de référence
-	-unaire
%	pourcentage
^	Exposant
* et /	Multiplication et division
+ et -	Addition et soustraction
&	Concaténation de deux chaînes de texte
= < > <= >=	Comparaison

Tableau 5 : Tableau 3.1 : Les opérateurs d'Excel rangés selon leur priorité décroissante



Attention

L'écriture d'une double inégalité comme $2,3 < b3 < 20$ dans une formule n'a pas le sens attendu. En effet, on la calcule dans l'ordre $(2,3 < b3) < 20$, et si $b3$ contient la valeur 20 par exemple, le calcul de $2,3 < b3$ produit le booléen VRAI, et c'est ce booléen qui est comparé à 20, et non $b3$. Elle doit être remplacée par $ET(2,3 < b3 ; b3 < 20)$, qui permet bien de contrôler à la fois si $b3$ est plus grand que 2,3 et plus petit que 20.

C. Les fonctions

Elles sont regroupées par catégories de façon légèrement différentes selon les versions de tableurs. A titre d'exemple, nous indiquons leur classement sous OpenOffice 3.1 et Excel 2003. Les catégories nommées de la même façon sous Excel et OpenOffice sont Base de données, Date & Heure, Finances, Information, Logique, Statistique, Texte. Les catégories Mathématique et Add-in d'OpenOffice correspondent respectivement aux catégories Math & trigo et Scientifiques d'Excel. Les catégories Matrice et Classeur d'openOffice sont fusionnées en la catégorie Recherche & Matrices d'Excel.

Voici une liste réduite de ces fonctions issue de la documentation d'OpenOffice disponible en ligne (http://wiki.services.openoffice.org/wiki/Documentation/FR/Calc/Fonctions_listees_par_categories)

Fonctions de base de données : utilisées pour extraire des informations de tables de données Calc, dans lesquelles les données sont organisées en lignes.

BDNB, BDNBVAL, BDLIRE, BDMAX, BDMIN, BDMOYENNE, BDPRODUIT, BDECARTYPE, BDECARTYPEP, BDSOMME, BDVAR, BDVARP

Fonctions classeur : trouvent des valeurs dans des tables ou des références de cellule.

ADRESSE, CHOISIR, RECHERCHEH, INDEX, INDIRECT, RECHERCHE, EQUIV, DECALER, RECHERCHEV, ZONES, COLONNE, COLONNES, TYPE.ERREUR, INFO, LIGNE, LIGNES, FEUILLE, FEUILLES, DDE, LIEN_HYPERTEXTE, STYLE

Fonctions de conversion de nombres : convertissent des nombres sous plusieurs formes.

BINDEC, BINHEX, BINOCT, DECBIN, DECHEX, DECOCT, HEXBIN, HEXDEC, HEXOCT, OCTBIN, OCTDEC, OCTHEX, ARABE, BAHTTEXT, DECIMAL, ROMAIN

Fonctions de date et d'heure

AUJOURDHUI, MAINTENANT, ANNEE, MOIS, JOUR, JOURSEM, HEURE, MINUTE, SECONDE, DATE, DATEVAL, TEMPS, TEMPSVAL, FIN.MOIS, MOIS.DECALER, SERIE.JOUR.OUVRE, NB.JOURS.OUVRES, JOURS, JOURS360, JOURSDANSMOIS, JOURSDANSANNEE, SEMAINES, SEMAINESDANSANNEE, NO.SEMAINE, NO.SEMAINE_ADD, NB.MOIS, ANNEES, ESTBISSEXTILE, DIMANCHEDEPAQUES, FRACTIONANNEE

Fonctions financières : fournissent des calculs courants dans le monde de la finance.

AMORDEGRC, AMORLINC, DB, DDB, AMORLIN, SYD, VDB, CUMUL.INTER, CUMUL.INTER_ADD, CUMUL.PRINCIPER, CUMUL.PRINCIPER_ADD, VC, VC.PAIEMENTS, INTPER, TRI, ISPMT, TRIM, NPM, VAN, VPM, PRINCIPER, VA, TAUX, RENTINVEST, TRI.PAIEMENTS, VAN.PAIEMENTS, INTERET.ACC, INTERET.ACC.MAT, TAUX.ESCOMPTE, DUREE, DUREE_ADD, TAUX.EFFECTIF, TAUX.EFFECTIF_ADD, TAUX.INTERET, DUREE.MODIFIEE, TAUX.NOMINAL, TAUX.NOMINAL_ADD, PRIX.PCOUPON.IRREG, REND.PCOUPON.IRREG, PRIX.DCOUPON.IRREG, REND.DCOUPON.IRREG, PRIX.TITRE, VALEUR.ENCAISSEMENT, PRIX.TITRE.ECHANGE, VALEUR.NOMINALE, TAUX.ESCOMPTE.R, PRIX.BON.TRESOR, RENDEMENT.BON.TRESOR, RENDEMENT.TITRE, RENDEMENT.SIMPLE, RENDEMENT.TITRE.ECHEANCE, NB.JOURS.COUPON.PREC, NB.JOURS.COUPONS, NB.JOURS.COUPON.SUIV, DATE.COUPON.SUIV, NB.COUPONS, DATE.COUPON.PREC, PRIX.DEC, PRIX.FRAC

Fonctions d'information : renvoient des informations à propos de la cellule, par exemple si elles contiennent du texte ou une formule, etc.

CELLULE, ACTUELLE, FORMULE, ESTVIDE, ESTERR, ESTERREUR, EST.PAIR, EST.PAIR_ADD, ESTFORMULE, ESTLOGIQUE, ESTNA, ESTNONTEXTE, ESTNUM, EST.IMPAIR, EST.IMPAIR_ADD, ESTREF, ESTTEXTE, N, NA, TYPE

Fonctions logiques

T, FAUX, SI, NON, OU, VRAI

Fonctions mathématiques

COS, SIN, TAN, COT, ACOS, ACOT, ASIN, ATAN, ATAN2, DEGRES, RADIANS, PI, COSH, SINH, TANH, COTH, ACOSH, ACOTH, ASINH, ATANH, TRONQUE, ARRONDI, ARRONDI.INF, ARRONDI.SUP, PLAFOND, PLANCHER, PAIR, IMPAIR, ARRONDI.AU.MULTIPLE, ENT, QUOTIENT, MOD, EXP, PUISSANCE, LOG, LN, LOG10, ABS, COMBIN, COMBINA, CONVERTIR, CONVERT_ADD, NB.VIDE, NB.SI, DELTA, ERF, ERFC, FACT, FACTDOUBLE, PGCD, PGCD_ADD, SUP.SEUIL, EST.PAIR, EST.IMPAIR, PPCM, PPCM_ADD, MULTINOMIALE, PRODUIT, ALEA, ALEA.ENTRE.BORNES, SOMMES.SERIES, SIGNE, RACINE, RACINEPI, SOUS.TOTAL, SOMME, SOMME.SI, SOMME.CARRES, BESSELJ, BESSELJ, BESSELK, BESSELY

Fonctions de matrice : renvoient et opèrent sur des tables entières de données.

FREQUENCE, CROISSANCE, DROITEREG, LOGREG, DETERMAT, INVERSEMAT, PRODUITMAT, MUNIT, SOMMEPROD, SOMME.X2MY2, SOMME.X2PY2, SOMME.XMY2, TRANSPONSE, TENDANCE

Fonctions de nombre complexe : opèrent sur des nombres complexes (mathématiques) tels que $2 + 3i$.

COMPLEXE, COMPLEXE.MODULE, COMPLEXE.IMAGINAIRE, COMPLEXE.PUISSANCE, COMPLEXE.ARGUMENT, COMPLEXE.COS, COMPLEXE.DIV, COMPLEXE.EXP, COMPLEXE.CONJUGUE, COMPLEXE.LN, COMPLEXE.LOG10, COMPLEXE.LOG2, COMPLEXE.PRODUIT, COMPLEXE.REEL, COMPLEXE.SIN, COMPLEXE.DIFFERENCE, COMPLEXE.SOMME, COMPLEXE.RACINE

Fonctions statistiques : fournissent des calculs de statistique et de probabilité.

ECART.MOYEN, MOYENNE, MOYENNEA, NB, NBVAL, MOYENNE.GEOMETRIQUE, MOYENNE.HARMONIQUE, MEDIANE, MODE, MOYENNE.REDUITE, SOMME.CARRES.ECARTS, KURTOSIS, COEFFICIENT.ASYMETRIE, ECARTYPE, ECARTYPEA, ECARTYPEP, ECARTYPEPA, VAR VARA VAR.P VAR.PA, B, LOI.BETA, BETA.INVERSE, LOI.BINOMIALE, KHIDEUX.INVERSE, TEST.KHIDEUX, LOI.KHIDEUX, INTERVALLE.CONFIANCE, CRITERE.LOI.BINOMIALE, LOI.EXPONENTIELLE, INVERSE.LOI.F, FISHER, FISHER.INVERSE, LOI.F, TEST.F, LOI.GAMMA.INVERSE, LNGAMMA, LOI.GAMMA, GAUSS, LOI.HYPERGEOMETRIQUE, LOI.LOGNORMALE.INVERSE, LOI.LOGNORMALE, LOI.BINOMIALE.NEG, LOI.NORMALE.INVERSE LOI.NORMALE, LOI.NORMALE.STANDARD.INVERSE, LOI.NORMALE.STANDARD, PERMUTATION, PERMUTATIONA, PHI, LOI.POISSON, PROBABILITE, CENTREE.REDUITE, LOI.STUDENT, LOI.STUDENT.INVERSE, TEST.STUDENT, LOI.WEIBULL, TEST.Z, COEFFICIENT.CORRELATION, COVARIANCE, PREVISION, ORDONNEE.ORIGINE, PEARSON, COEFFICIENT.DETERMINATION, PENTE, ERREUR.TYPE.XY, GRANDE.VALEUR, MAX, MAXA, MIN, MINA, CENTILE, RANG.POURCENTAGE, QUARTILE, RANG, PETITE.VALEUR

Fonctions de texte : manipulent ou renvoient un texte.

BASE, CAR, EPURAGE, CODE, CONCATENER, FRANC, EXACT, TROUVE, CTXT, GAUCHE, NBCAR, MINUSCULE, STXT, NOMPROPRE, REMPLACER, REPT, DROITE, CHERCHE, SUBSTITUE, T, TEXTE, SUPRESPACE, MAJUSCULE, CNU

Pour choisir la fonction adaptée à ses besoins, l'utilisateur peut consulter l'aide directement accessible depuis le menu. Pour Excel, il peut accéder depuis le sommaire à un résumé succinct de toutes les fonctions de chaque catégorie quand son intitulé n'est pas suffisamment précis (voir Fig. 3.8), puis suivre l'hyperlien de la fonction qui lui paraît la plus appropriée. Il dispose alors des détails de la fonction (définition succincte, syntaxe) et le plus souvent d'un exemple qui peut être copié directement dans une feuille de calcul.

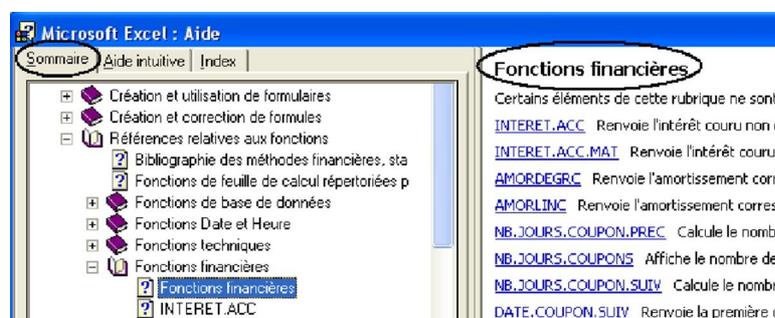


Figure 3.8 : Définitions sommaires des fonctions financières dans l'aide d'Excel 2003

Pour OpenOffice 3.1.0, l'aide du sommaire donne des informations détaillées par groupe de fonctions. Dans la figure 3.9, on a encadré un exemple de formule utilisant la fonction PRIX.PCOUPON.IRREG que l'utilisateur peut directement recopier dans une cellule de la feuille de calcul.

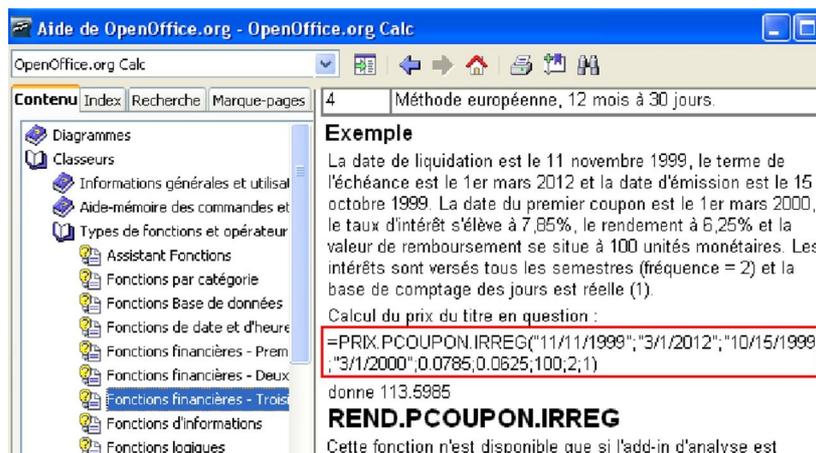


Figure 3.9 : Exemple d'utilisation d'une fonction financière dans l'aide d'OpenOffice 3.1



Exemple : Exemple corrigé de la question 3 du problème N°3 pour le calcul de remboursement de crédit

D. Formule conditionnelle

Trois fonctions permettent d'écrire des formules conditionnelles :

- SI(test_logique; valeur_si_vrai; valeur_si_faux)
- SOMME.SI(plage; critère; somme_plage)
- NB.SI(plage; critère)

1. La fonction SI

Elle est principalement utilisée pour écrire l'une ou l'autre de deux valeurs différentes dans une cellule, selon la valeur d'une autre cellule. Par exemple, si on écrit dans une cellule le genre (M/F), à l'aide de la fonction SI, on peut afficher la civilité qui sera « Monsieur » ou « Madame » selon les cas (voir fig.3.11).

fx =SI(C4="M","Monsieur","Madame")				
	A	B	C	D
3		Nom	Genre (M/F)	Civilité
4		Leduc	M	Monsieur
5		Leroux	F	Madame

Figure 3.11 : Utilisation de la fonction conditionnelle « SI »

Pour avoir le choix entre plus de deux alternatives, il suffit d'emboîter les fonctions « SI ». Un exemple d'utilisation de 2, SI imbriqués pour donner le choix entre 3 civilités (Mr., Mme ou Mlle) est en figure 3.12.

E4	=SI(C4="M";"Mr.";SI(D4="C";"Melle";"Mme"))			
	B	C	D	E
3	Nom	Genre (M/F)	Statut marital (M, C, D, V)	Civilité
4	Leduc	M	M	Mr.
5	Leroux	F	C	Melle
6	Poirier	F	D	Mme

Figure 3.12 : Utilisation de deux « SI » imbriqués

Dans la formule conditionnelle =SI(C4="M";"Mr.";SI(D4="C";"Melle";"Mme")) de la figure 3.11, c'est la valeur « valeur_si_faux » qui contient une formule, les 2 autres arguments de la fonction SI, « test_logique » et « valeur_si_vrai », étant de simples valeurs, booléenne pour la première, textuelle pour la seconde. Ce peuvent être également des formules. Toutefois les formules trop compliquées sont difficiles à corriger. Il est préférable alors de les découper en sous-formules chacune dans une cellule différente.

2. Exercice 1

Question 1

[Solution n°35 p 161]

1 :

Un nombre est écrit dans la cellule B3, trouver la formule à écrire en C3 qui affiche la racine carrée de ce nombre s'il est positif et un message « Le nombre doit être positif » dans le cas contraire

Question 2

[Solution n°36 p 161]

2 :

En D3, on veut afficher la valeur de fonction $x \log(x)$ quand x est positif, 0 quand il est nul, et un message d'erreur quand x est négatif, la valeur de x est prise dans la cellule B3.

Question 3

3 :

Le jeu « deviner un nombre » : un nombre est écrit dans une cellule, B3 par exemple, invisible car écrit en rouge sur fond rouge. L'utilisateur saisit une valeur en E3. Un message s'affiche en F3 pour donner une indication au joueur sur la position du nombre qu'il a proposé par rapport au nombre trouvé : selon que cette valeur est égale (resp. inférieure, supérieure) au nombre caché, le message est Bravo (resp. Trop petit, Trop grand). Il peut alors saisir une nouvelle valeur en E4. On ne prévoit que 5 essais, et la partie s'achève soit dès que l'utilisateur a gagné, soit au bout de 5 essais s'il a perdu les 4 premiers. Un bilan de la partie s'affiche alors en dessous, « Bravo, vous avez gagné en x parties » ou « Pas de chance, vous avez perdu au bout de 5 essais ».

Question 4

[Solution n°37 p 161]

4 :

On saisit par des codes les réponses de 9 personnes à la question « aimez-vous les sports d'hiver ? ». dans la colonne C (voir fig.3.14). Les réponses possibles sont : 0 :pas du tout, 1 : très peu, 2 :un peu, 3 : beaucoup, 4 : énormément (voir fig.3.14, colonne C). On désire que la signification de chaque code s'affiche à côté du code (colonne D). Quelle formule doit-on taper dans la cellule D4 ?

	C	D	E
1		2	un peu
2		1	très peu
3		3	beaucoup
4		4	énormément
5		0	pas du tout
6		5	énormément
7		7	énormément
8		2	un peu
9		3	beaucoup

Figure 3.14 : Affichage de libellés correspondant à des codes.

3. La fonction NB.SI

Comme son nom l'indique, elle permet de faire des comptages de cellules vérifiant une condition donnée. Sa syntaxe est la suivante :

NB.SI(plage; critère)

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1																								
2	Id	1	2	3	4	5	6	7	8	9	10	11	12	Comptages										
3	Sexe	M	F	F	M	M	F	F	F	F	M	F	F	Sexe	Rep	1	2	3	4	5	Tot			
4	Q1	3	2	5	1	4	4	2	2	2	4	3	1	M	4	Q1	2	4	2	3	1	12		
5	Q2	1	1	3	2	2	4	5		2	2	3	5	F	8	Q2	2	4	2	1	2	11		
6	Q3	2	4	3	2			2	1	2	3	5	2			Q3	1	5	2	1	1	10		

Figure 3.16 : Utilisation de la fonction NB.SI

Dans la figure 3.16, on a utilisé la fonction NB.SI pour faire des comptages de valeurs saisies à trois questions d'un questionnaire. La formule dans R4 est =NB.SI(D3:O3; "M") et dans U4 , =NB.SI(D4:O4;1). Dans le premier cas, on compte le nombre de personnes de sexe= M, et dans le second de Q1=1.

On peut aussi écrire des formules qui contiennent un critère plus « élaboré ». On peut donner comme critère une adresse de cellule :

dans R4 : =NB.SI(D3:O3;Q4), dans U4 : =NB.SI(D4:O4;U3)

Ce qui permet de les recopier automatiquement dans la cellule R5 pour la première formule, et dans les cellules U4 à Y6 pour la seconde, une fois les adresses corrigées comme il convient avec des \$.

Il y a une troisième façon d'écrire un critère : avec une formule comprise entre deux guillemets. Par exemple si on désire connaître le nombre de réponses « correctes » pour les trois question, c'est-à-dire inférieures ou égales à 5, on utilise la formule =NB.SI(D4:O6;"<=5"). La formule figurant entre les guillemets est très sommaire. Elle comporte un signe (ou deux, comme <> pour « différent de ») et un nombre. On ne peut pas compter le nombre de valeurs entre 1 et 5 par exemple. Il faut faire une soustraction : =NB.SI(D4:O6;"<=5") - NB.SI(D4:O6;"<1").

4. Exercice 2

Question 1

[Solution n°38 p 161]

1 :

Écrire la formule permettant de trouver le nombre de valeurs 5 de chaque personne aux 3 questions. Puis leur (voir fig.3.17).

Question 2

[Solution n°39 p 162]

2 :

Écrire la formule permettant de connaître le nombre de personnes ayant 0, 1, 2, ou 3 valeurs 5.

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
2	Id	1	2	3	4	5	6	7	8	9	10	11	12	Comptages									
3	Sexe	M	F	F	M	M	F	F	F	F	M	F	F	Sexe	Rep	1	2	3	4	5			
4	Q1	3	2	5	1	4	4	2	2	2	4	3	1	M	4	Q1	2	4	2	3	1		
5	Q2	1	1	3	2	2	4	5		2	2	3	5	F	8	Q2	2	4	2	1	2		
6	Q3	2	4	3	2				2	1	2	3	5	2		Q3	1	5	2	1	1		
7														Sexe	nombre de 5								
8	5	0	0	1	0	0	0	1	0	0	0	1	1	M	0			0	1	2	3		
9	2,3,4	2	2	2	2	2	2	2	1	3	3	2	1	F	1,9			8	4	0	0		

Figure 3.17: De D8 à O8 (resp. D9 :O9), nombre de réponses 5 (resp. 2,3 ou 4) aux 3 questions, en U8,

5. La fonction SOMME.SI

Elle fonctionne sur le même principe que la fonction NB.SI, mais les comptages sont remplacés par des sommes de valeurs. Sa syntaxe est la suivante :

SOMME.SI(plage; critère; somme_plage)

Par exemple, =SOMME.SI(D3:O3; "M"; D4:O4) affiche la somme des valeurs à la question 1 pour les hommes.

6. Exercice 3

Question

[Solution n°40 p 162]

Donner la formule qui permet de calculer la moyenne de valeurs 5 aux 3 questions pour les hommes et les femmes (cellules R8 et R9, fig. 3.17)

7. Exercice 4

Nous proposons maintenant un exercice d'application des ces diverses formules conditionnelles.

Question

[Solution n°41 p 162]

M. Paul Gagnant est chargé de gérer les comptes "actions" de cinq clients (voir Fig. 3.10). Il les a interrogés sur leur désir d'acquérir des actions « StartPlay », et la réponse de chacun figure en face de son nom dans la colonne D. Il dispose d'informations financières sur ses clients (colonnes B et C). Connaissant par ailleurs le montant unitaire de l'action (cellule B1), il écrit en E4 la formule conditionnelle permettant de calculer le nombre d'actions achetées par le premier client (Dupont Hortense) en fonction de tous ces éléments., formule recopiée automatiquement pour les autres clients. Puis dans les lignes 10 à 12, il fait un bilan.

	A	B	C	D	E
1	Action StartPlay :	125 €	Ventes : M. Paul Gagnant		
2					
3	Noms	Crédit du compte	Débit autorisé	accord	Quantité vendue
4	Dupont Hortense	1 800 €	0 €	oui	14
5	Leduc Adrien	200 €	300 €	non	0
6	Orpailleur Gilles	-200 €	1 000 €	oui	6
7	Fernandel Amélie	500 €	300 €	non	0
8	Charlot hubert	-200 €	200 €	oui	0
9					
10		nombre de personnes d'accord		3	
11		crédit total disponible sur accord		1 400 €	
12		nombre de personnes ayant acheté		2	

Figure 3.10 : Exemple d'utilisation d'une fonction financière dans l'aide d'OpenOffice 3.1

E. Calcul matriciel

Une partie des formules vues précédemment (+, moyenne, recherche, nb.si...) peuvent opérer sur des ensembles de cellules. Quand ces ensembles de cellules sont sous forme de tableaux rectangulaires et contiennent des nombres, on les appelle « matrices » et on peut leur utiliser dans des formules matricielles spécifiques. Quand la formule matricielle a pour résultat un nombre (par ex. moyenne), l'utilisateur manipule généralement les matrices de façon transparente. Par contre, quand le résultat est une matrice, qui s'écrit dans une plage rectangulaire de cellules, l'utilisateur doit faire les manipulations nécessaires.

Le calcul matriciel peut être vu comme une technique permettant de remplacer une série de calculs sur des « scalaires » (nom désignant les nombres ordinaires par opposition aux matrices) par un seul calcul sur des « matrices » ou des « vecteurs » (nom donné aux matrices qui comportent une seule ligne ou une seule colonne). C'est ce point de vue opératoire que nous privilégions ici, plutôt que de nous mettre dans le cadre formel de l'algèbre linéaire d'où les matrices sont issues. De ce fait, ce cours n'exige pas de connaissances spécifiques en mathématiques, la pratique étant privilégiée à la théorie.

1. Définitions et manipulations

Il y a 2 types de tableaux :

- les tableaux formés d'une plage rectangulaire pouvant comporter des cellules vides, et également des cellules contenant d'autres types de données que des nombres
- les matrices qui sont des tableaux ne contenant que des cellules avec des nombres

Quand la fonction rend un tableau ou une matrice, il convient de procéder en 3 étapes successives :

1. sélectionner la plage de cellules devant contenir le tableau
2. taper la formule de la fonction précédée du signe =
3. valider en appuyant simultanément sur CTRL+MAJ+ENTREE

la formule s'entoure alors d'accolades, et le tableau se remplit.

La figure 3.19 montre la réalisation d'une somme matricielle (somme termes à

termes de deux tableaux de nombres). L'image en haut à gauche correspond à l'étape 0, les deux en dessous à l'étape 1, et l'image à droite montre le résultat obtenu après l'étape 2.

Conseil



Écrire des formules matricielles nécessite un certain doigté plus qu'une profonde compréhension du principe, les manipulations étant parfois surprenantes, comme celles pour obtenir l'affichage du type matrice (voir figure 3.20) ! En cas d'accident, il ne faut pas se décourager et recommencer depuis le début : effacer toutes les cases du tableau, en appuyant sur SUPPR, et de recommencer la manipulation.

Le type matrice a pour code 64 (voir fig.3.20)

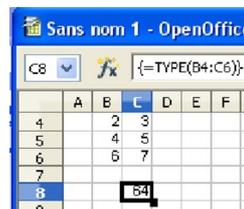


Figure 3.20 : Le type matrice, affiché sous OpenOffice 2.4

2. Opérateurs et fonctions matricielles

Les opérateurs arithmétiques fonctionnent aussi avec des matrices de même taille, et rendent les matrices obtenues en appliquant les opérateurs termes à termes (Fig. 3.21). Pour les utiliser on procède comme indiqué dans le paragraphe précédent pour l'addition de deux matrices (Fig. 3.20). Plus généralement, on peut aussi appliquer des fonctions numériques à une matrice (comme « sin ») et obtenir une matrice ayant comme termes les sinus des termes de la matrice d'origine.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	A	B	=A+B	=A-B	=A	=-A	=A*B	=A/B	=A^B	=sin(A)																				
2	1 4	-1 4	0 8	2 0	1 4	1 -4	-1 16	0 32	1 256	0,84 -0,76																				
3	3 5	3 -2	6 3	0 7	3 5	-3 2	9 -10	18 -6	27 0,04	0,14 -0,96																				
4	2 6	1 -1	3 5	1 7	2 6	-1 1	2 -6	3 -5	2 0,17	0,91 -0,28																				

Figure 3.21 : Les opérateurs de scalaires appliqués aux matrices

Remarque



On peut également utiliser les opérateurs arithmétiques de façon mixte, c'est-à-dire en mêlant scalaires et matrices dans les opérations. On obtient dans la plupart des cas des matrices alors que l'écriture des formules n'est pas toujours en adéquation avec les normes mathématiques usuelles (Fig. 3.22).

Figure 3.22 : Les opérateurs de scalaires utilisés de façon mixte scalaire/matrice

On dispose d'un « vrai » calcul matriciel (selon les normes mathématiques) sous forme de fonctions, dont les plus intéressantes sont les trois premières du tableau

ci-dessous, car elle permettent de réaliser en quelques clics des calculs qui peuvent être longs et complexes. Les autres fonctions du tableau sont plutôt utilisées pour les calculs statistiques.

Nom de la fonction	Rôle de la fonction	Types de données en entrée	Type de données en sortie	Conditions d'application
PRODUITMAT	Calcule le produit de deux matrices	2 matrices, la première (n,p), la seconde (p,q)	Matrice (n,q)	Dimensions des matrices compatibles
INVERSEMAT	Donne l'inverse d'une matrice	1 matrice (n,n)	Matrice (n,n)	Matrice carrée et inversible
DETERMAT	Donne le déterminant d'une matrice	1 matrice (n,n)	Scalaire	Matrice carrée
SOMMEPROD	Donne la somme des produits des éléments de matrice correspondants	2 matrices (n,p)	Scalaire	Dimensions des matrices identiques
SOMME.X2MY2	Renvoie la somme de la différence des carrés des valeurs correspondantes de deux matrices	2 matrices (n,p)	Scalaire	Dimensions des matrices identiques
SOMME.X2PY2	Renvoie la somme de la somme des carrés des valeurs correspondantes de deux matrices	2 matrices (n,p)	Scalaire	Dimensions des matrices identiques
SOMME.XMY2	Renvoie la somme des carrés des différences entre les valeurs correspondantes de deux matrices	2 matrices (n,p)	Scalaire	Dimensions des matrices identiques
TRANSPOSE	Donne la transposée d'un tableau	1 tableau (n,p)	Tableau (p,n)	Aucune

Tableau 6 : Tableau 6.2 : tableau de fonctions

À ces fonctions, s'ajoutent les fonctions de « classeur » qui permettent de chercher des informations dans la matrice ou dans le tableau de nombres. Par exemple la

fonction INDEX, qui utilise des coordonnées pour trouver une valeur à l'intérieur d'un tableau, s'utilise quand on veut accéder de façon plus individuelle aux cellules d'un tableau, la fonction RECHERCHE permet de « recoder » en remplaçant les valeurs présentes dans une colonne (resp. ligne) par les valeurs correspondantes présentes dans une autre colonne (resp. ligne), ces dernières, numériques ou textuelles, devant être rangées dans l'ordre croissant.

3. Exercice TD

Question 1

[Solution n°42 p 162]

1 :

On donne les coefficients de 4 épreuves (Fig 3.23, cellules C2 à F2) composant un examen et les notes de 3 étudiants à ces épreuves (C3 à F5). Utiliser des fonctions matricielles de type « somme » pour trouver pour chaque étudiant 1) sa moyenne pondérée à l'examen (colonne H), 2) sa distance euclidienne aux autres étudiants (cellules K3 à M3). Trouver les formules matricielles permettant de lier les intitulés de ligne et de colonnes du tableau de distances euclidiennes aux 3 noms saisis en colonne B.

	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Notes	Coefficients				Total	Moyenne	Distances euclidiennes					
2	Etudiants	2	4	3	1	10				Jean	Marie	Paul	
3	Jean	8	10	15	16	117	11,7	Jean	0	235	63		
4	Marie	12	11	8	3	95	9,5	Marie	235	0	70		
5	Paul	11	13	12	10	120	12	Paul	63	70	0		

Figure 3.23 : Calculs illustrant l'utilisation de fonctions matricielles de type « somme »

Question 2

[Solution n°43 p 163]

2 :

Résolution d'un système de n équations à n inconnues. : On le met sous de produit matriciel $AX=B$, et la solution, si elle existe et est unique, est donnée par $X=A^{-1}B$. Résoudre le système de 3 équations $2x + 4y = 2$; $3x + z = -1$; $x - 2y + 4z = 3$. Une fois trouvée la valeur de la matrice X, calculer $AX-B$ qui doit être nul.

F. Exercices appliqués à l'économie avec correction

1. Problème 7 : Entreprise, deux usines, effet de structure

Une entreprise possède deux usines, l'une en France et l'autre en Chine. Les salaires différents selon le pays sont donnés par les tableaux suivants, en France et en Chine pour les années 2000 et 2010.

	année 2000		année 2010	
Cadres	20	[2400-3600[10	[2600-4000[
Agents de Maîtrise	80	[1600-2400[30	[2000-2600[
Ouvriers	100	[1400- 1600[60	[1400-2000[

Tableau 7 : Usine en France

	année 2000		année 2010	
Cadres	20	[600-1400[40	[750-1450[
Agents de Maîtrise	80	[200-600[160	[250-750[
Ouvriers	100	[100-200[200	[120-240[

Tableau 8 : Usine en Chine

La répartition des salaires est homogène dans chaque groupe.

Question 1

[Solution n°44 p 163]

Préparer la structure du tableau

Question 2

[Solution n°45 p 163]

1) Tracer l'histogramme des salaires de l'entreprise (les deux usines) en 2000.

Question 3

[Solution n°46 p 164]

2) Déterminez le salaire médian en 2000

Question 4

[Solution n°47 p 164]

3) Calculez le salaire moyen en 2000, en France, en Chine et de l'entreprise.

Les salaires ont augmenté entre 2000 et 2010 aussi bien en France qu'en Chine.

Question 5

[Solution n°48 p 164]

4) Calculez l'augmentation moyenne des salaires en France, en Chine. Qu'en pensez-vous ?

Question 6

[Solution n°49 p 164]

5) Calculez le salaire moyen de l'ensemble de l'entreprise. Qu'en pensez-vous ?

Question 7

[Solution n°50 p 164]

6) Expliquez les résultats obtenus aux questions 4 et 5

2. Problème 8 : Règles conditionnelles

Soit le tableau suivant donnant les notes des 820 étudiants de première année d'économie et gestion ayant présenté les épreuves.

Les notes de mathématiques sont comprises entre 5 et 17, et les notes d'économie entre 7 et 15. Le tableau indique les effectifs de chaque catégorie.

		Mathématiques									
	Notes	5	7	9	10	15	17				
	7	28	36	17	25	2	1				
	9	24	57	59	12	25	3				
Economie	11	11	21	27	34	53	17				
	12	2	5	15	28	34	15				
	14	1	5	12	15	32	64				
	15	1	3	11	18	49	58				

Tableau 9 : Notes des étudiants de 1ère année d'économie et gestion ayant présenté les épreuves

Question 1

[Solution n°51 p 164]

Préparer la structure du tableau

Question 2

[Solution n°52 p 164]

1) Déterminez les effectifs d'étudiants ayant obtenu 10 en mathématiques.

Question 3

[Solution n°53 p 164]

2) Parmi ces derniers, combien ont obtenu plus de la moyenne en économie ?

Question 4

[Solution n°54 p 164]

3) Quelle est la note moyenne en économie des étudiants ayant obtenu 10 en mathématiques ?

Question 5

[Solution n°55 p 164]

4) Quelle est la note moyenne en économie des étudiants ayant obtenu la moyenne en mathématiques (10 ou plus de 10) ?

Question 6

[Solution n°56 p 164]

5) Quelle est la note moyenne en économie de l'ensemble des étudiants ?

Question 7

[Solution n°57 p 164]

6) Quelle est la note moyenne en mathématiques de l'ensemble des étudiants ?

Question 8

[Solution n°58 p 164]

7) Parmi les étudiants qui ont la moyenne en économie, quel est le pourcentage d'étudiants qui ont la moyenne en mathématiques ?

3. Problème 9 : Calcul des prix de revient à partir d'Inputs

Une entreprise de menuiserie fabrique trois types de produits : des tables (T), des chaises (C) et des armoires (A).

Cette entreprise a des frais fixes de 10000 € par mois, quelle a décidé de répartir entre les unités produites (exemples, elle produit 200 objets, les coûts fixes représentent 50 € par objet).

Chaque objet utilise un certain nombre de facteurs variables (les inputs) : des planches en bois, des paquets de vis et de chevilles, des tiges métalliques, pour les pieds de tables et de chaises.

Les quantités nécessaires sont données par le tableau suivant :

	Planches bois	Vis	Tiges
Table	8	3	4
Chaise	1	2	2
Armoire	30	9	0

Tableau 10 : Quantités nécessaires

Lire : pour une table il faut : 4 planches en bois, 3 vis et 4 tiges

Le prix de chaque input est donné par le tableau suivant :

	Coût Unitaire (en €)
Planches	10
Visses	2
Tiges	4

Tableau 11 : Prix de chaque input

Question 1

[Solution n°59 p 164]

1) Sachant que l'entreprise à fabriqué le mois dernier 50 tables, 320 chaises et 30 armoires, calculez le prix de revient de chaque article.

Question 2

[Solution n°60 p 165]

2) Les prix de vente des produits sont donnés par le tableau suivant.

	Prix de vente (€)
Tables	130
Chaises	50
Armoires	350

Tableau 12 : Prix de vente des produits

Calculez le bénéfice de l'entreprise le mois dernier.

Question 3

[Solution n°61 p 165]

3) L'entreprise a trouvé un sous-traitant qui propose de fabriquer des tables pour elle, lui permettant de fabriquer plus de chaises et d'armoires. Le sous-traitant réclame 131 € par table, qu'en pensez-vous ?

Question 4

[Solution n°62 p 165]

4) Le mois suivant, le chef d'entreprise décide de sous-traiter les tables et avec la même équipe produira : 750 chaises et 50 armoires. Calculez les prix de revient. Qu'en pensez-vous ?

Question 5

[Solution n°63 p 165]

5) Calculez le profit de l'entreprise, sachant qu'elle doit vendre 100 tables, avec ses chaises. Expliquez le résultat obtenu.

4. Problème 10 : Fichier d'une enquête sur le premier salaire

On se propose d'étudier le premier salaire de jeunes sortis, depuis 18 mois, d'un BTS en apprentissages, à partir d'un extrait d'une enquête (voir enquête premier salaire feuille 2).

Ont été retirées : les non réponses à certaines questions, les jeunes qui poursuivent des études ou qui sont à la recherche d'un emploi, ainsi que les résultats aberrant. Nous traiterons donc les 1560 réponses à l'enquête qui nous concernent.

Question 1

[Solution n°64 p 165]

1) A partir de ces données, construire un tableau indiquant les salaires des hommes et des femmes.

Question 2

[Solution n°65 p 165]

2) Calculez le salaire moyen de chaque sexe, sachant que la distribution est homogène à l'intérieur de chaque classe.

Cet exercice sera repris en chapitre 4, pour des applications de tableaux croisés dynamiques et pour des représentations graphiques.

G. QCM de la leçon 3

Exercice 1

[Solution n°6 p 198]

Question 1 :

	B	C	D	E	F
4	x	y	z	t	Formules
5	4	0	2	8	=somme(B\$7:B\$10)
6	3	4	3	6	=SOMME(C5:C8)
7	0	3	3	9	=SI(B7+E7<=5;B4;SI(C7+D7>5;10;0))
8	2	2	0	d	=NB.SI(C5:E13;2)
9	1	5	0	36	=MOYENNE(\$B\$5:\$D\$5)
10	2	b	c	-3	=NB(C9;D11)
11	1	1	0	1	
12	0	1	1	0	
13	2	0	1	2	

Figure 15 : Qcm feuille de calcul

Il y a , dans la feuille de calcul ci-dessus, des données dans 4 colonnes, et des formules dans la cinquième. On demande le résultat que contiennent les cellules, une fois la formule entrée. Si c'est un nombre, sélectionner le nombre le plus proche. S'il tombe exactement entre 2 nombres proposés, les sélectionner tous les 2, si c'est une lettre, sélectionner la réponse lettre, sinon sélectionner la réponse "autre", que ce soit un texte, une erreur, ou autre chose.

Dans la cellule F5 :

- 0
- 1
- 2
- 3
- 4
- 5
- 10
- 30
- lettre
- autre

Exercice 2

[Solution n°7 p 198]

Question 2 :

	B	C	D	E	F
4	x	y	z	t	Formules
5	4	0	2	8	=somme(B\$7:B\$10)
6	3	4	3	6	=SOMME(C5:C8)
7	0	3	3	9	=SI(B7+E7<=5;B4;SI(C7+D7>5;10;0))
8	2	2	0	d	=NB.SI(C5:E13;2)
9	1	5	0	36	=MOYENNE(\$B\$5:\$D\$5)
10	2	b	c	-3	=NB(C9;D11)
11	1	1	0	1	
12	0	1	1	0	
13	2	0	1	2	

Figure 15 : Qcm feuille de calcul

Il y a , dans la feuille de calcul ci-dessus, des données dans 4 colonnes, et des formules dans la cinquième. On demande le résultat que contiennent les cellules, une fois la formule entrée. Si c'est un nombre, sélectionner le nombre le plus proche. S'il tombe exactement entre 2 nombres proposés, les sélectionner tous les 2, si c'est une lettre, sélectionner la réponse lettre, sinon sélectionner la réponse "autre", que ce soit un texte, une erreur, ou autre chose.

Dans la cellule F6 :

Leçon 3 :

- 0
- 1
- 2
- 3
- 4
- 5
- 10
- 30
- lettre
- autre

Exercice 3

[Solution n°8 p 199]

Question 3 :

	B	C	D	E	F
4	x	y	z	t	Formules
5	4	0	2	8	=SOMME(B\$7:B\$10)
6	3	4	3	6	=SOMME(C5:C8)
7	0	3	3	9	=SI(B7+E7<=5;B4;SI(C7+D7>5;10;0))
8	2	2	0	d	=NB.SI(C5:E13;2)
9	1	5	0	36	=MOYENNE(\$B\$5:\$D\$5)
10	2	b	c	-3	=NB(C9;D11)
11	1	1	0	1	
12	0	1	1	0	
13	2	0	1	2	

Figure 15 : Qcm feuille de calcul

Il y a , dans la feuille de calcul ci-dessus, des données dans 4 colonnes, et des formules dans la cinquième. On demande le résultat que contiennent les cellules, une fois la formule entrée. Si c'est un nombre, sélectionner le nombre le plus proche. S'il tombe exactement entre 2 nombres proposés, les sélectionner tous les 2, si c'est une lettre, sélectionner la réponse lettre, sinon sélectionner la réponse "autre", que ce soit un texte, une erreur, ou autre chose.

Dans la cellule F7 :

- 0
- 1
- 2
- 3
- 4
- 5
- 10
- 30
- lettre
- autre

Exercice 4

[Solution n°9 p 199]

Question 4 :

	B	C	D	E	F
4	x	y	z	t	Formules
5	4	0	2	8	=somme(B\$7:B\$10)
6	3	4	3	6	=SOMME(C5:C8)
7	0	3	3	9	=SI(B7+E7<=5;B4;SI(C7+D7>5;10;0))
8	2	2	0	d	=NB.SI(C5:E13;2)
9	1	5	0	36	=MOYENNE(\$B\$5:\$D\$5)
10	2	b	c	-3	=NB(C9;D11)
11	1	1	0	1	
12	0	1	1	0	
13	2	0	1	2	

Figure 15 : Qcm feuille de calcul

Il y a , dans la feuille de calcul ci-dessus, des données dans 4 colonnes, et des formules dans la cinquième. On demande le résultat que contiennent les cellules, une fois la formule entrée. Si c'est un nombre, sélectionner le nombre le plus proche. S'il tombe exactement entre 2 nombres proposés, les sélectionner tous les 2, si c'est une lettre, sélectionner la réponse lettre, sinon sélectionner la réponse "autre", que ce soit un texte, une erreur, ou autre chose.

Dans la cellule F8 :

Leçon 3 :

0

1

2

3

4

5

10

30

lettre

autre

Exercice 5

[Solution n°10 p 200]

Question 5 :

	B	C	D	E	F
4	x	y	z	t	Formules
5	4	0	2	8	=SOMME(B\$7:B\$10)
6	3	4	3	6	=SOMME(C5:C8)
7	0	3	3	9	=SI(B7+E7<=5;B4;SI(C7+D7>5;10;0))
8	2	2	0	d	=NB.SI(C5:E13;2)
9	1	5	0	36	=MOYENNE(\$B\$5:\$D\$5)
10	2	b	c	-3	=NB(C9;D11)
11	1	1	0	1	
12	0	1	1	0	
13	2	0	1	2	

Figure 15 : Qcm feuille de calcul

Il y a , dans la feuille de calcul ci-dessus, des données dans 4 colonnes, et des formules dans la cinquième. On demande le résultat que contiennent les cellules, une fois la formule entrée. Si c'est un nombre, sélectionner le nombre le plus proche. S'il tombe exactement entre 2 nombres proposés, les sélectionner tous les 2, si c'est une lettre, sélectionner la réponse lettre, sinon sélectionner la réponse "autre", que ce soit un texte, une erreur, ou autre chose.

Dans la cellule F9 :

- 0
- 1
- 2
- 3
- 4
- 5
- 10
- 30
- lettre
- autre

Exercice 6

[Solution n°11 p 200]

Question 6 :

On désire résoudre un système de 3 équations à 3 inconnues. Pour cela, on tape la formule suivante :

=PRODUITMAT(INVERSEMAT(B11:D13);E11:E13))

Sélectionner le système dont le triplet solution sera calculé grâce à cette formule :

- système n° 1 : $y + 2t = 1$ $y + z = 0$ $z + t = 2$
- système n° 2 : $y + t = 1$ $z + t = 0$ $2y + t = 2$
- système n° 3 : $y + z = 1$ $z + t = 0$ $y + 2t = 2$
- système n° 4 : $y + t = 1$ $y + z = 0$ $z + 2t = 2$

Exercice 7

[Solution n°12 p 200]

Question 7 :

On désire résoudre un système de 3 équations à 3 inconnues. Pour cela, on tape la formule suivante :

=PRODUITMAT(INVERSEMAT(B11:D13);E11:E13))

Pour obtenir la solution, entourer la bonne méthode :

- méthode n°1 : taper la formule dans une cellule, puis recopier vers le bas pour avoir les 3 nombres.
- méthode n°2 : sélectionner une zone de 3 cellules, puis taper la formule.
- méthode n°3 : taper la formule dans une cellule, et utiliser le "copier-coller" du menu "édition", ou les combinaisons de touches Ctrl C, Ctrl V



Leçon 4 : Les outils supplémentaires de bureautique

Exercice TD	99
Les tris, les filtres	100
Les sous-totaux et tableaux croisés dynamiques	104
La recherche de valeurs cibles, le solveur	116
Exercices appliqués à l'économie avec correction	121

Dans cette leçon, nous étudions quelques outils du tableur qui permettent de construire de la connaissance à partir des données. Dans la première partie, nous disposons de tableaux, avec une première ligne d'intitulés suivie de nombreuses lignes de données numériques ou textuelles, comme les données « adults » importées dans la leçon 2, et nous réordonnons, filtrons les lignes de données afin de faire ressortir certaines informations. Dans la deuxième partie, à partir des mêmes tableaux, nous faisons des regroupements, des croisements afin de créer des vues synthétiques des données, des résumés de celles-ci. Dans la troisième et dernière partie nous réalisons des modèles de données afin de trouver des valeurs cibles, et des optima sous contraintes. L'important, dans cette dernière partie est d'être capable d'organiser l'information dans les cellules du tableur afin de faciliter la résolution du problème à l'aide des outils adaptés du tableur.

A. Exercice TD

Les données « Adults »

Dans la leçon précédente, nous avons créé le classeur correspondant à ces données, avec deux feuilles de calcul, l'une nommée « data » et l'autre « names ». Nous allons travailler avec la feuille de données « data », qui contient les valeurs selon 15 caractéristiques de 32561 personnes. Avant de travailler avec ces données, nous allons ajouter une colonne de numéros qui nous permettra de faire des contrôles chaque fois que les résultats surprenants nous feront penser à de fausses manipulations.

Question

Insérer une nouvelle colonne, avant la première colonne. Cette colonne vide est la colonne A, et nous écrivons l'intitulé « Numéro » dans la cellule A1, avant de la remplir des numéros allant de 1 en 1.

B. Les tris, les filtres

1. Les tris

Le tri du tableau de données a pour but de ranger les individus selon une de leurs caractéristiques, par exemple l'âge, dans le sens croissant ou décroissant. Les individus sont actuellement rangés par ordre croissant de numéro. Pour pouvoir changer l'ordre des individus sans changer les caractéristiques de chacun, il faut sélectionner tout le tableau (penser au raccourci clavier Ctrl *), soit les cellules de A1 à P32562, comme on peut le voir à gauche de la figure 4.3, puis faire apparaître la fenêtre de l'outil de tri par dans Données>Tri. Si les intitulés de colonnes n'apparaissent pas dans la fenêtre, il faut les faire apparaître afin que la ligne d'intitulés ne soit pas triée avec les autres lignes, mais reste bien en première ligne.

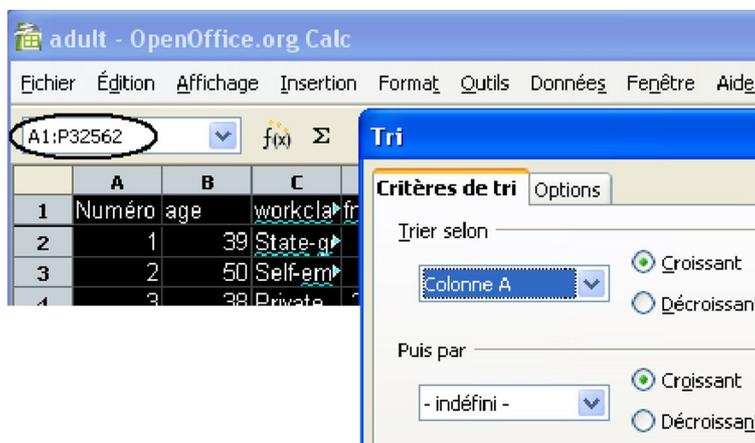


Figure 4.3 : Sélection du tableau à trier, et ouverture de la fenêtre de tri sous OpenOffice

Une fois les intitulés reconnus, on peut choisir de trier selon plusieurs critères, en général 3 au maximum (voir Fig.4.4 et 4.5). Nous choisissons de trier selon l'âge décroissant en premier. Les individus sont alors triés du plus âgé (90 ans) au plus jeune (17 ans). Comme l'âge est donné en années, et qu'il y a plus de 32000 personnes, nombreuses sont celles qui ont le même âge. Le deuxième critère que nous avons choisi est le sexe. Ce n'est pas un critère quantitatif mais qualitatif, les deux sexes étant notés par 'F' et 'M'. On peut toutefois les ranger par ordre alphabétique, et comme on a choisi le sens croissant, les femmes de 90 ans sont en premier, suivies par les hommes du même âge, etc. Puis on a pris encore un troisième critère, le niveau d'études.

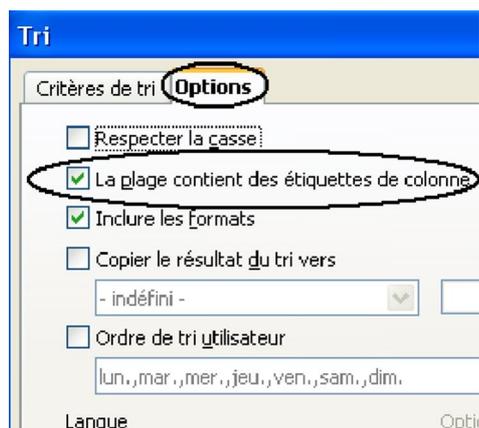


Figure 4.4 : Sélection des options du tri sous OpenOffice



Figures 4.5 : Choix des critères de tri

On remarque que dans le tableau trié, les numéros ne sont plus dans le même ordre, bien que le numéro ne fasse pas partie des critères de tri. C'est tout à fait normal car ce sont les lignes complètes qui se sont déplacées, et la personne n°1, âgée de 39 ans se retrouve maintenant avec les personnes de son âge.

adult - OpenOffice.org Calc												
	A	B	C	D	E	F	G	H	I	J	K	L
1	Numéro	age	workcla	fnlwgt	educati	educ	marital	occupa	relation	race	sex	cap
2	24239	90?		166343	1st-4th	2	Widowe	?	Not-in-f	Black	Female	
3	25304	90?		175444	7th-8th	4	Separat	?	Not-in-f	White	Female	
4	5273	90	Private	141758	9th	5	Never-n	Adm-cl	Not-in-f	White	Female	
5	11513	90	Private	87285	HS-grad	9	Never-n	Other-s	Own-ch	White	Female	
6	1041	90	Private	137018	HS-grad	9	Never-n	Other-s	Not-in-f	White	Female	
7	8964	90?		77053	HS-grad	9	Widowe	?	Not-in-f	White	Female	
8	32278	90	Private	313749	HS-grad	9	Widowe	Adm-cl	Unmarr	White	Female	
9	19213	90	Private	139660	Some-o	10	Divorce	Sales	Unmarr	Black	Female	
10	2892	90	Private	171956	Some-o	10	Separat	Adm-cl	Own-ch	White	Female	
11	4110	90?		256514	Bachel	13	Widowe	?	Other-r	White	Female	9:
12	15893	90	Private	88991	Bachel	13	Married	Exec-m	Wife	White	Female	
13	18414	90	Private	313749	Bachel	13	Never-n	Prof-sp	Own-ch	White	Female	
14	20611	90	Private	206667	Masters	14	Married	Prof-sp	Wife	White	Female	
15	18833	90	Private	115306	Masters	14	Never-n	Exec-m	Own-ch	White	Female	
16	19748	90	Private	226968	7th-8th	4	Married	Machin	Husban	White	Male	
17	32368	90	Local-g	214594	7th-8th	4	Married	Protect	Husban	White	Male	26:
18	12976	90	Private	250832	10th	6	Married	Exec-m	Husban	White	Male	
19	4071	90	Private	313986	11th	7	Never-n	Handler	Own-ch	White	Male	
32560	10282	17	Private	234780	HS-grad	9	Never-n	Farming	Own-ch	Black	Male	
32561	28631	17	Private	271837	Some-o	10	Never-n	Craft-re	Own-ch	White	Male	
32562	6889	17	Private	100828	Some-o	10	Never-n	Adm-cl	Own-ch	White	Male	

Figure 4.6 : Effet du tri à 3 critères sur les données.

2. Exercice TD 1

Question

Rerouter la personne de numéro 1 dans les données triées avec Edition>Rechercher.

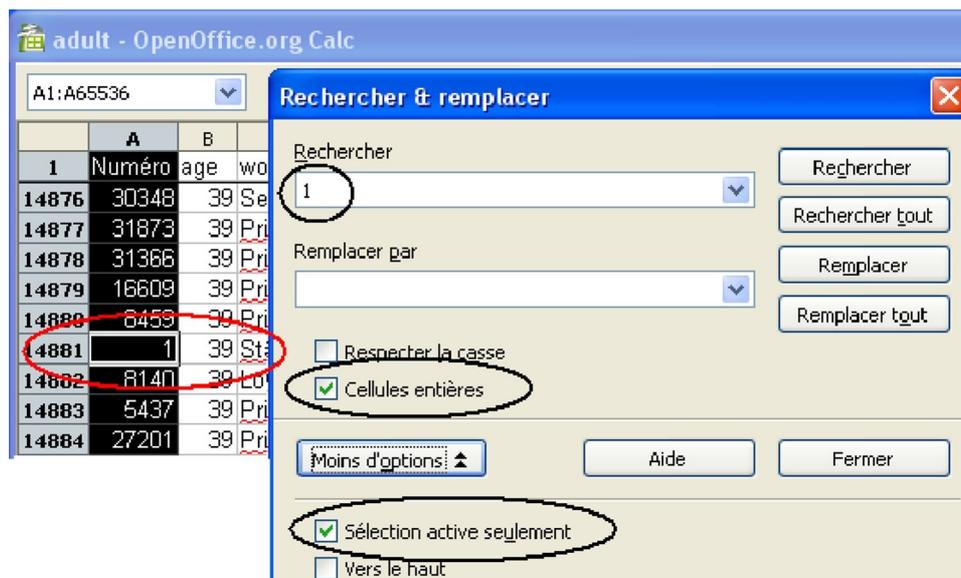


Figure 4.7 : Paramétrage de la recherche sous OpenOffice.

3. Les filtres

Il a deux façons d'utiliser le filtre, la façon automatique, et une façon plus élaborée.

a) Filtre automatique

Dans le premier cas, on sélectionne le tableau de données, puis on choisit Données>Filtrer> Filtre Automatique. Il apparaît alors une flèche avec liste de choix au sommet de chaque colonne, dans la ligne d'intitulés. Nous avons choisi de filtrer les personnes ayant 90 ans, de sexe Féminin, et ayant fait 9 années d'études, comme entouré en rouge dans la figure 4.8. Il est ressorti 4 personnes, qui sont les seules affichées dans le tableau.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Num	age	work	fnlwg	educ	educ	mar	occ	rela	ract	sex	cap	ca
1042	1041	90	Private	137018	(Tous)	Never	Other	Not-in	White	Female	0		
8965	8964	90	?	77053	(10 premiers...)	Widow?		Not-in	White	Female	0	4	
11514	11513	90	Private	87285	(Personnalisé...)	Never	Other	Own-c	White	Female	0		
32279	32278	90	Private	313749	2	Widow	Adm-c	Unmar	White	Female	0		
32563					5								
32564					9								
					10								
					13								
					14								

Figure 4.8 : Utilisation du filtre automatique sous Excel

On retourne à l'affichage normal, sans le filtre, en reprenant Données>Filtre>Automatique.

b) Filtre élaboré

Pour utiliser le filtre élaboré, on écrit les critères dans une zone supplémentaire. Dans la figure 4.9, il y a 4 critères définis à l'aide des lignes 1 et 2.

- age = 90
- sex = Female
- education-num < 10
- education-num >4

Puis on choisit Données>Filtrer>Filtre élaboré sous Excel et puis on choisit Données>Filtre>Filtre spécial sous OpenOffice, et on trouve 6 personnes vérifiant simultanément ces 4 critères.

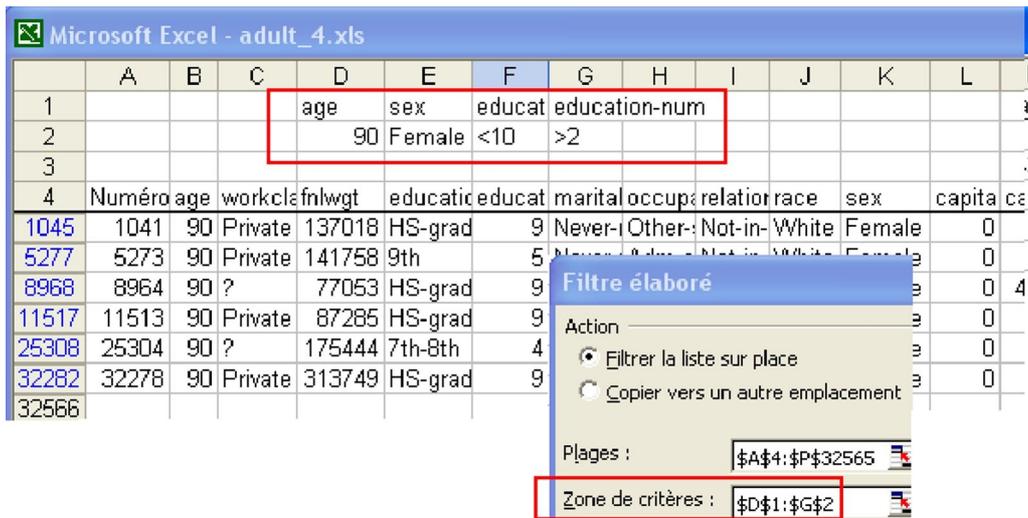


Figure 4.9 : Utilisation du filtre élaboré sous Excel

Il existe également sous Excel comme sous OpenOffice des possibilités de définir des filtres avec des critères plus complexes. Nous renvoyons le lecteur intéressé à l'aide des tableurs.

C. Les sous-totaux et tableaux croisés dynamiques

Alors que dans la partie précédente nous ne faisons que modifier l'affichage des données par des tris, ou en extraire une partie, dans cette partie nous allons créer de nouvelles informations de type statistique (comptages, min, max, moyennes, écart-types) à partir du tableau de données en faisant des regroupements selon certaines caractéristiques prises séparément, ou « croisées ». Dans le premier cas, ce sont les sous-totaux, et dans le second, les tableaux croisés dynamiques.

1. Les sous-totaux

Le tableau de données peut être regroupé par valeurs de certaines caractéristiques, par exemple l'âge, le sexe et le nombre d'années d'études. Selon les tableurs, on doit ou non le trier avant de créer des groupes de personnes selon ces caractéristiques. On calcule certains éléments par groupe de personnes, comme leur nombre, la somme de leurs revenus, la moyenne de leurs âges, etc. Ce sont ces nombres qui sont les « sous-totaux »

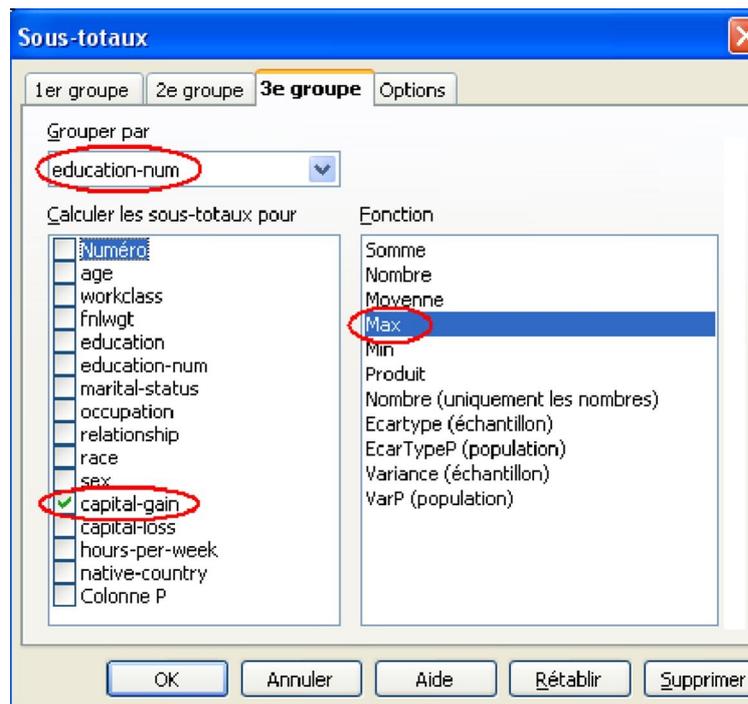


Figure 4.10 : Paramétrage des sous-totaux sous OpenOffice

adult - OpenOffice.org Calc												
	A	B	C	D	E	F	G	H	I	J	K	L
1	Numéro	age	workclass	fnlwgt	education	educ	marital	occupa	relation	race	sex	capita
2	13044	17	?	275778	9th	5	Never-m	?	Own-ch	White	Female	0
3	20431	17	Private	53367	9th	5	Never-m	Other-s	Own-ch	White	Female	0
4	20926	17	Private	60562	9th	5	Never-m	Other-s	Own-ch	White	Female	0
5	26899	17	Private	220562	9th	5	Never-m	Sales	Other-re	Other	Female	0
6	25503	17	Private	317702	9th	5	Never-m	Other-s	Own-ch	Black	Female	0
7	23290	17	Private	98675	9th	5	Never-m	Other-s	Unmarr	White	Female	0
8	14852	17	Self-em	413557	9th	5	Never-m	Sales	Own-ch	White	Female	0
9	14194	17	Private	166290	9th	5	Never-m	Other-s	Own-ch	White	Female	0
10	14545	17	Private	73145	9th	5	Never-m	Craft-re	Own-ch	White	Female	0
11						5 Max						0
12	19135	17	?	280670	10th	6	Never-m	?	Own-ch	White	Female	0
13	28946	17	Private	147069	10th	6	Never-m	Sales	Own-ch	White	Female	0

Figure 4.11 : Utilisation du filtre élaboré sous OpenOffice, tous les niveaux sont dépliés

Dans la figure 4.11, on voit en ligne 11 un « sous-total » qui s'est ajouté aux données, il correspond au max de capital-gain (indiqué par « 0 » en colonne L) des neuf personnes juste au dessus ayant fait 5 années d'études (indiqué par « 5 Max » en colonne F), comme spécifié dans les paramètres de la figure 4.10, qui est égal à 0. A gauche de la feuille de calcul, on a des « traits » et des signes « - ». Quand on clique sur un -, il se transforme en + et toutes les lignes jointes par le trait disparaissent, il ne reste plus que le sous-total correspondant. Le premier groupe est formé des personnes de 17 ans, de sexe féminin, qui ont fait 5 ans d'études. Les zones numérotées de 1 à 5 sur le côté correspondent aux niveaux d'imbrications des groupes. Le trait le plus à gauche est pour le groupe total, le deuxième trait définit les groupes par âge, le troisième fait des groupes de sexe à l'intérieur de chaque groupe d'âge, le quatrième fait des groupes de niveau d'études à l'intérieur de ces derniers, et le cinquième correspond aux personnes isolées. Si on clique sur un de ces nombres en haut à gauche, on replie le niveau

correspondant pour toutes les lignes (voir Fig4.12).

	1	A	B	C	D	E	F	G	H	I	J	K	L
	1	Numéro	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain
	11						5 Max						0
	70						6 Max						34095
	167						7 Max						1055
	187						8 Max						0
	191						9 Max						0
	193						10 Max						0
	194											186	0
	196						3 Max						0
	200						4 Max						0

Figure 4.12 : Utilisation du filtre élaboré sous OpenOffice, niveau 4 replié

Suivant les tableurs, on autorise plus ou moins de niveaux d'imbrication, et les paramètres sont plus ou moins détaillés. Le tableau croisé dynamique permet de faire une synthèse plus élaborée des données.

2. Le tableau croisé dynamique

Son nom diffère selon les tableurs. Appelé « tableau croisé dynamique » sous excel, il devient un « Pilote de données » sous OpenOffice. On sélectionne tout le tableau de données, y compris les intitulés, puis on lance cet outil depuis le menu « Données », et on obtient un tableau de résultats.

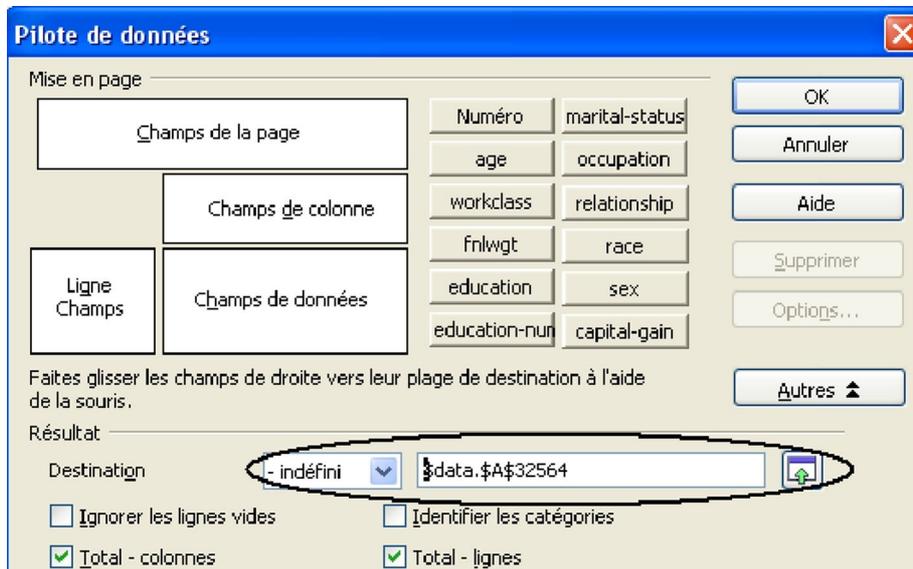


Figure 4.13 : Appel à l'outil de tableau croisé dynamique sous OpenOffice



Attention

Sous Excel, le choix de l'emplacement du tableau créé est directement proposé dans la fenêtre de l'assistant. Il est situé par défaut dans une nouvelle feuille, ce qui permet d'enchaîner la création de plusieurs tableaux croisés dynamiques. Sous les versions 2.4 et précédentes d'OpenOffice, si on désire en faire plusieurs, il faut penser à aller dans les options de la fenêtre pour rendre le choix accessible (voir

Fig. 4.13), sinon chaque nouveau tableau écrase le précédent, son emplacement par défaut étant situé quelques lignes sous le tableau de données.

Si on désire utiliser seulement certaines caractéristiques des données, on peut se contenter de sélectionner la partie correspondante des données, à condition que les colonnes soient consécutives. Avant de lancer cet outil, il faut vérifier que toutes les colonnes concernées ont des intitulés sur la première ligne sélectionnée. L'assistant permet alors de disposer les intitulés dans les 4 champs du tableau. Il ne reste plus qu'à choisir les intitulés adaptés aux informations qu'on veut faire apparaître. Nous proposons maintenant de découvrir concrètement cet outil à travers un exercice TD.

Remarque



Mise à jour

Le tableau croisé dynamique est un outil efficace, mais il mobilise des ressources informatiques importantes, tant en capacité de calcul, qu'en mémoire et en espace disque. Pour l'économiser, la mise à jour en cas de changement dans le tableau de données, n'est pas automatique, elle se fait seulement quand l'utilisateur la demande par la commande « actualiser », accessible par un clic droit si elle n'apparaît pas dans la fenêtre. Et plusieurs tableaux croisés dynamiques sur les mêmes données (obtenus ou non par copier-coller) peuvent être liés, toute modification de l'un (comme un groupement de valeurs) se faisant également sur les autres.

Du dynamique au statique

Le tableau croisé dynamique n'est pas un tableau ordinaire, formé de valeurs indépendantes écrites dans des cellules. Cela a pour conséquence que, par exemple, la modification des intitulés, ou l'insertion de certains graphiques, ne sont pas possibles dans certaines versions du tableur. Pour remédier à cela, il faut le rendre statique en faisant un « copier-collage spécial>Valeurs ». Mais il n'a alors plus aucun lien avec les données, et on ne peut plus utiliser ses fonctionnalités, comme la mise à jour, l'ajout d'intitulés dans les champs, etc.

3. Exercice TD 2

1) Création du tableau

Ouvrir le fichier adult.xls, écrire « Salaire » en P1, sélectionner les données puis lancer l'outil comme indiqué sur la figure 4.14 pour Excel, et valider toutes les étapes de l'assistant d'Excel avec les valeurs par défaut.

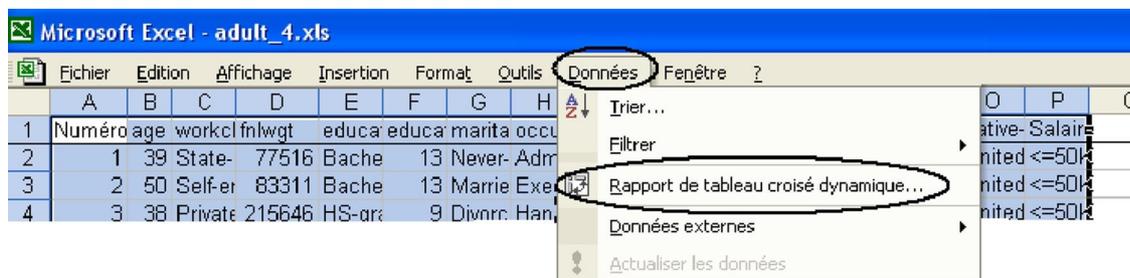


Figure 4.14 : Appel à l'outil de tableau croisé dynamique sous Excel

Une nouvelle feuille a été ajoutée qui contient une maquette de tableau à compléter. Dans la figure 4.15, on a entouré en rouge les menus : en haut le menu général du tableau croisé dynamique, en bas la liste des caractéristiques des

données sélectionnées. Pour déplacer une caractéristique vers une des 4 zones du tableau entourées de noir, on la sélectionne par un clic gauche, puis on déplace la souris, clic gauche enfoncé, vers la zone choisie.

Aide :

Même si l'étiquette ne suit pas le parcours de la souris, il faut continuer le mouvement, puis lâcher le clic une fois arrivé dans la bonne zone. En cas d'erreur, il faut faire le mouvement dans l'autre sens : partir de l'étiquette mise dans la zone inappropriée, et la déplacer vers la liste d'étiquettes.

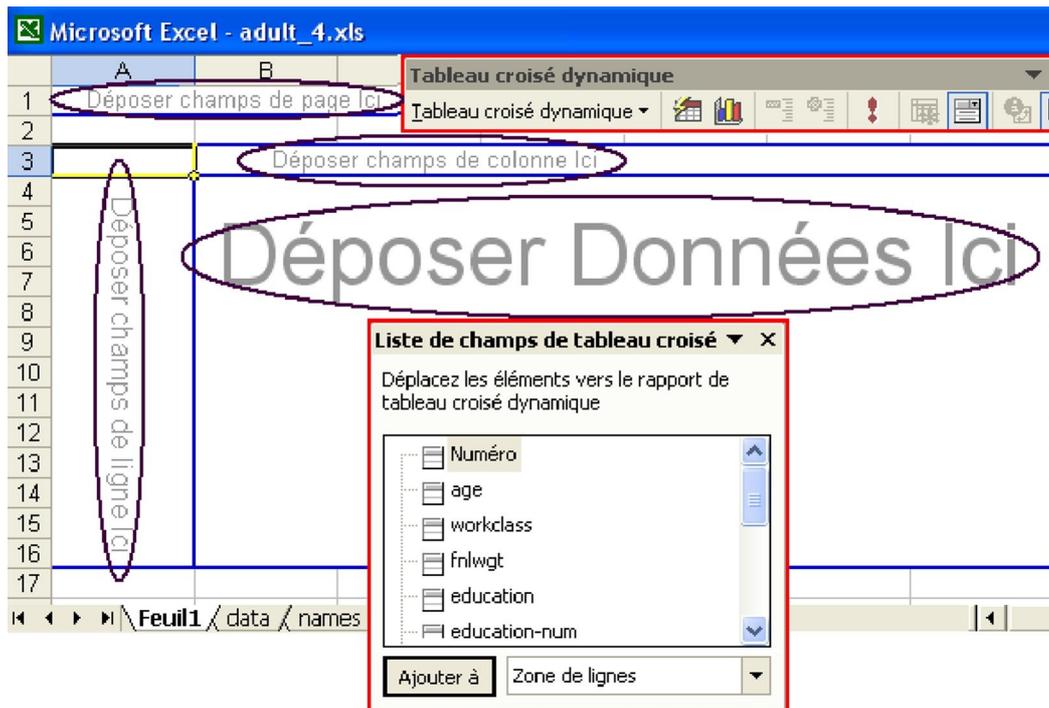


Figure 4.15 : Appel à l'outil de tableau croisé dynamique sous Excel

Il y a 2 types de zones, 1) les champs de page, ligne, ou colonne, où on souhaite voir apparaître la caractéristique selon laquelle les données seront regroupées, et 2) la zone de données qui contient le genre de calcul souhaité. Dans cette dernière, une fois la caractéristique placée, l'utilisateur peut choisir s'il désire faire une somme, proposée par défaut si les données sont numériques, un comptage, comme proposé par défaut si les données sont textuelles, ou une moyenne, etc. le type de calcul étant le même que pour l'outil « sous-totaux » exposé dans la partie précédente (voir Fig. 4.16). On peut aussi intervenir sur les zones du premier type, comme on le voit dans la figure 4.17 pour OpenOffice, si on désire avoir les sous-totaux d'un champ de ligne, qui sont fournis par défaut sous Excel.



Figure 4.16 : Des options d'affichage sous OpenOffice. Ici pour le champ de données entouré de bleu.

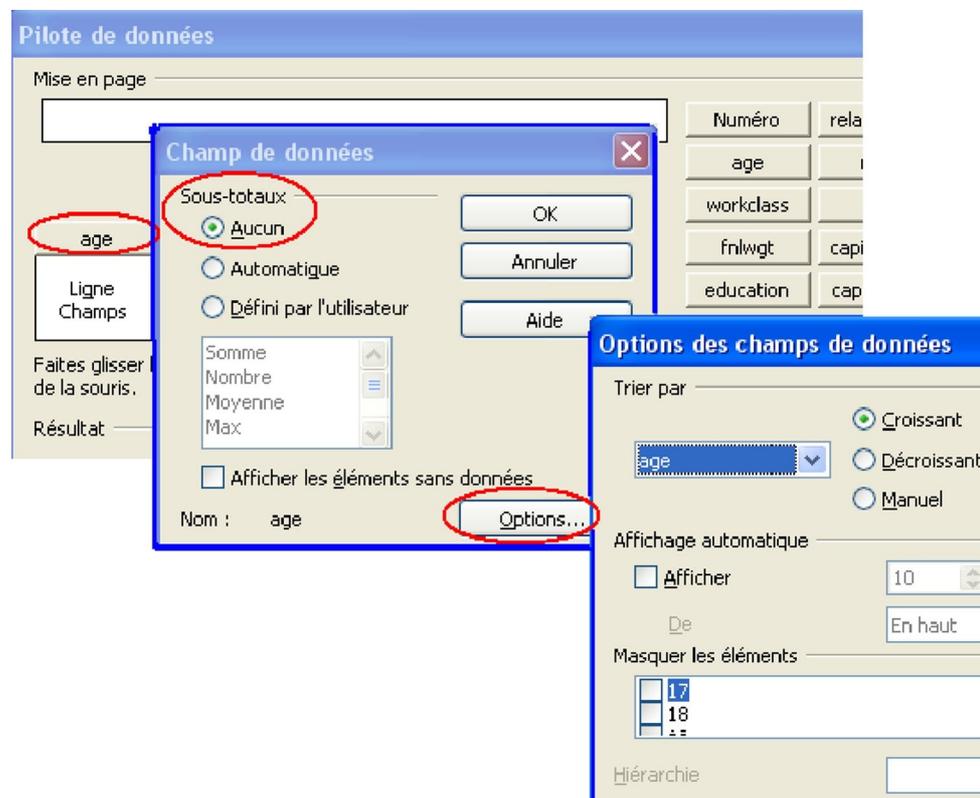


Figure 4.17 : De nombreuses options d'affichage sous OpenOffice. Ici pour le champ de ligne « age »

Remarque

Dans cette leçon, notre objectif est de faire découvrir à l'utilisateur l'existence des fonctionnalités les plus utiles de cet outil, sachant que selon les divers tableurs, et mêmes les diverses versions d'un même tableur, ces fonctionnalités peuvent se situer à des endroits très différents. Connaissant leur existence et leur rôle, l'utilisateur qui ne les aura pas repérées dans les fenêtres de l'assistant saura les retrouver en parcourant le dédale des fenêtres qui apparaissent suite à un clic droit sur les zones correspondantes du tableau, ou en interrogeant l'aide.

2) Utilisation du tableau

Pour utiliser le tableau obtenu, aller sur la feuille de calcul où il apparaît. S'il n'est pas dans une nouvelle feuille, cherchez-le en bas, en dessous des données, ou en haut, à droite des données.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Filter																		
2	occupation	Sales																	
3																			
4	Max - capital-gain		education-num																
5	age	sex	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total	Résultat
6	17	Female			0	0	1055	0	0										1055
7		Male			0	0	0	0	0										0
8	18	Female			0	0	0	1055	2176										2176
9		Male			0	0	0	594	0	2176									2176
10	19	Female				0	0	0	0	0									0
11		Male					0	1055	4101	0									4101

Figure 4.18 : Appel à l'outil de tableau croisé dynamique sous OpenOffice

Dans la zone de page, on avait placé la caractéristique « occupation », qui se retrouve en haut du tableau (dans la cellule A2 de la Figure 4.18). Dans la liste déroulante qui est à côté, choisissez « Sales ». Ainsi, on n'a qu'une partie des données, celles relatives aux personnes qui travaillent dans le secteur des ventes.

On a entouré en bleu dans cette figure les contenus de 4 cellules :

- en E6 : les personnes
 - travaillant dans le secteur des ventes
 - âgées de 17 ans
 - de sexe féminin
 - ayant fait 4 années d'études

La cellule est vide, aucune personne figurant dans les données n'est dans ce cas.

- en F8 : les personnes
 - travaillant dans le secteur des ventes
 - âgées de 18 ans
 - de sexe féminin
 - ayant fait 5 années d'études

La cellule F8 contient le nombre 0, qui est le max du capital-gain pour ces personnes. En faisant un double-clic sur cette cellule, on fait apparaître toutes les personnes correspondantes dans une nouvelle feuille (Voir Fig. 4.19). Elles sont au nombre de deux (numéros 8517 et 22671). On peut constater qu'elles vérifient bien les 4 conditions pointées ci-dessus. Et le capital-gain de chacune étant 0, leur max est donc bien 0.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Numéro	age	work	fnl	wgt	educ	educa	marital	occup	relatio	race	sex	capital-gain
2	8517	18	Privat	135315	9th	5	Never-r	Sales	Own-o	Other	Female	0	
3	22671	18	Privat	193166	9th	5	Never-r	Sales	Own-o	White	Female	0	

Figure 4.18

- en I8 : les personnes
 - travaillant dans le secteur des ventes
 - âgées de 18 ans
 - de sexe masculin
 - ayant fait 8 années d'études

La cellule I8 contient le nombre 594, qui est le max du capital-gain pour ces personnes. En faisant un double-clic sur cette cellule, on fait apparaître toutes les personnes correspondantes dans une nouvelle feuille. Elles sont également au nombre de deux (numéros 4470 et 26882). Le capital-gain de la première est 0 alors que celui de la seconde est 594, leur max est donc bien 594.

- en R11 : les personnes
 - travaillant dans le secteur des ventes
 - âgées de 19 ans
 - de sexe masculin

Leur niveau d'études n'est pas spécifié, car cette cellule fait partie de la colonne de totaux. On peut remarquer que le niveau d'études de ces personnes sera de 7, 8, 9 ou 10, et le maximum de leurs capital-gains sera de 4101. En double-cliquant sur cette cellule, on obtient les données des 37 personnes concernées. En parcourant les lignes correspondant aux 37 personnes, on peut constater que toutes sauf 2 ont un capital-gain de 0. C'est la personne numéro 23494 qui atteint le maximum de capital-gain de 4101.

Conseil :

Jouer sur la mise en forme du tableau croisé dynamique pour rendre l'information plus lisible.

1) Réduire des catégories à une seule ligne. Pour cela il suffit de double-cliquer sur une cellule contenant la catégorie. Par exemple un double-clic sur la cellule A6 de la figure 4.20 fera disparaître le détail (Male/Female) pour la catégorie « 16 » de la caractéristique « age ». C'est ce qui a été fait pour les lignes 12 (age=40) et 13 (age=19) de cette figure, ainsi que pour les colonnes F à I (« education-num » de 5, 6, 7 et 8)

	A	B	C	D	E	F	G	H	I	J	K	L	M
2													
3	Max de capital-gain		education-num	4	Total	4	5	6	7	8	9	Total	9
4													
5	age	sex	Priv-house-serv	Sales									Priv-house-serv
6		17	Female										
7			Male										
8	Total	17											
9			24	Female									
10				Male									
11	Total	24											
12			40										
13			90										
14	Total												
15													

Figure 4.20 : Utilisation des paramètres du tableau croisé dynamique sous Excel

Voir définition du "glisser/déposer"*

2) Ne faire apparaître que les catégories intéressantes. Il faut faire apparaître ces informations puis cocher sur celles qu'on désire garder (voir Fig. 4.20 pour Excel) ou on contraire retirer (voir Fig. 4.17 pour OpenOffice).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	occupat	Sales																
2																		
3	Max de		educés															
4	age	sex		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	1-20	Female		0	0	1055	0	1055	0	1055	2176	0	0	0	0	0	0	2176
6		Male		0	0	0	0	0	0	1055	4101	2176						4101
7	Total	1-20		0	0	0	0	1055	1055	4101	2176	0	0	0	0	0	0	4101
8	21-40	Female		0	0	0	0	0	0	7688	7298	594	7688	15024	0	13650	15024	
9		Male		0	0	4064	4508	0	0	0	0	0	0	0	0	0	0	99999
10	Total	21-40		0	0	4064	4508	0	0	0	0	0	0	0	0	0	0	99999
11	41-60	Female		0	0	0	0	0	0	0	0	0	0	0	0	0	0	27828
12		Male		0	0	0	0	0	7688	0	0	0	0	0	0	0	0	99999
13	Total	41-60		0	0	0	0	0	7688	0	0	0	0	0	0	0	0	99999
14	61-80	Female		0	0	0	0	0	0	0	0	0	0	0	0	0	0	25124
15		Male		0	0	0	0	0	0	0	0	0	0	0	0	0	0	99999
16	Total	61-80		0	0	0	0	0	0	0	0	0	0	0	0	0	0	99999
17	81-100	Female		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18		Male		0	0	0	0	0	0	0	0	0	0	0	0	0	0	9386
19	Total	81-100		0	0	0	0	0	0	0	0	0	0	0	0	0	0	9386
20	Total			0	0	4064	4508	7688	7430	18481	99999	99999	27828	15024	99999	99999	15024	99999

Figure 4.21 : Regroupement des âges du tableau croisé dynamique sous Excel

3) Regrouper les catégories numériques si le tableau le permet. Pour cela, on sélectionne une catégorie du tableau (cellule A5 de la figure 4.18), et on va dans Données>Plan>Grouper. On peut voir l'effet de cette commande dans la figure 4.21 : On a maintenant 5 tranches d'âge au lieu d'autant de catégories qu'il y a d'âges différents dans les données.

4) Agir sur le graphique plutôt que sur le tableau. Quand le tableau croisé dynamique est actif (sélectionner une cellule de ce tableau, faire un clic droit et l'activer s'il ne l'est pas), on peut représenter automatiquement les données par un graphique, plus ou moins dynamique. Celui figurant sur la figure 4.22 est très dynamique : on y retrouve dans les encadrés rouges les menus de la figure 4.15, et

entourées en bleu les 5 zones de la figure 4.18. On peut déplacer les éléments comme on l'a fait dans la figure 4.15, et le graphique ainsi que le tableau croisé dynamique sur lequel il est construit se modifient en conséquence. Une remarque toutefois : le choix par défaut consistant à faire le cumul des valeurs n'est pas approprié à un maximum, mais à une somme, ou un comptage.

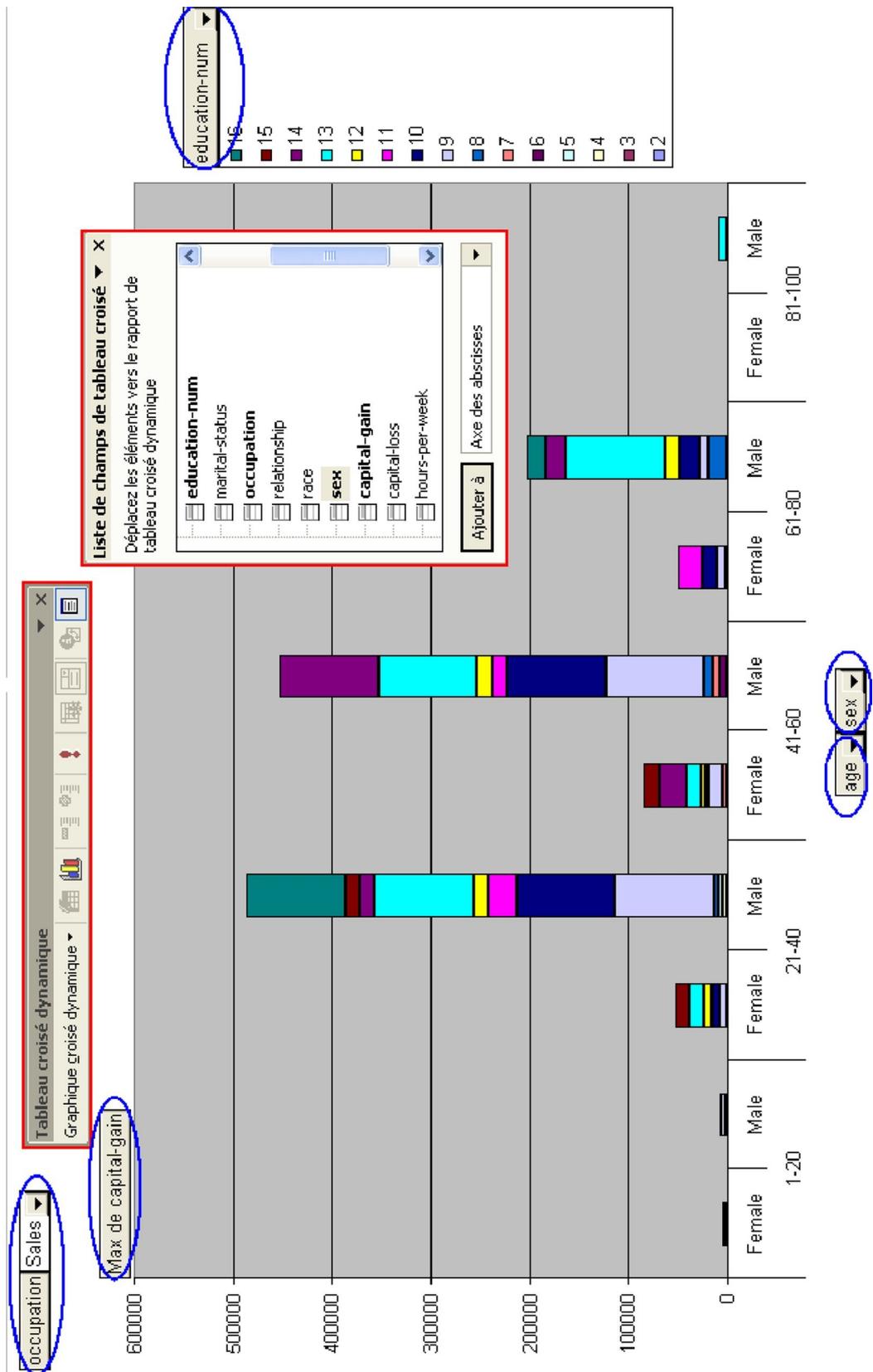


Figure 4.22 : Graphique du tableau croisé dynamique sous Excel

4. Exercices

Question 1

[Solution n°66 p 165]

1) Réaliser le tableau croisé dynamique tel qu'il apparaît sur la figure 4.20

Question 2

2) Réaliser le graphique tel qu'il apparaît sur la figure 4.22, mais avec des valeurs non cumulées.

Remarque : le choix par défaut consistant à faire le cumul des valeurs n'est pas approprié à un maximum, mais à une somme, ou un comptage.

Question 3

3) Disposer l'intitulé du champ page du tableau croisé dynamique de la figure 4.18 sur le champ ligne, à la suite de l'âge et le sexe, et remplacer dans le calcul la fonction « max » par la fonction « nombre ». Regrouper les âges par tranches de 10 ans et les années d'études en deux tranches de votre choix couvrant l'ensemble des années d'études.

Question 4

4) Ajouter le champ « Salaire », et en utilisant les autres champs de vos choix, trouver une règle du type : Si caract1=valeur1 et caract2=val2 et Alors Salaire = »>50K

Et une règle donnant le résultat contraire.

Question 5

5) faire un exercice avec plusieurs valeurs dans le champ de données

D. La recherche de valeurs cibles, le solveur

Dans cette partie nous ne partons plus des données pour en tirer des informations, en déduire des règles, mais nous partons de relations, équations, inéquations, et nous en cherchons des solutions. Les méthodes de résolution implantées dans le tableur sont « numériques » et non « formelles ». Pour résoudre $3x+5=2$, le tableur n'utilise pas la formule $x=(2-5)/3$, mais une méthode de type algorithmique, consistant à essayer un certain nombre de valeurs de x jusqu'à ce que la valeur de $3x+5$ soit assez proche de 2.

1. La recherche de valeurs cibles

Dans la figure 4.23, on montre comment résoudre cette équation simple $3a+5=2$ en utilisant l'outil « recherche de valeur cible ». Il convient de faire les opérations suivantes

- écrire dans une cellule une valeur quelconque de la variable a (on a mis « 1 » dans B4)
- écrire dans une cellule la formule en fonction de a (on a mis « $3*B4+5$ » dans C4)

- sélectionner la cellule de la formule (C4)
- lancer l'outil « recherche de valeur cible ». La cellule de formule est pré-remplie (par C4)
- remplir la valeur cible et la cellule variable puis valider (valeurs saisies entourées en bleu)

La solution (encadrée en rouge) trouvée est bien égale à « -1 ». Si on accepte que la valeur soit insérée, le contenu de la cellule B4 deviendra 1, et le contenu de la cellule C4 deviendra 2.

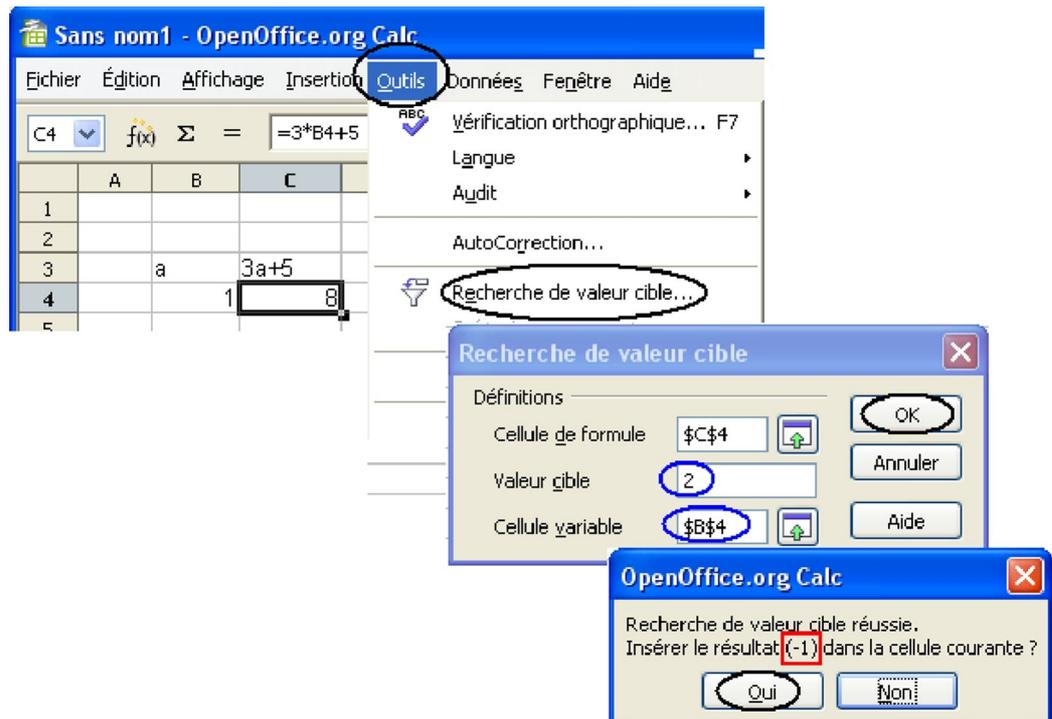


Figure 4.23 : Recherche de la solution approchée de $3a+5=2$ par tableur



Remarque

Les valeurs écrites dans les cellules B3 et C3 ne sont pas utilisées par l'outil. Elles servent seulement comme intitulés pour rappeler le contenu des cellules du dessous.

2. Exercice TD 3

1) Résoudre de façon approchée l'équation $x^3-5x^2+4x+1=0$

On doit déjà avoir une idée du nombre de solutions avant d'en rechercher des valeurs approchées. On sait que c'est le nombre de fois que la courbe d'équation $y = x^3-5x^2+4x+1$ rencontre l'axe des x. Pour dessiner cette courbe, écrire dans la colonne A à partir de la ligne 3 une suite de nombres de valeur initiale -1 et de raison 0,1. Puis dans B3 la formule « $=A3^3-5*A3^2+4*A3+1$ » ou « $=A3*A3*A3-5*A3*A3+4*A3+1$ ». Et « tirer » vers le bas afin de la recopier tant qu'il y a des valeurs dans la colonne A. Sélectionner ce bloc de valeurs x et y et insérer un graphique (nuage de points pour Excel, diagrammeXY pour OpenOffice).

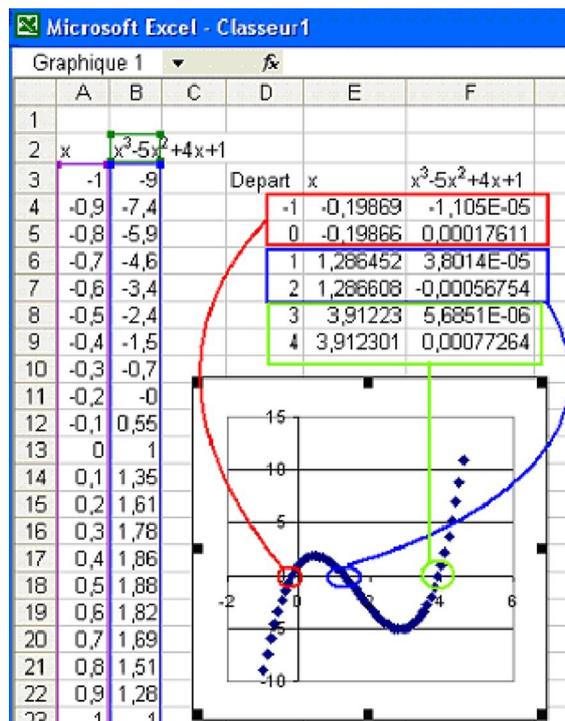


Figure 4.24 : Recherche de valeurs approchées des 3 racines du polynôme $P(x) = x^3 - 5x^2 + 4x + 1$

On a entouré les 3 points où la courbe rencontre l'axe des x : en rouge pour x entre -1 et 0, en bleu pour x entre 1 et 2, en vert pour x entre 3 et 4.

2) On écrit en E4 la valeur -1, et on fait un copier-coller de B3 vers F4. Puis on sélectionne F4 et on appelle l'outil « valeur cible ». La solution approchée apparaît dans la cellule E4 et remplace -1. Elle produit une valeur de F4 presque nulle. Comme -1 a disparu, on l'écrit dans la cellule D4 pour garder trace de la valeur dont est partie le tableur pour la recherche de solution.

3) On tape 0, 1, etc. sous la cellule D4, on les recopie sous la cellule E4, et on fait une copie de la formule de F4 dans les cellules du dessous. Puis on appelle l'outil « valeur cible » sur ces 5 cellules. On peut observer qu'on trouve à chaque fois une valeur approchée de la solution la plus proche de la valeur de départ.

3. Exercice 3

Question

[Solution n°67 p 165]

On place une somme de 500 euros pendant 10 ans à un taux annuel de 2,30%. Combien obtient-on au bout de ces 10 ans ? Quelle somme initiale permettrait d'obtenir la somme de 700 euros dans les mêmes conditions ? Même question en faisant varier l'année puis en faisant varier le taux.

4. Le solveur

Il figure dans le menu « Outils » de la plupart des versions actuelles du tableur. Toutefois, s'il ne figure pas dans ce menu sous Excel, il est possible de le faire apparaître en allant dans « Outils > Macro complémentaire », et en cochant la case du solveur (Voir Fig. 4.27), puis en validant l'installation de cet outil. Le solveur est

alors ajouté dans le menu « Outils ».

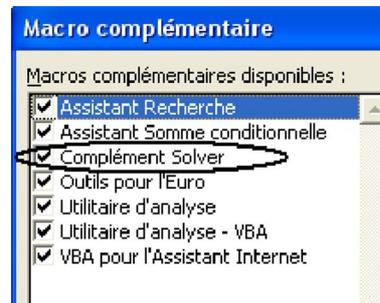


Figure 4.27 : Les macros complémentaires à installer sous Excel 2003 pour disposer de tous les outils.

Étant donnée une fonction d'une ou plusieurs variables, le solveur a pour but de trouver des valeurs de ces variables qui rendent la fonction maximale, minimale, ou égale à une valeur donnée, dans un domaine particulier des variables, défini par des contraintes (inéquations, équations, valeurs entières). Comme pour la valeur cible, le solveur ne peut fournir qu'une solution, la plus proche des valeurs des variables utilisées pour écrire le problème.

Un problème simple pour débiter

Reprenons avec le solveur le problème traité dans la figure 4.25 avec l'outil « Valeur cible ». Pour cela on sélectionne la cellule H27, puis on appelle le solveur par « Outils>Solver »,

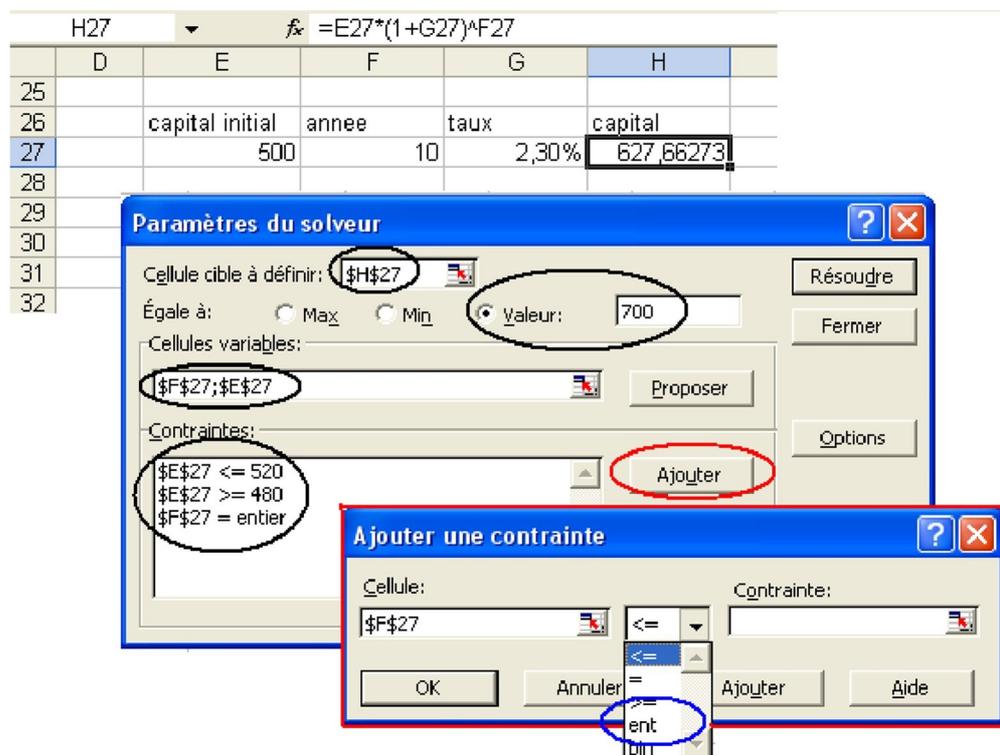


Figure 4.28 : L'utilisation du solveur sous Excel : choix des paramètres

On a indiqué dans la figure 4.27 les paramètres du solveur à fixer. Pour cela, on se place dans la zone de saisie et on clique sur la cellule nécessaire et son adresse est recopiée dans la ligne de saisie. Tout en haut, l'adresse de la cellule cible est proposée, car on l'avait sélectionnée. On doit alors cocher la case « valeur » et

taper le nombre 700 à côté de cette case. On peut maintenant choisir plus d'une cellule variable. Celle qui nous intéresse est la cellule F27 qui contient le nombre d'années. Une fois cette adresse écrite, on tape un « ; » et on clique sur la cellule E27 contenant le capital. En effet, bien qu'on ne désire changer que le nombre d'années, comme on souhaite qu'il soit entier, on autorise une petite modification du capital initial, sans laquelle le solveur signale qu'il ne peut pas trouver de solution. On ajoute alors une à une les contraintes. Le capital est autorisé à varier entre 480 et 520, et le nombre d'années doit être entier.

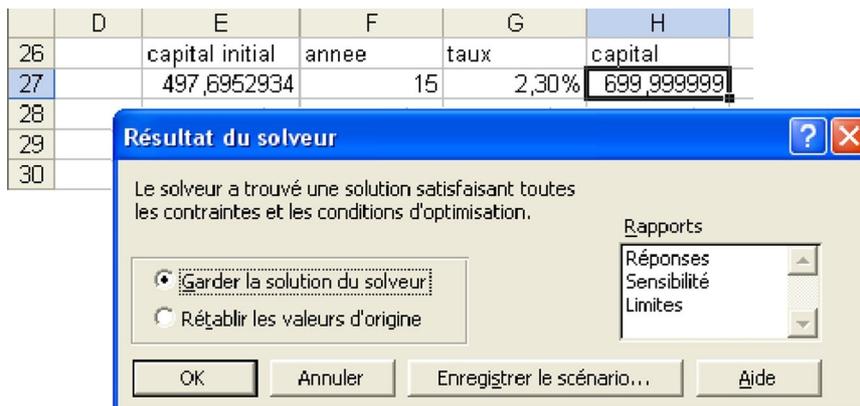


Figure 4.29 : L'utilisation du solveur sous Excel : choix des paramètres

On enfonce alors le bouton « résoudre ». Une solution est proposée (voir Fig. 4.29), et si on choisit de la garder on retrouve bien les 15 ans d'intérêts pour que le capital obtenu soit de 700.

Un problème d'optimisation simple : la programmation linéaire

Les problèmes d'optimisation que peut résoudre le solveur sont nombreux. Pour pouvoir les exprimer de façon adaptée à cet outil, il faut les mettre sous la forme d'une fonction « objectif » de plusieurs variables, dont on recherche le maximum (ou le minimum), et surtout la valeur de chacune des variables permettant de l'atteindre, les variables étant soumises à des contraintes (équations, inéquations ou valeurs entières).

Le problème le plus simple d'optimisation appartient au domaine de la programmation linéaire :

- la fonction objectif est une combinaison linéaire de ses variables
- les variables appartiennent à un domaine de valeurs défini de façon linéaire.

C'est sur un exercice TD que nous allons voir le fonctionnement du solveur.



Exemple : Exemple corrigé sur la maximisation du profit (Question N°4 du problème N°4)

5. Exercice TD 3.2

Une entreprise qui vient de se créer désire s'équiper en ordinateurs en minimisant le coût de revient de son parc informatique sur 5 ans. Pour cela elle a évalué les coûts liés à 4 types de machines, que nous appellerons A, B, C et D, en répartissant ces coûts en trois types : coût d'achat, coût d'entretien (réparation, voire remplacement), coût de fonctionnement (électricité, personnel qualifié, logiciels adaptés) car elle a des limites à ne pas dépasser pour les trois types de dépenses. Voici un tableau résumant ces informations.

Coûts unitaires sur 5 ans			
Type	Achat	Entretien	Fonctionnement
A	600,00 €	1 900,00 €	1 000,00 €
B	1 000,00 €	1 350,00 €	1 350,00 €
C	1 150,00 €	1 200,00 €	1 450,00 €
D	1 800,00 €	1 100,00 €	1 100,00 €

Sommes à ne pas dépasser			
	18 000,00 €	25 000,00 €	21 000,00 €

Tableau 13 : Coûts unitaires sur 5 ans et sommes à ne pas dépasser

Aux contraintes financières s'ajoute une contrainte sur le nombre d'ordinateurs : elle souhaite que leur nombre soit compris entre 17 et 20. Le problème se résout en plusieurs étapes

Question 1

[Solution n°68 p 166]

- 1) Écriture des informations dans le tableur

Question 2

[Solution n°69 p 166]

- 2) Écriture des formules permettant d'utiliser le solveur

Question 3

[Solution n°70 p 167]

- 3) Utilisation du solveur

Question 4

[Solution n°71 p 168]

- 4) Éventuelles modifications des contraintes pour d'autres essais avec le solveur

E. Exercices appliqués à l'économie avec correction

1. Problème 11 : Maximisation du profit

Une entreprise produit des bouteilles de boisson rafraîchissante d'un seul modèle. Elle a connaissance de la fonction de demande des consommateurs, qui dépend de la température extérieure et du comportement de son principal concurrent.

La fonction de demande quotidienne est donnée par la fonction :

$$D = 5000 * t^{0,5} - p^{0,6}$$

Dans laquelle t est égal à la température extérieure (en ° C) et p est le prix de la bouteille (en €), fixé à 2 €.

Question 1

[Solution n°72 p 169]

- 1) Calculez la quantité demandée lorsque la température extérieure est de 20° C, puis pour 30 ° C.

Question 2

[Solution n°73 p 169]

2) Quel est le profit de l'entreprise dans les deux situations précédentes si le prix de revient est de 1,50 € ?

Question 3

[Solution n°74 p 169]

3) L'entreprise envisage de réaliser une promotion en proposant des lots de trois bouteilles à 5 €, elle espère ainsi doubler ses ventes. Intuitivement, qu'en pensez-vous ?

Question 4

[Solution n°75 p 170]

4) Calculez le nouveau bénéfice.

Question 5

[Solution n°76 p 170]

5) Compte tenu de cette situation, l'entreprise fixe un nouveau tarif des lots à 5,50 €, elle espère toujours doubler ses ventes. qu'en pensez-vous ?

Question 6

[Solution n°77 p 170]

6) Devant cette situation, le principal concurrent, qui voit ses ventes baisser, décide aussi de faire des lots pour retrouver ses parts de marché. L'augmentation des ventes est alors ramenée à 1,2 fois les ventes antérieures ; Qu'en pensez-vous ?

Question 7

[Solution n°78 p 170]

7) L'entreprise décide de revenir aux tarifs de départ, mais le concurrent poursuit ses promotions, ce qui aboutit à une baisse des ventes (moitié de la dernière situation). Calculez les nouveaux profits pour les températures de 20°, 25° et 30°.

2. Problème 12 : Données chronologiques

Chiffre d'affaires d'un hypermarché en 100 000 €

	Données : chiffre d'affaires			
CA	2006	2007	2008	2009
janvier	22	22	23	23
février	10	10	11	10
mars	11	11	12	12
avril	12	12	13	13
mai	17	19	18	21
juin	20	20	21	21
juillet	17	17	18	20
août	10	10	11	12
septembre	21	20	21	22
octobre	17	17	19	19
novembre	15	15	16	16
décembre	28	30	30	33

Tableau 14 : Chiffre d'affaires d'un hypermarché

Nombre de jours ouvrables :

Jours ouvrables	2006	2007	2008	2009
janvier	22	21	22	21
février	19	20	19	20
mars	25	23	25	23
avril	23	23	23	23
mai	17	20	17	20
juin	20	23	20	23
juillet	22	21	22	21
août	21	22	21	22
septembre	23	23	23	23
octobre	23	22	23	24
novembre	24	22	25	22
décembre	23	25	23	24

Tableau 15 : Nombre de jours ouvrables

Dans le traitement de données chronologiques, le temps est une variable discrète qui prend des valeurs de 1 à n.

Question 1

[Solution n°79 p 170]

1) En janvier 2007 le responsable du magasin constate une stabilisation, par rapport à 2006, en mai il perçoit une augmentation, en novembre une baisse et en décembre une augmentation, pourtant il a une impression contraire. Pourquoi ?

Question 2

[Solution n°80 p 170]

2) Calculez le chiffre d'affaires par année, qu'en pensez-vous ?

Question 3

[Solution n°81 p 170]

3) Calculez les chiffres d'affaires des mêmes mois qu'à la question 1 par jour ouvrable

Question 4

[Solution n°82 p 170]

4) Le chef d'entreprise décide de se faire une idée en calculant les chiffres d'affaires après correction des jours ouvrables (cette correction est opérée en calculant le chiffre d'affaires par jour ouvrable, multiplié par le nombre moyen de jours ouvrables)

3. Problème 13 : Enquête premier salaire

Suite de l'exercice 4 de la leçon 3.

On se propose d'étudier le premier salaire de jeunes sortis, depuis 18 mois, d'un BTS en apprentissages, à partir d'un extrait d'une enquête (voir enquête premier salaire feuille 2).

Question 1

[Solution n°83 p 170]

1) A l'aide d'un tableau croisé dynamique, déterminez la répartition des salaires des salariés à temps plein, selon le sexe (réponses feuille 1). Déterminez le salaire moyen des hommes, des femmes et total.

Question 2

[Solution n°84 p 170]

2) Mêmes questions pour les salariés à temps partiel supérieur ou égal à un mi temps.

Leçon 5 : Les diverses représentations graphiques des données

Introduction	125
Diagrammes de statistiques descriptives	126
Une modélisation des données « à la main » utilisant le côté dynamique des graphiques	137
Exercice TD : Ajustement graphique de lois de probabilités à des données	138
Tendances pour modéliser les données	150
Exercices TD pour la leçon 5	163
Exercices appliqués à l'économie avec correction	170

A. Introduction

Dès que la taille d'un tableau de données devient importante, une représentation par un diagramme/graphique permet d'avoir une vue d'ensemble des données, de visualiser certaines particularités. Nous étudions d'abord les principaux types de diagrammes utilisés en statistiques descriptives, puis nous passons aux outils de modélisation des données à base de graphiques les plus courants : ajustement « à la main » de modèles paramétriques simples aux données, découverte de tendances à l'aide de moyennes mobiles et de régressions linéaires.



Attention

Les diagrammes de ce cours ont été réalisés essentiellement sous OpenOffice 3.1, mais devraient pouvoir être faits sous toutes les versions de tableur, sauf indication contraire de notre part. Toutefois les détails de leur réalisation sont souvent spécifiques à OpenOffice 3.1 et si on utilise une autre version de tableur, il faudra alors procéder différemment pour arriver au même résultat, ou à un résultat proche.

B. Diagrammes de statistiques descriptives

1. Introduction

Si on dispose de données sous la forme d'un tableau de valeurs pour un ensemble d'individus ou d'objets, à raison d'une ligne par individu et d'une colonne par variable ou caractéristique, on peut représenter ces données sous forme de diagrammes. Pour cela, on doit disposer de valeurs numériques, qui peuvent des variables quantitatives, ou être obtenues par comptages (effectifs, pourcentages) à partir de variables qualitatives (M/F pour le sexe, jeune/adulte/vieux pour l'âge, etc.). Dans le premier cas, quand on a peu d'individus, il suffit de dessiner un point par individu, mais quand on en a plus, on regroupe très souvent les lignes selon d'autres variables (sexe, âge, ...) et on calcule alors une valeur numérique par groupe qui peut être la moyenne, médiane, variance, écart-type, etc. des individus du groupe. Nous avons vu dans la leçon 4 comment utiliser les tableaux croisés dynamiques sous Excel et OpenOffice pour créer de tels tableaux de synthèse des données.



Remarque

On peut créer sous Excel des effectifs en utilisant l'outil « histogramme ». On a accès à cet outil et à d'autres bien utiles en statistiques (analyse de variance à un ou deux facteurs, de corrélation, de covariance, test de l'égalité de moyennes pour échantillons indépendants ou appariés, d'égalité de variances, génération de nombres aléatoires suivant diverses loi de probabilités, etc.), en suivant le chemin Outils>Utilitaire d'analyse>Histogramme, chemin présent si on a choisi d'installer la macro complémentaire « utilitaire d'analyse ».

2. Une série de valeurs numériques, répartition en groupes selon une variable qualitative

Quand la valeur numérique est un effectif, on privilégie les diagrammes à barres ou colonnes, les diagrammes circulaires ou « camemberts ». On procède en plusieurs étapes pour les obtenir :

- Sélection de toute la zone contenant les nombres (disposés en une ligne ou une colonne) et les intitulés ou étiquettes (textes présents dans une ligne et une colonne supplémentaire). Pour l'exemple de la figure 5.1, G4 : J5 pour la répartition selon l'âge et L2 : M4 selon le sexe.
- Appel de l'assistant graphique par insertion>Diagramme
- Choix du diagramme approprié (voir figure 5.2), qui se fait lui-même en plusieurs étapes
- Sélection éventuelle d'objets du diagramme pour les modifier. Par exemple :
 - figure 5.3, la légende a été sélectionnée puis supprimée par appui sur la touche « Suppr » du clavier.
 - figure 5.4, l'axe a été sélectionné puis par un clic droit, on a pu accéder à ses propriétés et modifier son échelle pour faire apparaître la valeur initiale 0.
 - figure 5.5, le diagramme a été sélectionné et on a pu ainsi accéder au menu « d'insertion » contextuel au diagramme. Ce menu diffère selon

qu'on sélectionne un diagramme ou une cellule de la feuille de calcul ; c'est pour cela qu'on l'appelle contextuel. Quand on utilise un clic droit sur un objet, c'est également un menu contextuel qui s'affiche, comme on vient de le faire pour l'axe du graphique.

- figure 5.6, des étiquettes ont été ajoutées à la série de données, contenant l'effectif. Puis ces étiquettes ont été sélectionnées et leur format a été modifié pour les rendre plus lisibles.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		Jeune	Adulte	Vieux	Total							Sexe	Effectif
3	Feminin	10	22	30	62							Feminin	62
4	Masculin	15	18	40	73	Age	Jeune	Adulte	Vieux			Masculin	73
5	Total	25	40	70	135	Effectif	25	40	70				

Figure 5.1 : Répartition de 135 personnes selon le sexe et l'âge

The image shows a spreadsheet with a bar chart and the OpenOffice Assistant dialog. The chart displays two bars: 'Feminin' with a value of 62 and 'Masculin' with a value of 73. The Assistant dialog is in the 'Choisissez les paramètres des titres, de la légende et de la grille' step. The title is 'Répartition par sexe', the Y-axis is 'Effectifs', and the legend is checked. The 'Terminer' button is highlighted.

Figure 5.2 : Utilisation de l'assistant d'OpenOffice pour la réalisation d'un diagramme « colonne »

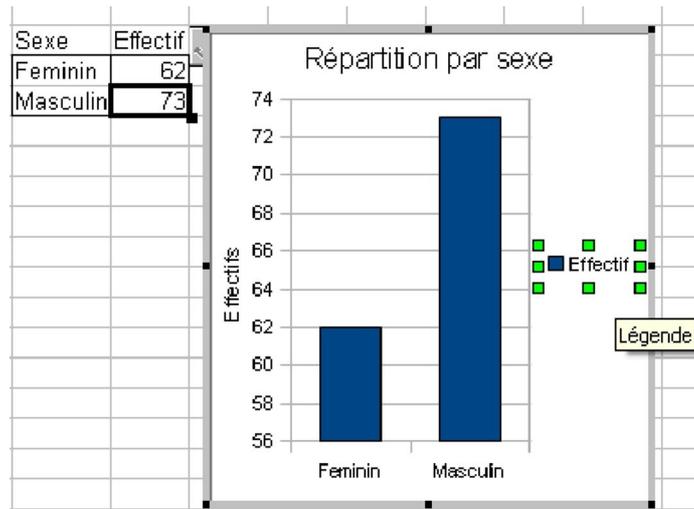


Figure 5.3 : Sélection de la légende du digramme

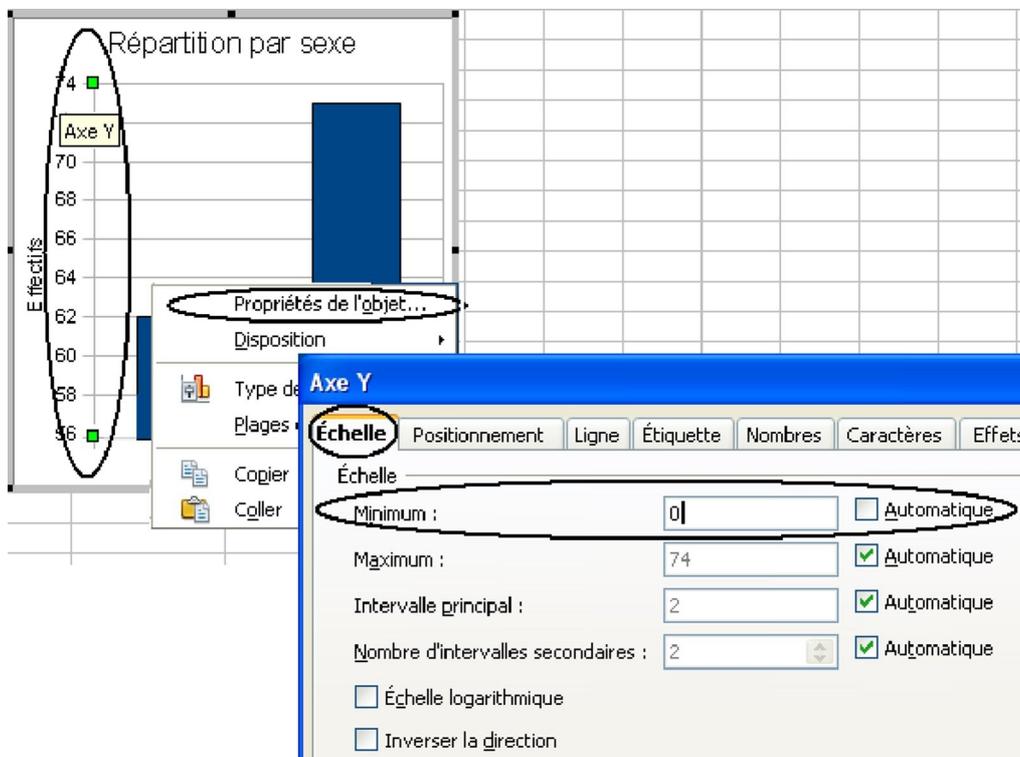


Figure 5.4 : Modification de l'échelle de l'axe des Y

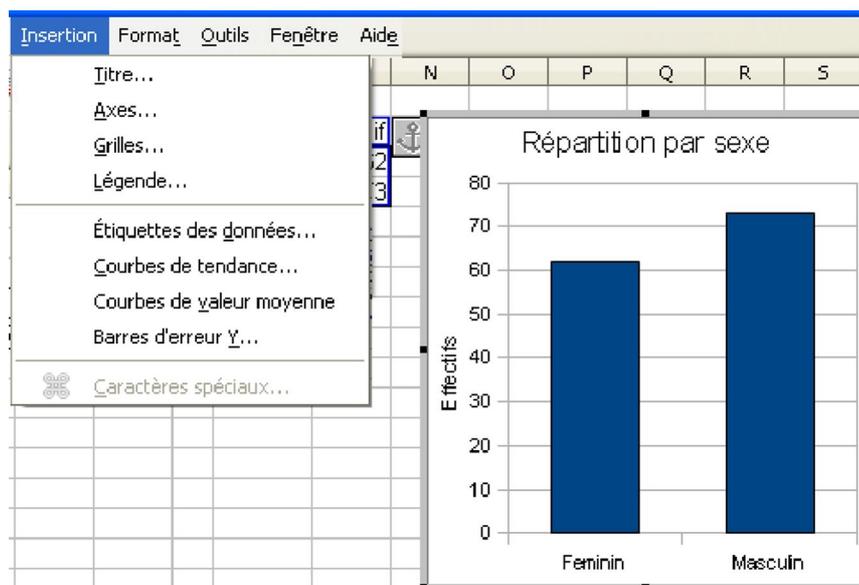


Figure 5.5 : Accès au « menu contextuel » lorsque le graphique est sélectionné

Remarque



Sous OpenOffice, on ne peut accéder aux objets du diagramme que s'il est entouré d'un trait épais, comme on peut le voir dans les figures 5.2 à 5.5. Quand on le crée, il est sélectionné de cette façon, mais si on sélectionne un autre objet du tableur ensuite, il faut alors faire un double-clic gauche dessus pour le re-sélectionner ainsi. Quand il est entouré de points verts (lors de sa sélection par un simple clic gauche), on ne peut pas accéder à ses objets mais on change sa taille ou sa position dans la feuille de calcul. Dans les dernières versions d'Excel, il n'est pas nécessaire de double-cliquer sur le graphique pour atteindre ses objets, un simple clic est suffisant. Notons également que dans les versions actuelles, il est possible de modifier un graphique déjà créé de façon beaucoup plus souple sous Excel que sous OpenOffice.

Les diagrammes sont organisés en types et sous-types. Par exemple dans le type colonne, on peut choisir le type 3D, et des cônes plutôt que des barres.

Quand la valeur numérique est une mesure (âge en années, salaire en euros, poids en kg, note à un examen, etc.) on privilégie les « lignes », les « toiles », comme dans la figure 5.7.

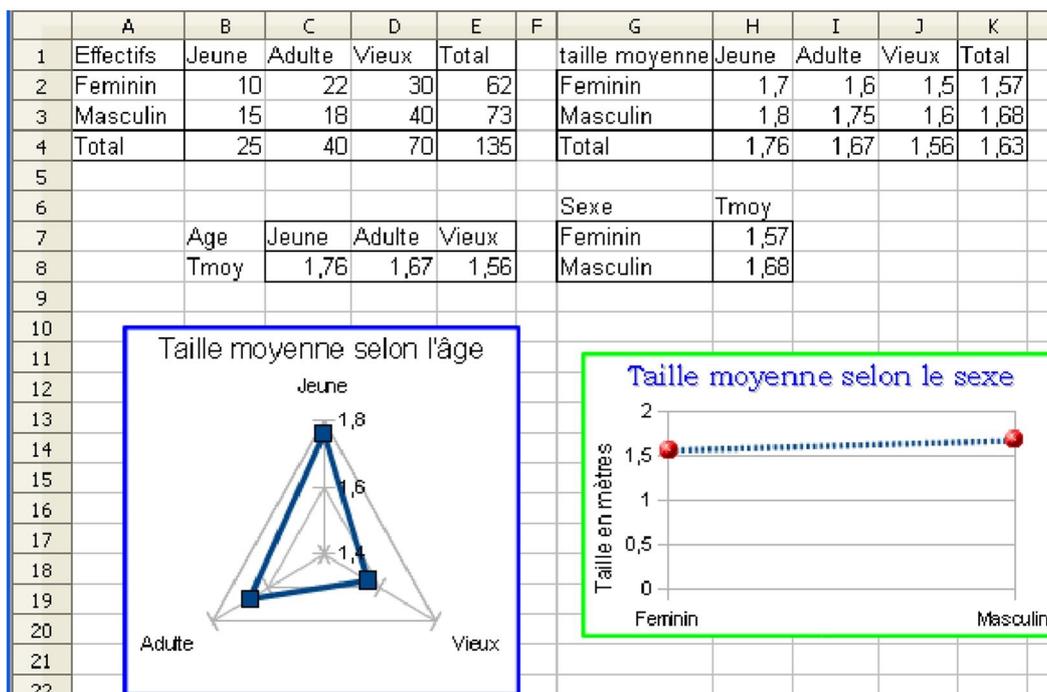


Figure 5.7 : Réalisation de deux diagrammes représentant la taille moyenne selon l'âge (toile) et le sexe (lignes)

3. Exercice1

Question

[Solution n°85 p 170]

Réaliser les deux diagrammes de la figure 5.7 à partir des données des 6 effectifs et des 6 tailles des groupes obtenus en faisant varier le sexe et l'âge, les autres valeurs étant obtenues par formules. On veillera à reproduire exactement le format des graphiques de la figure : zone de diagramme avec un trait bleu épais pour le diagramme de gauche, et pour le diagramme de droite, zone de diagramme avec un trait vert épais, titre en bleu avec relief et une police de caractère de taille inférieure, ligne bleue en pointillés fins, symboles en boule rouge, axe des y gradué à partir de 0.

4. Une série de valeurs numériques, répartition en groupes selon deux variables qualitatives

En suivant le même principe que dans le paragraphe 1.1, et en l'appliquant aux données de A1 : D4 et G1 : K4, on peut disposer des mêmes diagrammes dans lesquels une des deux variables qualitatives est en axe des x, comme précédemment, pendant que l'autre est en légende, avec autant de codes couleurs différents que de modalités. Dans la figure 5.8, le diagramme de type « colonnes » contient trois colonnes de couleurs différentes, une par tranche d'âge. Dans le langage du tableur, ce sont trois « séries » de 2 nombres (un effectif par sexe) qui sont représentées.

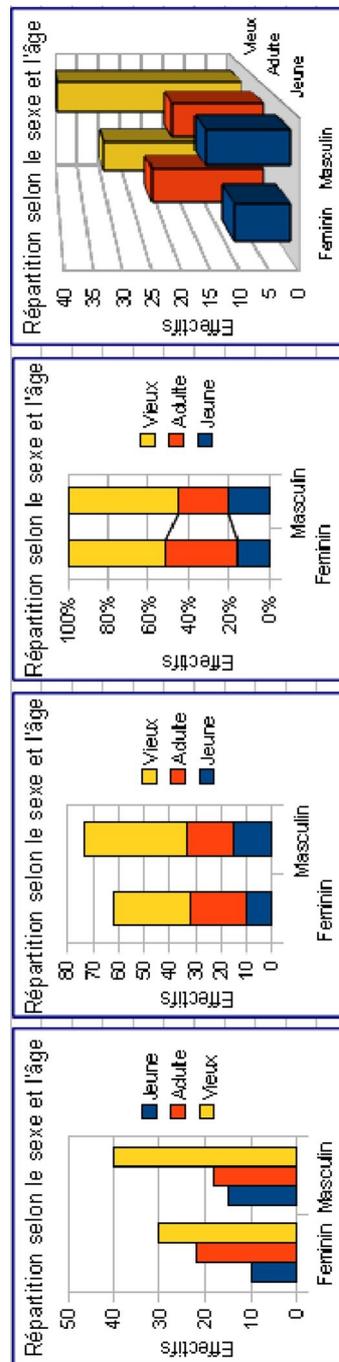


Figure 5.8 : De gauche à droite les sous-types normal, empilé, pourcentage empilé, 3D simple du type barre.

Pour réaliser les quatre diagrammes de la figure 5.8 sur la même page A1 : D4, on a fait un copier-coller du premier, puis après avoir sélectionné la copie par un double-clic, on a modifié par un clic droit en suivant le chemin « Type de diagramme > colonnes ». Dans le troisième graphique de la figure 5.8, on a ajouté des « lignes de connexion » grâce à la fenêtre de propriétés de la figure 5.9. Dans le quatrième graphique, les polices de caractères des deux axes ont été diminuées afin que les modalités des deux variables qualitatives apparaissent au complet.

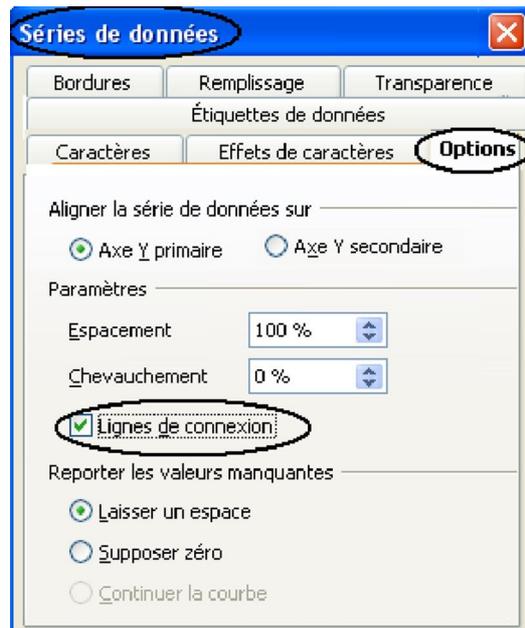


Figure 5.9 : Propriété des séries de données d'un diagramme à colonnes

C'est également ainsi que le diagramme de la figure 5.10 a été créé, en modifiant non seulement le type de diagramme, mais également la plage des données sources. Puis le rôle des deux variables qualitatives a été échangé, en cochant l'option « série de données en colonnes ». Les 3 séries de 2 nombres ont ainsi été remplacées par deux séries de 3 nombres, l'axe des x indiquant l'âge et la légende le sexe.

Ces diagrammes, colonnes et lignes, sont les plus courants pour ce type de données et les plus faciles à réaliser et à modifier selon l'effet attendu. D'autres types de diagrammes sont possibles, différant selon les versions de tableur. Et les possibilités de modifications sont également variées, par exemple les points peuvent être remplacés par des symboles variés, voire des images, les types de lignes sont nombreux, les colonnes peuvent être hachurées, etc.

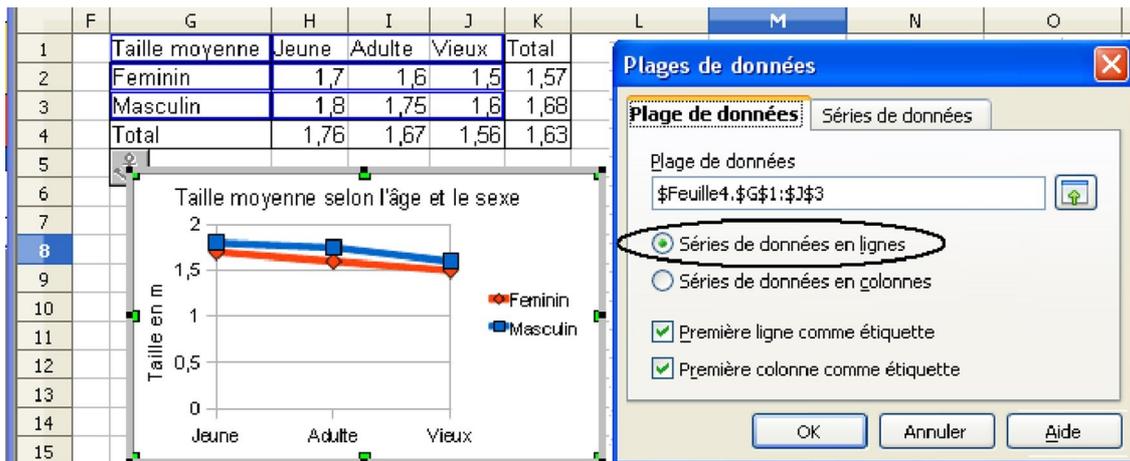


Figure 5.10 : Interprétation des différents éléments du tableau de données pour leur représentation

Par exemple, les séries boursières peuvent être représentées par des diagrammes spécifiques, comme celui de la figure 5.12 représentant les données de la figure 5.11 trouvées à l'aide de la requête « indice cac 40 historique ». Dans Excel 2003

comme dans OpenOffice3.1, on dispose de 4 sous-types de graphiques boursiers selon les informations dont on dispose.



Figure 5.11 : Interprétation des différents éléments du tableau de données pour leur représentation

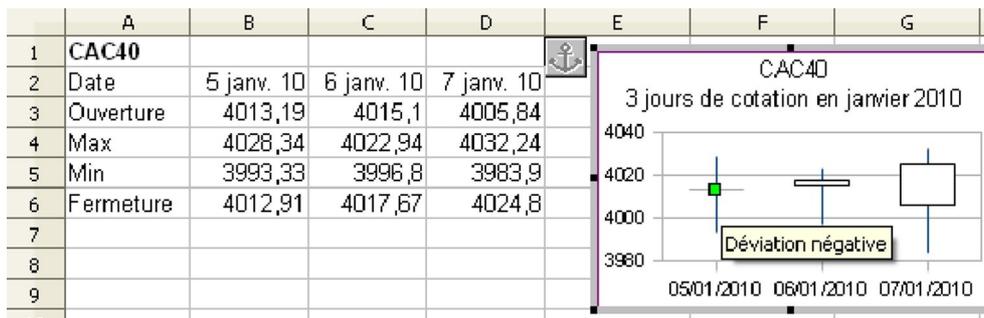


Figure 5.12 : Interprétation des différents éléments du tableau de données pour leur représentation

Remarque



On peut noter en observant la figure 5.12 que

1. la baisse (fermeture < ouverture) est représentée par un simple trait alors que les hausses sont représentées par des boîtes.
2. l'axe des x contient bien des dates, mais sous forme d'étiquette dont le format ne peut pas être modifié comme il l'a été dans la ligne 2 de la feuille de calcul.

5. Exercice2

Question

Réaliser le graphique de la figure 5.12 après avoir récupéré des données de même type sur Internet. Attention de bien veiller au format des données une fois mises dans les cellules : on doit avoir des formats de date et des formats de nombre et non du texte.

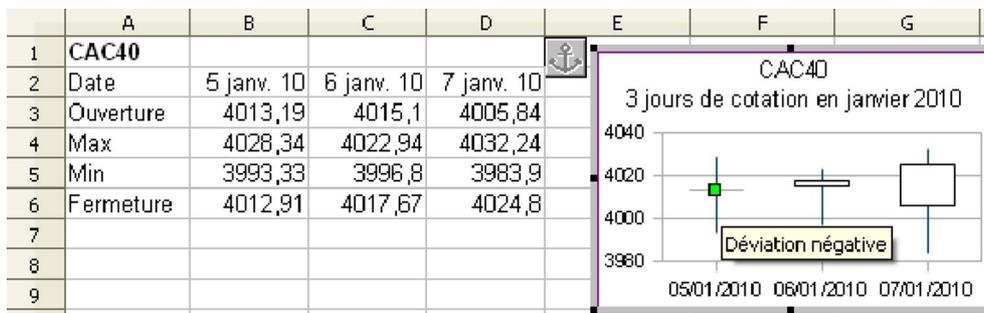


Figure 5.12 : Interprétation des différents éléments du tableau de données pour leur représentation

6. Deux variables quantitatives

Alors que dans les diagrammes précédents un seul axe était numérique, les autres étant munis d'étiquettes, on a maintenant deux axes numériques, un par variable. Ce type de diagramme est appelé selon les tableurs « nuage de points », ou « diagramme XY ». Des diagrammes permettant de dessiner des surfaces, comme des courbes de niveau, existent en Excel, mais nous nous limiterons ici aux simples nuages de points communs à toutes les versions de tableur.

Exemple : le fichier « 2actions_cac40.txt » contient 3 colonnes, la première avec

les dates de 255 jours ouvrables de la bourse et les deux autres avec les valeurs de clôture de ce jour de deux actions du CAC40. La représentation graphique des deux colonnes de valeurs est dans la figure 5.13.

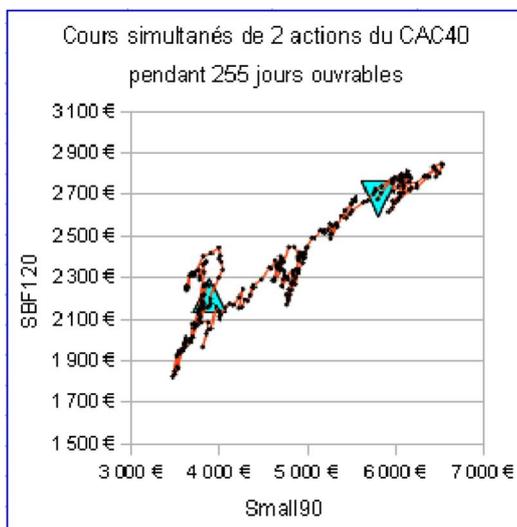


Figure 5.13 : Diagramme XY (nuage de points) pour les données du fichier « 2actions_cac40.txt »

Pour ce graphique, à chaque jour de cotation (sauf le premier et le dernier) correspond un point dont la valeur en x est le cours de la première action et dont la valeur y est le cours de la deuxième. Les points sont indiqués par un symbole de petite taille de couleur bleue et joint par une ligne rouge dans l'ordre chronologique. On a ajouté deux grands triangles bleus pour les points extrêmes de la série, celui qui pointe vers le haut correspond au premier point et celui qui pointe vers le bas au dernier point.

7. Exercice3

Question

[Solution n°86 p 170]

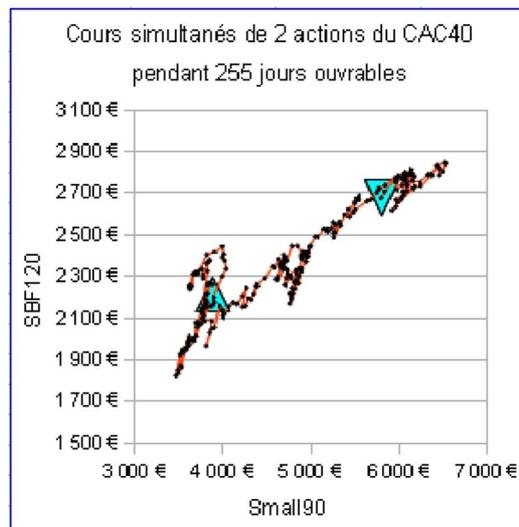


Figure 5.13 : Diagramme XY (nuage de points) pour les données du fichier « 2actions_cac40.txt »

Réaliser le graphique de la figure 5.13 à partir des données du fichier « 2actions_cac40.txt ».

8. Exercice4

Question

[Solution n°87 p 171]

À partir des données du fichier « Cac40_1an.txt », issues d'un fichier Excel trouvé sur le site <http://www.club-du-cac40.com>, réaliser le graphique de la figure 5.15. On veillera en particulier à ce que l'axe des x soit gradué comme indiqué sur cette figure.

9. Plus de variables quantitatives et qualitatives

Il est toujours possible de représenter plusieurs séries sur un même graphique, mais il faut généralement se limiter à 2 axes, dont 1 ou 2 peuvent être quantitatifs. La légende peut permettre de prendre en compte une variable qualitative de plus (voir figure 5.8). Dans certaines versions de tableur, on peut utiliser deux graduations différentes pour un même axe. Notons que plusieurs graphiques côte à côte sont souvent plus lisibles qu'un seul graphique réunissant l'information de tous. Une bonne synthèse des données se doit d'être compréhensible.

C. Une modélisation des données « à la main » utilisant le côté dynamique des graphiques

Les diagrammes faits dans une feuille de calcul sont attachés aux cellules de celles-ci. Si on modifie les contenus des cellules, ils sont mis à jour en temps réel. Cela permet d'ajuster « à la main » un modèle paramétrique à des données, si ce modèle est représentable graphiquement.

Les modèles statistiques les plus courants sont ceux dont la loi de distribution est

1) unidimensionnelle : concernant une variable unique, 2) paramétrique : définie par une formule dépendant de la valeur de la variable et de quelques valeurs appelées paramètres qui sont fixées arbitrairement ou estimées d'après les données. Ces lois peuvent être continues, c'est-à-dire prendre toute valeur d'un intervalle (ou d'une réunion d'intervalles) de \mathbb{R} , ou discrètes et prendre leurs valeurs sur un ensemble fini ou dénombrable de nombres, le plus souvent entiers.

Parmi ces lois, les plus utilisées sont :

- la loi uniforme continue définie sur un intervalle $I=[a, b]$ donné, ou discrète définie sur un ensemble fini E de valeurs données. Dans le premier cas, la probabilité que la valeur de la variable soit dans un intervalle quelconque inclus dans I ne dépend que de la longueur de l'intervalle et pas de ses bornes, et dans le second cas la probabilité qu'elle soit dans un sous-ensemble de E ne dépend que du nombre d'éléments de ce sous-ensemble, pas de ses éléments.
- la loi normale $N(m,s)$ où les deux paramètres sont m , l'espérance (moyenne théorique) et s , l'écart-type. Cette loi est continue et permet d'établir, comme son nom l'indique (on l'appelle aussi loi de Laplace-Gauss), des « normes » : par exemple, son utilisation permet de prédire le nombre d'individus d'une population entre deux tailles données connaissant la taille moyenne de la population et son écart-type.
- la loi de Poisson $P(\lambda)$, son seul paramètre λ est à la fois l'espérance et l'écart-type. C'est la loi des événements rares, son utilisation permet par exemple de prédire le nombre de naissance d'enfants atteints d'une maladie rare. C'est une loi discrète, qui peut prendre toute valeur entière positive ou nulle, les grandes valeurs étant de probabilité quasiment nulle.
- la loi binomiale $B(n,p)$ qui donne la probabilité des divers nombres de succès pour n tirages quand p est la probabilité de succès pour un tirage. C'est également une loi discrète qui prend ses valeurs entières dans l'intervalle $[0,n]$.

Le but de l'exercice TD qui suit est d'ajuster au mieux des lois de probabilités sur des données en faisant varier leurs paramètres. Pour les lois continues, on essaiera de rendre la distance de Kolmogorov entre les courbes (fréquences cumulées des données et les fonctions de répartition théoriques) la plus petite possible, et pour les lois discrètes, la distance du Chi2 entre l'histogramme et les fonctions de densité.

D. Exercice TD : Ajustement graphique de lois de probabilités à des données

Dans ce TP, on cherche à ajuster graphiquement quelques lois de probabilités usuelles à une série de données représentant une variable aléatoire X . Selon le type des données, continues ou discrètes, on procède de façon différente. Dans le cas de données continues, on compare les fréquences cumulées de X observées aux lois de répartitions théoriques, et dans le cas de données discrètes les effectifs observés d'intervalles de valeurs de X aux effectifs attendus selon les lois de probabilités.

1. Les données sont continues

a) Représentation des données brutes

On dispose dans le fichier « data_continues.txt » de données X pouvant prendre toute valeur comprise entre 0 et 5 recueillies auprès d'un échantillon de N personnes. Les copier dans une feuille Excel, les trier par valeurs croissantes, et calculer leurs caractéristiques statistiques N, moyenne, écart-type (Fig. 6.1 : B11 à B13) ainsi que leurs fréquences cumulées (colonne B).

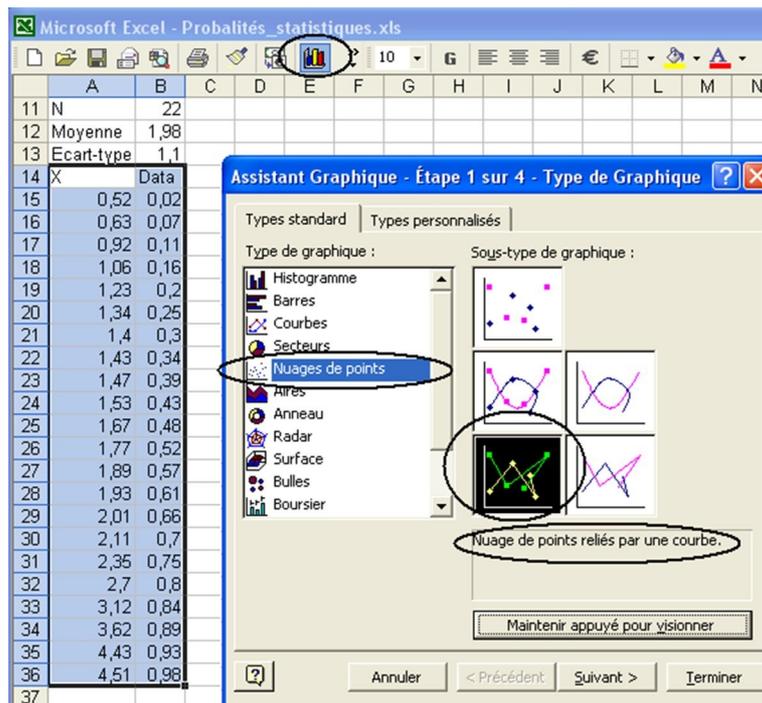


Figure 6.1 : Représentation graphique des fréquences cumulées de X (Étape 1)

Aide :

Les fréquences cumulées sont habituellement les valeurs k/N avec $k = 0, 1, \dots, N-1, N$. Pour faciliter l'ajustement aux lois de probabilités continues, on a choisi $k=0,5 ; 1,5 ; N-0,5$.

Après sélection des deux colonnes de données, on appelle l'assistant graphique (en cliquant sur l'icône correspondant, ou en allant dans les menus déroulants Insertion>graphique sous Excel, ou insertion>diagramme sous OpenOffice).

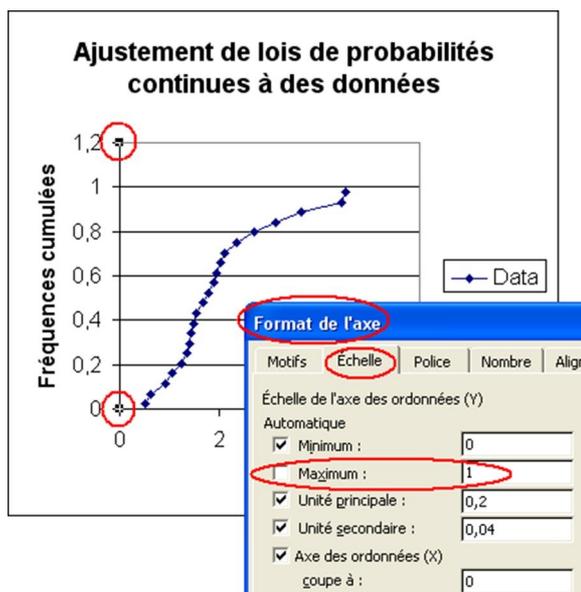


Figure 6.2 : Représentation graphique des fréquences cumulées de X (Étape 5)

Dans l'étape 1 de l'assistant graphique, on choisit la représentation graphique adaptée (« nuage de points sous Excel », « diagramme XY » sous OpenOffice), et on remplit pour chaque étape les champs demandés, puis on sélectionne le graphique afin qu'il ressemble à celui de la figure 6.2.



Conseil

Si vous ne savez pas comment remplir les champs ou choisir les options parmi celles proposées dans les diverses fenêtres de l'aide, vous pouvez valider sans remplir aucun champ. Il suffit alors de sélectionner le graphique obtenu (par un simple clic sous Excel, et un double clic sous OpenOffice) et de corriger un à un les divers éléments en les sélectionnant par un clic gauche puis en choisissant leur format par un clic droit. On peut voir dans la figure 6.2 la modification du format de l'axe des y.

Les objets du graphique sont atteignables par un clic gauche une fois le graphique sélectionné, à condition qu'ils existent dans le graphique. Si ce n'est pas le cas, par exemple pour le titre, les intitulés des axes, la légende, il faut d'abord les ajouter. Pour cela dérouler le menu « insertion » une fois le graphique sélectionné, et remplir alors les champs demandés.

Sous Excel, on peut faire défiler les divers objets du graphique (zone de graphique, zone de traçage, titre du graphique, légende, axes, séries de points, grille secondaire) par les touches de déplacement vers le haut et vers le bas du clavier, puis les divers éléments d'un objet par les touches de déplacement gauche et droite, par exemple les points successifs d'une série si on veut changer la couleur de l'un d'eux, les divers éléments de la légende, chacun des axes, etc.

b) Ajustement de la loi normale aux données

On choisit une espérance et un écart-type quelconques (par exemple dans C12 une espérance de 2,5 et dans C13 un écart-type de 10, voir Fig. 6.3), et on crée dans la colonne C les données de la loi de répartition normale $N(2,5 ; 10)$ en écrivant dans C15 la formule « =LOI.NORMALE(A15;C12;C13;1) », puis en la recopiant vers le bas après avoir ajouté les \$ aux endroits appropriés. Puis on ajoute la courbe

correspondante à celle des données brutes.

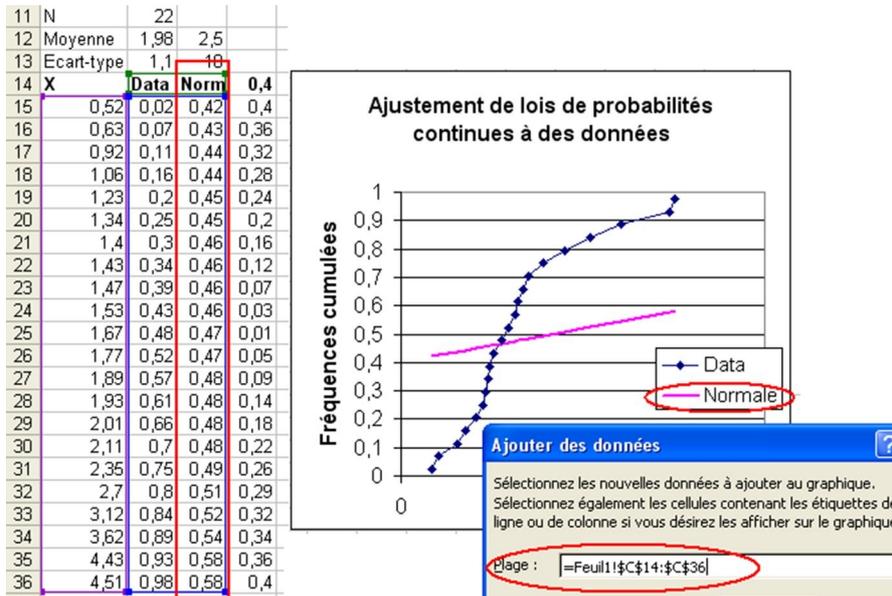


Figure 6.3 : Ajout de la fonction de répartition de la loi normale $N(2,5 ; 10)$

On peut voir sur la figure 6.3 que la courbe de la loi normale diffère beaucoup de la courbe des données. Cela est dû au mauvais choix des paramètres de la loi normale (espérance, écart-type). Nous allons essayer d'améliorer l'ajustement en faisant un choix plus judicieux.

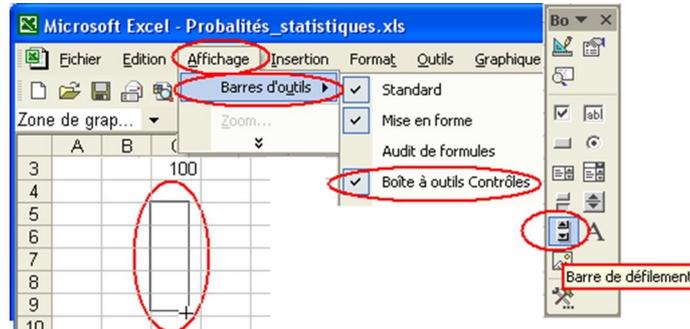


Figure 6.4 : création de la barre de défilement pour ajuster l'espérance de la loi normale

Plutôt que de saisir successivement de nombreuses valeurs, nous créons une barre de défilement (voir Fig. 6.4) dont le rôle est de remplir la cellule C2 d'un nombre entier compris entre 0 et 100, en allant de 1 en 1 (petit changement : SmallChange) ou de 20 en 20 (grand changement : LargeChange). Un clic droit sur la barre de défilement créée permet d'atteindre les propriétés de celle-ci et de saisir ses paramètres (voir en rouge sur la figure 6.5).

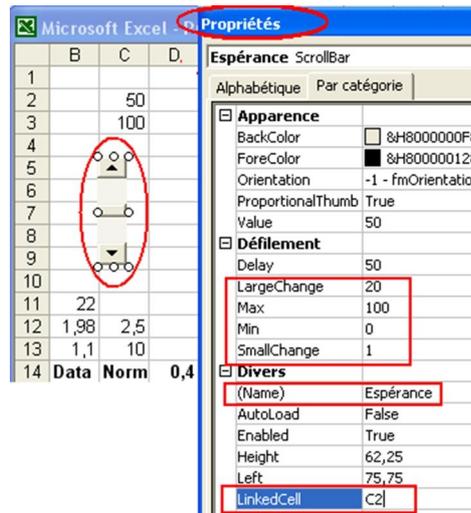


Figure 6.5 : choix des paramètres de la barre de défilement de l'espérance

Après un clic sur l'icône du « mode création » (en haut à gauche de la « boîte à outils contrôles ») pour le désactiver, on utilise la barre de défilement pour modifier le contenu de la cellule C2, comme indiqué figure 6.6.

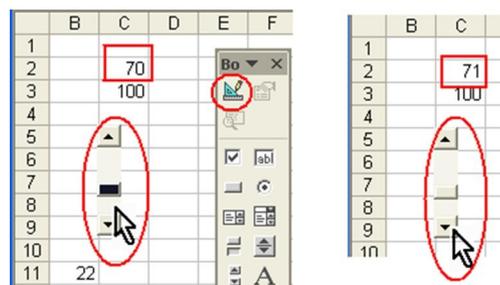


Figure 6.6 : à gauche clic pour un « LargeChange », à droite pour un « SmallChange »

On crée ensuite la barre de défilement pour la cellule C3. On peut procéder comme précédemment, ou activer le « mode création », faire un copier-coller de la barre de défilement précédente, et modifier ses propriétés (voir Figure 6.7).

Les deux cellules C2 et C3 contenant des nombres entiers qui peuvent varier par pas de 1, on peut faire varier par pas de 0,05 l'espérance en tapant dans la cellule C12 la formule « =C2*0,05 », et par pas de 0,1 l'écart-type en tapant dans la cellule C13 la formule « =C3*0,1 ».

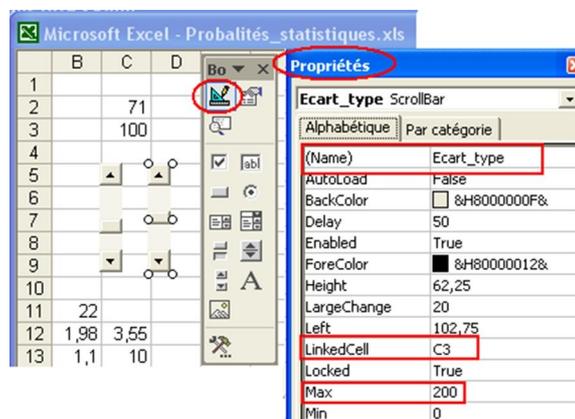


Figure 6.7 : choix des paramètres pour la deuxième barre de défilement

Il ne reste plus qu'à déplacer les barres de défilement pour obtenir la courbe la plus ajustée aux données. Pour rendre la recherche plus facile, on peut s'aider d'un indicateur d'écart entre les deux courbes, par exemple est la plus grande distance verticale. Pour cela on a écrit dans la colonne D (voir Fig. 6.8) pour chaque point les écarts verticaux entre les deux courbes, et en D14 le maximum de ces écarts, qui est à minimiser.

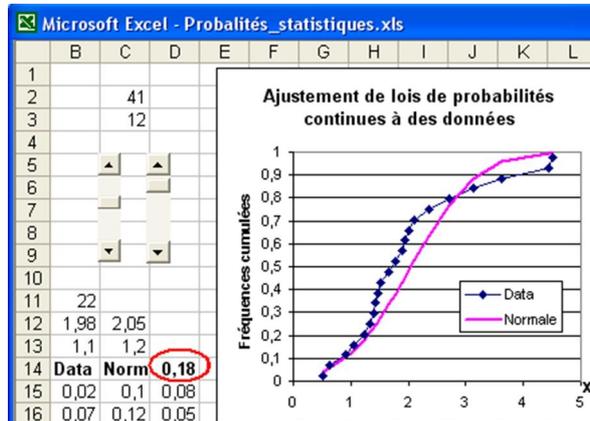


Figure 6.8 : Utilisation des deux barres de défilement pour choisir des paramètres optimaux

On peut voir dans la figure 6.8 le résultat sur le graphique du choix de bons paramètres pour cette loi, choix facilité par les barres de défilement. A ce choix correspond un écart vertical de 0,18. Par la méthode des moments, pouvez-vous obtenir mieux ? Et par tâtonnement ? En utilisant le solveur ?

c) Ajustement de deux autres lois continues aux données

On propose d'ajuster de la même façon la loi de Weibull et la loi exponentielle aux données. On peut voir le résultat dans la figure 6.9 ci-dessous. La loi la plus ajustée aux données semble être la loi de Weibull. Réaliser les calculs et le graphique.

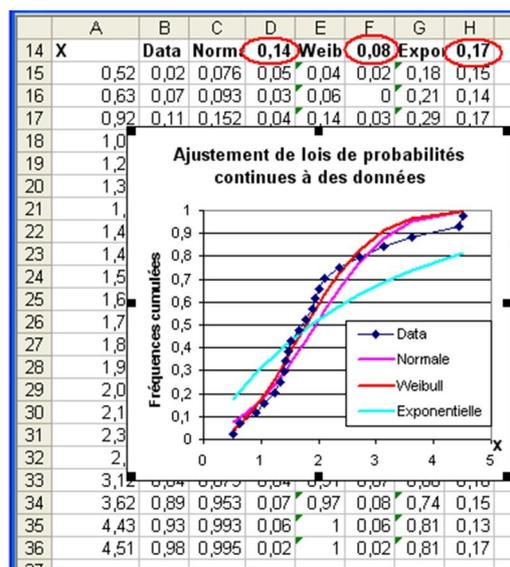


Figure 6.9 : ajustement des données par les lois normale, de Weibull, exponentielle

d) Correction

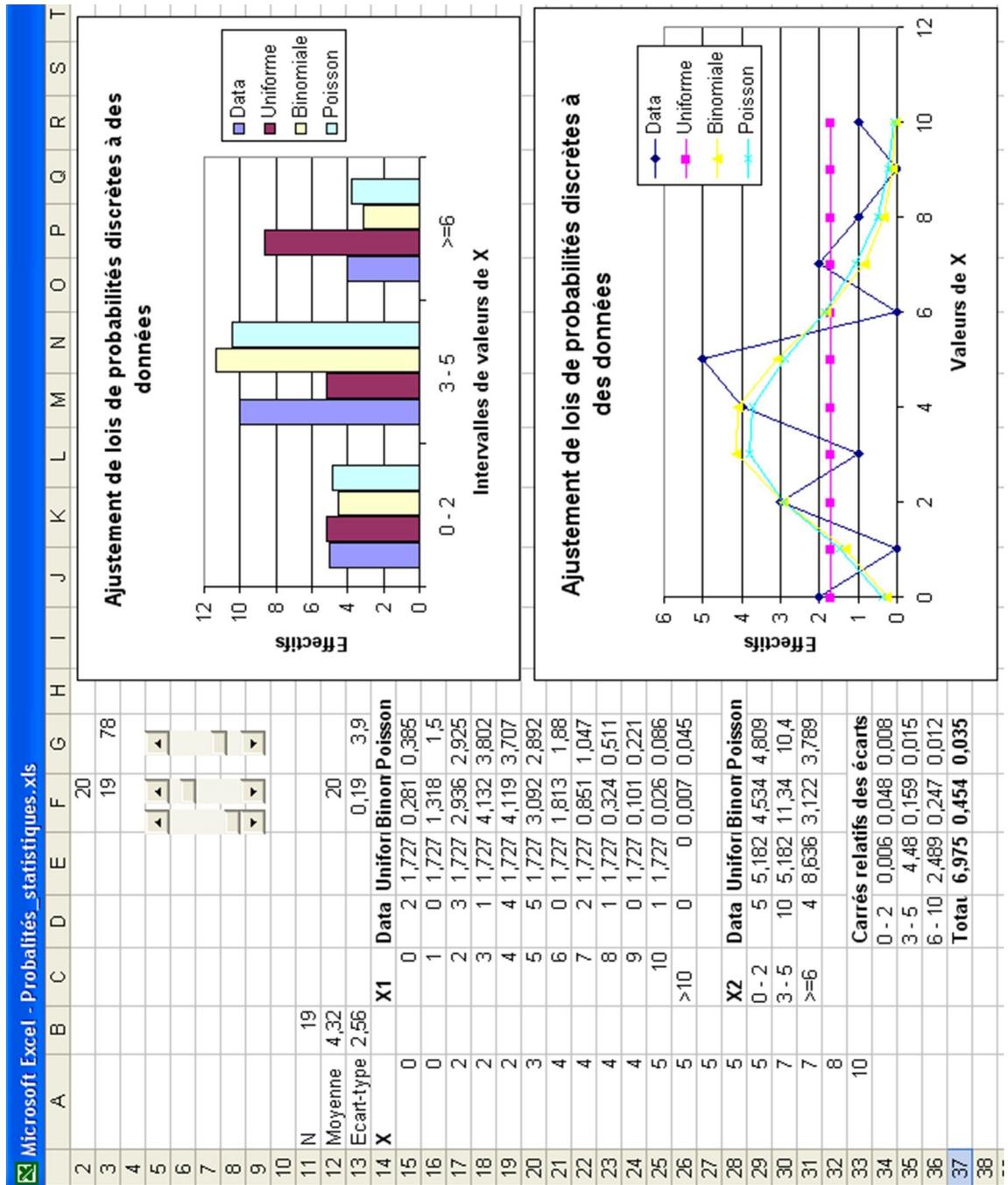


Figure 6.11 : Les paramètres optimaux d'ajustements obtenus avec le solveur

Pour la loi de répartition des données

Cellules saisies directement : colonne A, B14, C14, lignes 2 et 3.

B11 =NB(A15:A36)

B12 =MOYENNE(A15:A36)

B13 =ECARTYPE(A15:A36)

B15 =0,5/B11

$B16 = B15 + 1/B\$11$

copier-coller de B16 jusqu'à B36

Réaliser le graphique avec les données des colonnes A et B (sélectionner la plage de A14 à B36).

Pour la loi normale :

en C2 et C3 des valeurs entières quelconques, par exemple 50 et 100

$C12 = C2 * 0,05$

$C13 = C3 * 0,1$

$C15 = \text{LOI.NORMALE}(\$A15; C\$12; C\$13; 1)$

copier-coller de C15 jusqu'à C36

$D14 = \text{MAX}(D15:D36)$

$D15 = \text{ABS}(B15 - C15)$

copier-coller de D15 jusqu'à D36

Ajouter au graphique la fonction de répartition de la loi normale

lignes 2 et 3 remplies avec "Barre de défilement de nombres entiers utilisés pour faire varier les paramètres (ces nombres multipliés par 0,1 ou 0,01, ou ...) des lois situés dans les lignes 12 et 13.

Faire varier les valeurs de l'espérance et de l'écart-type pour trouver une courbe de loi normale qui ait un écart minimum (contenu de la cellule D14) avec celle des données.

Etirer la plage C2:D36 pour qu'elle aille de C2 à H36

Remplacer l'intitulé "Normale" par Weibull puis Exponentielle

et la formule de la loi normale par les lois correspondantes :

$E15 = \text{LOI.WEIBULL}(\$A15; E\$12; E\$13; 1)$

$G15 = \text{LOI.EXPONENTIELLE}(\$A15; G\$13; 1)$

et les étirer vers le bas.

Puis supprimer le contenu des cellules G2 et G12.

Insérer ces nouvelles lois de répartition dans le graphique.

2. Les données sont discrètes

a) Représentation des données brutes

On dispose dans le fichier « data_discrettes.txt » de données X pouvant prendre des valeurs entières à partir de 0, recueillies auprès d'un échantillon de N personnes. Les copier dans une feuille Excel, les trier par valeurs croissantes, et calculer leurs caractéristiques statistiques N, moyenne, écart-type (Fig. 6.1 : B11 à B13). Puis trouver les effectifs correspondant à chaque valeur (plage C14 :D26).

b) L'ajustement de trois lois théoriques

Les lois de probabilités utilisées pour ajuster ces données sont :

- la loi uniforme
- la loi binomiale
- la loi de Poisson

La loi uniforme a pour seul paramètre le nombre de valeurs, que nous avons choisi

ici égal à 11. Le somme des effectifs pour chaque valeur de X1 est la même et l'effectif total est N, ce qui fait que chaque effectif est de nombre_classes/N.

La loi binomiale représente le nombre de succès obtenus en répétant n fois de façons indépendantes une expérience ayant une probabilité p de succès. Nous prenons pour n un entier entre 10 et 20, et pour p une valeur entre 0 et 1 variant avec un pas de 0,01.

La loi de Poisson représente le nombre de survenues d'un événement rare pendant une période donnée, sachant qu'une nouvelle occurrence de cet événement peut arriver à tout moment de la période, indépendamment de l'arrivée de l'occurrence précédente, avec en moyenne λ occurrences pendant la période.

Comme dans la partie précédente, on utilise des barres de défilement pour ajuster les paramètres des lois. Pour pouvoir comparer les lois théoriques à la distribution empirique des données, on regroupe les valeurs en classes, puis on compare les effectifs selon la distance du Chi2. Le regroupement est fait dans les cellules C28 à G31 et donne la série X2, représentée par un histogramme dans le graphique du haut de la figure 6.10, alors que la série X1 est représentée dans celle du bas par une courbe très accidentée, donc peu exploitable pour l'ajustement. Les écarts correspondant à X2 sont calculés dans la plage D33 :G37. L'ajustement se fait en observant les valeurs de la ligne 37, somme des 3 valeurs au dessus, la cellule E34 contenant la formule « $=(E29-D29)^2/E29$ », qui est recopiée vers la plage E34 :G36 une fois les \$ ajoutés si nécessaire.

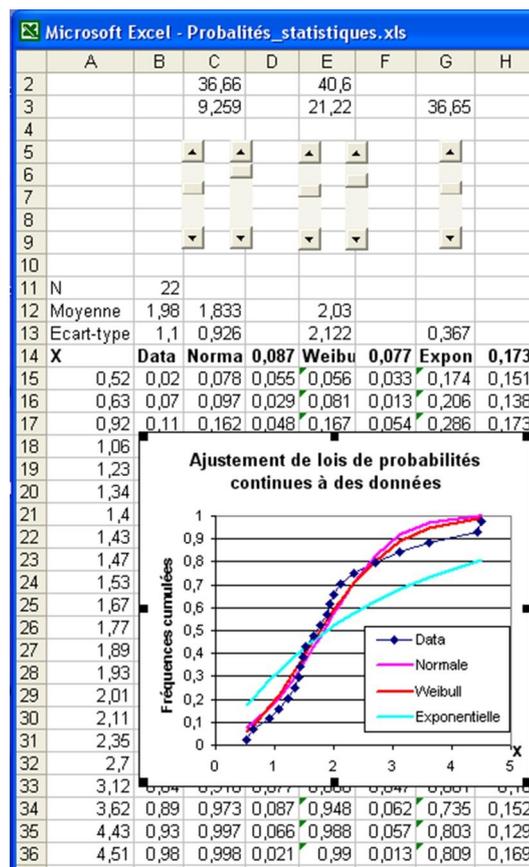


Figure 6.10 : ajustement des données par les lois uniforme, binomiale et de Poisson

Réaliser le travail correspondant à la figure 6.11. La loi de Poisson paraît la plus ajustée aux données. Aurait-on pu obtenir mieux par la méthode des moments ? Et

que donne le solveur ?

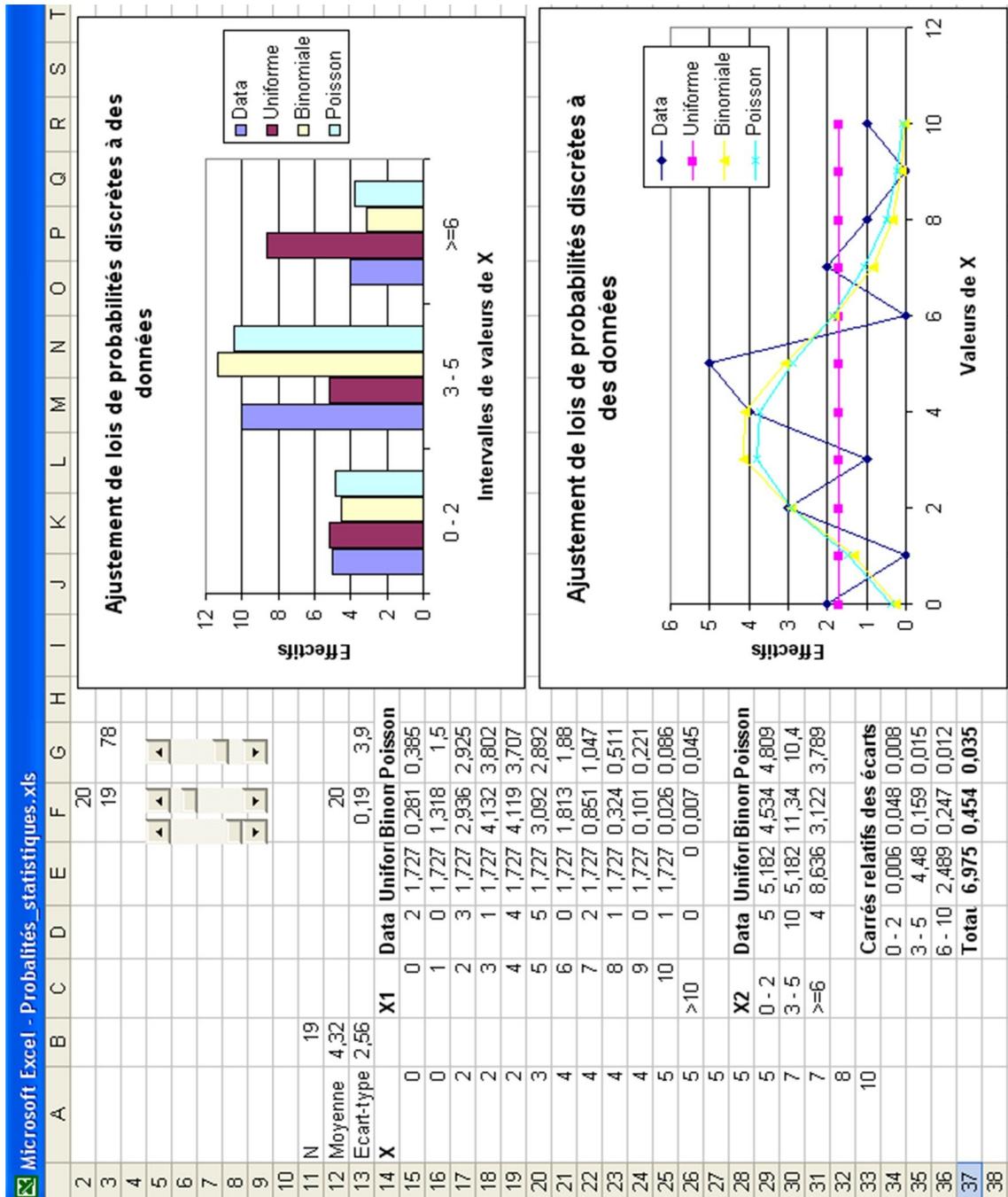


Figure 6.11 : Les paramètres optimaux d'ajustements obtenus avec le solveur

c) Correction

Dans F12 =F2

Dans F13 =F3*0,01

Dans E15 = =B\$11/NB(\$C\$15:\$C\$25)

Dans F15 =LOI.BINOMIALE(C15;F\$12;F\$13;0)*\$B\$11

Dans G15 =LOI.POISSON(C15;G\$13;0)*\$B\$11

Dans D29 =SOMME(D15:D17)

Dans E34 =(E29-\$D29)*(E29-\$D29)/E29

Dans E37 =SOMME(E34:E36)

E. Tendances pour modéliser les données

Cette modélisation est automatique. L'ajustement ne se fait pas comme précédemment « à la main », mais en utilisant les formules obtenues à partir de calculs formels. Particulièrement adaptée aux données temporelles, elle permet de dégager dans ce cas une tendance afin de prédire les valeurs qu'on pourrait obtenir dans un futur proche si la tendance se maintenait. Pour un autre usage, on peut aussi prédire des valeurs intermédiaires. Nous voyons dans cette partie les deux plus courantes : les moyennes mobiles et la régression linéaire. Pour le premier, les valeurs initiales sont remplacées par leurs moyennes sur un certain nombre de valeurs consécutives, ceci afin de lisser la courbe des données, et de faire apparaître visuellement la tendance. Les modèles de régression permettent, une fois un certain nombre d'hypothèses statistiques admises, de remplacer les données par une fonction simple de celles-ci, linéaire, affine, parabolique, cubique, logarithmique, etc. la courbe produite représentant la tendance.

1. Les moyennes mobiles

On choisit de préférence une période de k jours (ou mois, années, etc.) avec k impair : par exemple 3, 5, 7, 15 ou 31 jours consécutifs (on ne tiendra pas compte des dates, deux jours consécutifs représentant 2 jours ouvrés: samedi et lundi, lundi et mardi, etc.). La valeur à la date i est alors remplacée par la moyenne de sa valeur avec celle des $(k-1)/2$ dates avant et des $(k-1)/2$ dates après. On ne fait pas le calcul si $i < (k-1)/2$ ou si $i > n - (k-1)/2$ et la série lissée comporte donc moins de valeurs que la série brute. On peut voir figure 5.17 et 5.18 l'effet de plusieurs lissages sur les données de clôture du CAC40. Plus la période de lissage augmente, moins la courbe prend en compte de petites variations, les « zigzags » sont remplacés par des ondulations.

On peut être contraint de choisir un nombre pair k de valeurs (12 mois par exemple ou 4 trimestres si on désire un modèle annuel des données). Dans ce cas, on considère $k+1$ valeurs consécutives, $k/2$ avant la valeur à lisser et $k/2$ après, et on les ramène à k valeurs en remplaçant la première et la dernière par leur moyenne.

	A	B	C	D	E	F
1	Date	Clôture	Lissage 5	Lissage 7	Lissage 15	Lissage 31
2	3 déc. 03	3501,93				
3	4 déc. 03	3496,55				
4	5 déc. 03	3457,14	3469,33			
5	8 déc. 03	3434,91	3456,71	3464,77		
6	9 déc. 03	3456,12	3450,98	3460,3		
7	10 déc. 03	3438,85	3453,68	3459,42		
8	11 déc. 03	3467,9	3464,78	3463,63		
9	12 déc. 03	3470,6	3470,87	3470,05	3478,82	
10	15 déc. 03	3490,42	3479,08	3476,78	3479,38	
11	16 déc. 03	3486,6	3486,14	3485,81	3480,99	
12	17 déc. 03	3479,87	3492,43	3489,83	3485,8	
13	18 déc. 03	3503,21	3493,56	3494,04	3494	
14	19 déc. 03	3502,04	3496,25	3496,88	3503,37	
15	22 déc. 03	3496,06	3502,34	3501,77	3514,67	
16	23 déc. 03	3500,09	3505,86	3508,81	3523,2	
17	24 déc. 03	3510,3	3511,28	3516,62	3529,39	3534,52
18	29 déc. 03	3520,79	3523,65	3530,16	3536,21	3539,62
19	30 déc. 03	3529,18	3542,99	3546,19	3542,09	3545,42

Figure 5.17 : Calcul des divers lissages des données de « Cac40_1an.txt »

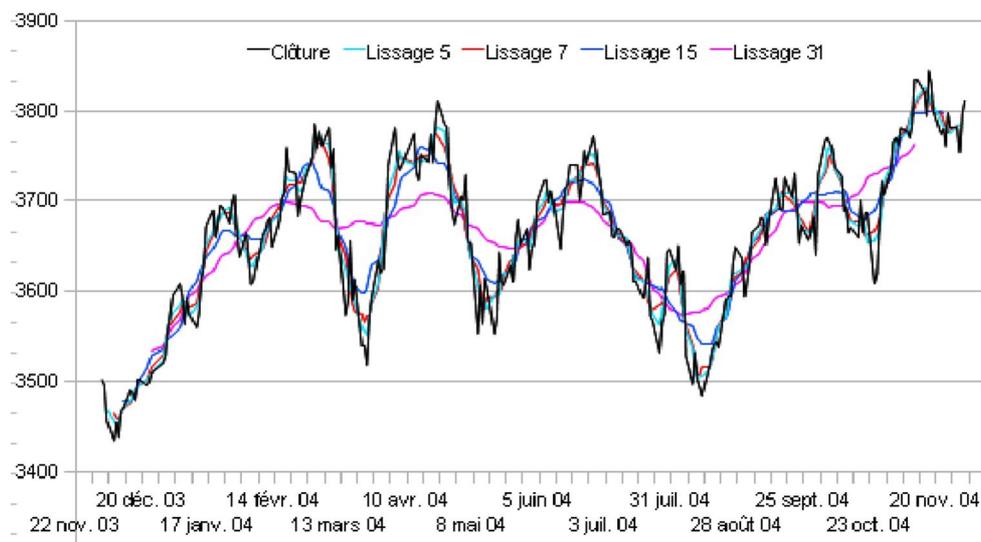


Figure 5.18 : Graphiques des divers lissages des données de « Cac40_1an.txt »

2. La régression linéaire

Définitions de la régression linéaire

Les modèles de régression permettent d'ajuster une variable quantitative y , appelée variable à expliquer (ou variable dépendante), à une fonction f de variables quantitatives ou même qualitatives, appelées variables explicatives (ou variables indépendantes, ou régresseurs).

Par exemple, on donne parfois comme relation entre la taille x d'un individu adulte et son poids idéal y l'équation $y = (x-100) - (x-150)/4$ soit $y = 0,75x - 62,5$. Parmi ces modèles, $f(x)=ax+b$ est le modèle le plus simple de régression linéaire, qui permet d'ajuster le nuage de points (x, y) , où x et y sont deux variables quantitatives, à la droite d'équation $y_{est.} = ax+b$, également appelée droite de

régression de y sur x . Dans cette équation, $y_{est.}$ est la valeur de y estimée selon le modèle, et non la valeur de y observée dans les données, leur différence $y_{obs.} - y_{est.}$ étant ε , appelée résidu (ou erreur, ou écart au modèle). Dans l'exemple précédent, $y_{est.}$ est le poids idéal d'un individu de taille x , qui diffère généralement de son poids réel $y_{obs.}$. Les paramètres a et b du modèle, respectivement appelés pente de la droite et ordonnée à l'origine sont calculés à partir de toutes les valeurs de x et de y avec la méthode des moindres carrés ordinaires, que nous ne détaillerons pas ici, le tableur se chargeant de tous les calculs. La qualité explicative globale de ce modèle est mesurée en partie par la corrélation linéaire (coefficient de corrélation linéaire de Bravais-Pearson) existant entre la valeur estimée et la valeur réelle, ou plutôt par son carré R^2 , appelé coefficient de détermination, qui exprime le pourcentage de variation de y expliqué par le modèle.

Modèles emboîtés de régression linéaire

Si le modèle de régression linéaire simple paraît insuffisant pour expliquer y , on peut ajouter d'autres variables explicatives. Avec p variables, l'une à expliquer et les autres explicatives, le nuage de points dans l'espace de dimension p peut être ajusté à un hyperplan de dimension $p-1$ d'équation $y_{est.} = a_1x_1 + a_2x_2 + \dots + a_px_p + b$, le modèle correspondant étant appelé modèle de régression linéaire multiple. Quand $p=3$, on dit qu'on a un plan de régression. Plus le nombre de variables augmente, plus la valeur de R^2 augmente également. Malgré tout, le meilleur modèle n'est pas généralement celui comportant le plus grand nombre de variables explicatives.

a) La régression linéaire avec le tableur

Si on dispose d'un tableau de données de p variables pour n individus, la régression de l'une des variables sur les $p-1$ autres s'obtient par la fonction DROITEREG(y ; x ; constante; statistiques), où y est la colonne de n valeurs de la variable à expliquer, x sont les $p-1$ colonnes de n valeurs des variables explicatives, constante est à VRAI si on accepte une valeur de b non nulle, et statistiques est à VRAI si on désire avoir des informations statistiques autres que les valeurs de paramètres. Le résultat de cette fonction est le tableau ayant les p colonnes et 5 lignes suivantes quand constante; statistiques sont à VRAI ($p=4$ dans celui-ci):

a_3	a_2	a_1	b
e.t.(a_3)	e.t.(a_2)	e.t.(a_1)	e.t.(b)
R^2	e.t.(e_e)		
F	Ddl		
Ssreg	Ssres		

Tableau 16 : Régression linéaire avec le tableur

Dans la première ligne on a les paramètres du modèle, dans la ligne suivante, leurs écarts-types, et R^2 en dessous. Les autres valeurs sont utilisées pour construire des tests statistiques. Les cases grisées ne sont pas remplies de valeurs, mais de codes d'erreur. Pour obtenir le tableau de résultats, comme on désire plusieurs valeurs, on procède comme pour le calcul matriciel du tableur : on sélectionne la zone de p colonnes et 5 lignes, on tape =droitereg(, on sélectionne alors la colonne des y , puis les colonnes des x , et vrai et vrai, ces 4 arguments étant séparés par des point-virgule, et après avoir fermé la parenthèse, on valide en enfonçant simultanément les 3 touches adéquates, CTR, MAJ, Entrée..

Exemple



Dans la figure 5.19, on a 5 colonnes de valeurs pour 5 individus de B4 à F8. Le premier modèle de régression, M1, est écrit dans les cellules de H4 à I8. Sa formule est =DROITEREG(B4:B8;C4:C8;1;1). Avec le résultat figurant dans les cellules H4 et I4, on peut écrire le poids en fonction de la taille par la formule :

$$\text{Poids est.} = 1,14 * \text{Taille} - 125,22$$

La qualité de ce modèle est élevée car $R^2 = 0,92$ (cellule H6), valeur proche de 1.

Modèle M2 : $\text{Poids}_{\text{est.}} = 0,83 * \text{Taille} - 0,3 * \text{Age} - 57,65$, $R^2=0,94$, =DROITEREG(B4:B8;C4:D8;1;1).

Modèle M3 : $\text{Poids}_{\text{est.}} = 0,86 * \text{Taille} - 0,18 * \text{Age} - 0,003 * \text{Salaire} - 64,44$, $R^2=0,95$, =DROITEREG(B4:B8;C4:E8;1;1).

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	Poids	Taille	Age	Salaire	QI		M1			M2				M3			
4	75	180	45	1400	110		1,14	-125,22		-0,3	0,83	-57,65		0	-0,18	0,86	-64,44
5	48	152	68	2500	120		0,19	33,44		0,34	0,41	83,99		0,01	0,81	0,59	122,28
6	70	168	37	800	90		0,92	5,23		0,94	5,43	#N/D		0,95	7,55	#N/D	#N/D
7	80	182	28	1200	100		35,31	3		16,76	2	#N/D		5,79	1	#N/D	#N/D
8	92	185	30	1000	98		965,94	82,06		988,98	59,02	#N/D		990,97	57,03	#N/D	#N/D

Figure 5.19 : Quelques données et quelques modèles de régression linéaire

On peut remarquer que la valeur de R^2 est très forte pour M1, et qu'elle augmente très légèrement quand on passe de M1 à M2 puis M3. Ces trois modèles sont dits emboîtés car les ensembles de variables qu'ils utilisent sont inclus les uns dans les autres, et on peut calculer la significativité, au sens statistique du terme, de ces augmentations (les lecteurs intéressés peuvent consulter les ouvrages de G. Baillargeon sur la régression linéaire, par exemple celui intitulé « La régression linéaire » aux éditions des Trois Sources, Québec, 2000). Seul le premier modèle a vraiment un sens.

Dans la figure 5.20, on a analysé les apports du modèle M1. La colonne D a été remplie avec les valeurs estimées de P selon ce modèle (en D3, on a =C3*G\$3+H\$3), en colonne F les écarts P-P_{est.} ont été calculés (en C3, on a =B3-D3). À droite de la figure, on a représenté graphiquement trois séries :

- la série des points observés, indiqués par des carrés bleus, avec la taille en abscisse et le poids observé en ordonnée
- la série des points estimés, indiqués par des losanges rouges, avec la taille en abscisse et le poids estimé d'après le modèle M1 en ordonnée
- la droite de régression dessinée automatiquement par le tableur (sélection de la série de valeurs observées puis ensuite insertion> tendance linéaire)

On peut contrôler que la droite de régression passe bien par tous les points estimés, et que le nuage de points bleus est très allongé, ce qui est cohérent avec un R^2 proche de 1.

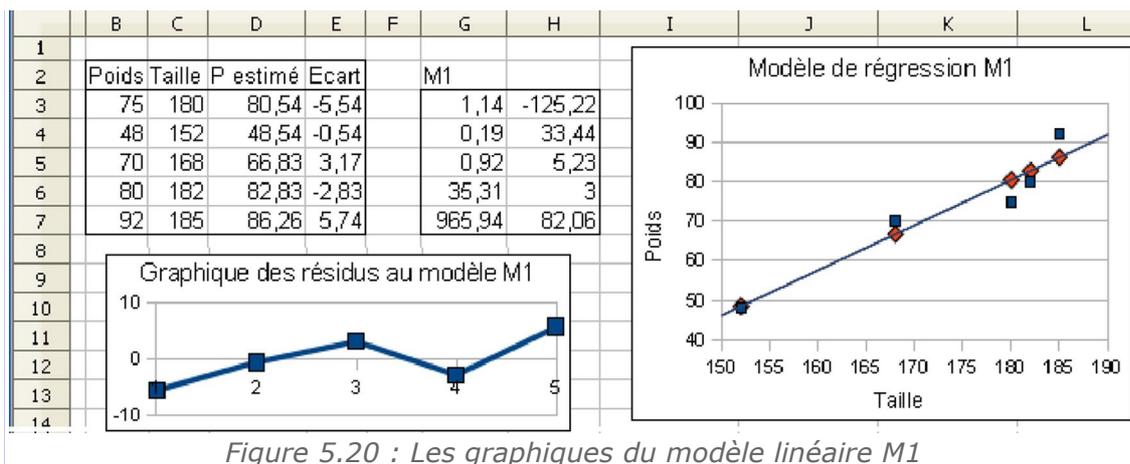


Figure 5.20 : Les graphiques du modèle linéaire M1

b) Les conditions de validité d'un modèle de régression linéaire

La connaissance de R^2 est importante pour apprécier la qualité de ces modèles mais elle ne suffit pas. En effet ces modèles de régression linéaire ainsi estimés d'après les données ne sont vraiment utiles que s'ils permettent de faire des prédictions réalistes, par exemple de « prédire » réellement le poids d'un individu sachant sa taille. Pour cela il faut que les hypothèses statistiques sur lesquelles s'appuient les modèles de régression linéaire soient respectées (pour plus de détails, se référer à l'ouvrage « Comment interpréter les résultats d'une régression linéaire? » de R. Tomassone, édité par l'ITCF, Paris, 1997). Voici les trois plus importantes, les deux premières sont préalables à la construction du modèle, la dernière doit être vérifiée une fois celui-ci construit, avant son utilisation.

1. La plus importante pour un modèle statistique, difficile à contrôler après coup, impose que les données aient été recueillies et soient utilisées selon la théorie de l'échantillonnage. Dans notre cas très simple de régression linéaire, il faut que les personnes interrogées soient un échantillon représentatif de la population sur laquelle on compte appliquer le modèle. Rappelons que faire des statistiques sur des données dont on ignore la provenance ne peut conduire qu'à des conclusions sans aucune valeur scientifique.
2. Même si les données ont été collectées en respectant les règles de la théorie de l'échantillonnage, il faut vérifier qu'il n'y a pas de données aberrantes (un homme de 300 cm, de 20 kg, etc.) dues à des erreurs humaines de saisie. Par exemple, pour le modèle M1 de la figure 5.20, un graphique permet de contrôler qu'il n'y a pas de données aberrantes, c'est-à-dire qui « dépassent » du nuage de points de dimension 2 (points bleus dans le graphique à droite de la figure 5.20). Quand la dimension du nuage de points augmente, il faut examiner la distribution des variables une à une.
3. Les résidus doivent être i.i.d. (Indépendamment Identiquement Distribués). On peut essayer de vérifier cela par un certain nombre de graphiques. Par exemple dans la figure 5.20, en bas à gauche, on a représenté la succession des 5 résidus. On peut voir qu'il y en a à peu près autant de positifs que de négatifs, mais avec si peu de valeurs on ne peut pas vraiment dire que les valeurs proches de 0 sont plus nombreuses que celles éloignées de 0, comme l'impose la loi normale. Cette loi est souvent postulée pour permettre d'obtenir des intervalles de confiance des valeurs prédites.

3. Exercice TD

Le but de ce TD est d'écrire différents modèles d'estimation du salaire d'une entreprise par l'ancienneté, la responsabilité et le sexe en utilisant des « régressions linéaires » sur des données artificielles du fichier salaires.txt. Cet exercice correspond à une démarche habituelle de recherche de modèle statistique des données : les relations linéaires sont les plus simples et donc celles qu'on « essaie » en premier, si on ne dispose pas d'autres modèles fournis par des théories des sciences humaines. Suivant cette logique, on commence habituellement par le modèle le plus simple, c'est-à-dire avec une seule variable explicative, et on essaie de l'améliorer. Nous allons réaliser ces différentes étapes à l'aide du tableur. Nous utilisons pour cela principalement la fonction « DROITEREG » du tableur. Bien que nous utilisions les formules matricielles et la loi normale dans ce TP, leur maîtrise n'est pas indispensable car tous les rappels utiles sont faits dans le TP.

Modèle 1

Le salaire en fonction de l'ancienneté

Question 1

[Solution n°88 p 171]

Q1 : Recherche du modèle M1 : $y=ax+b$

En utilisant la fonction DROITEREG du tableur, trouver les valeurs a et b correspondant à l'équation de régression Salaire = a Ancienneté + b + erreur minimisant les erreurs (voir figure 5.21). Le paramètre b représente le salaire à l'embauche (0 année d'ancienneté) et la valeur de a représente l'augmentation annuelle de salaire, si ce modèle correspond à la réalité, bien sûr.

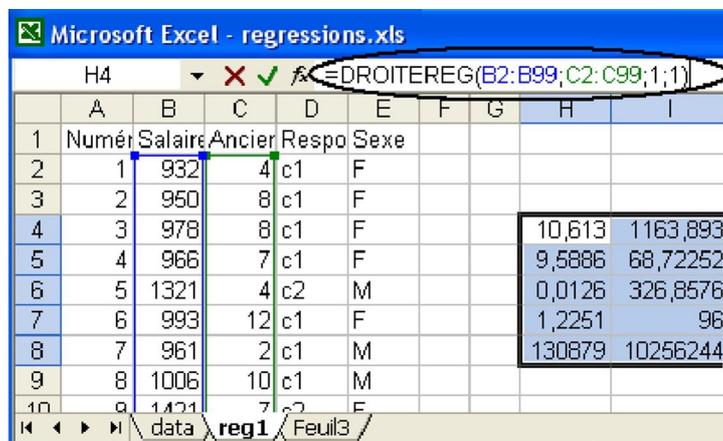


Figure 5.21 : modèle de régression Salaire=a.Ancienneté + b sur les données du fichier « salaires.txt »

Question 2

[Solution n°89 p 172]

Q2 : Utilisation du modèle M1 trouvé pour les prédictions et la visualisation

Si on remplace l'ancienneté par un nombre quelconque, on a une estimation du salaire correspondant à cette ancienneté selon le modèle. Écrire dans le tableur quelques valeurs d'ancienneté puis calculer le salaire estimé correspondant en utilisant les valeurs de a et de b obtenues (voir colonnes K et L de la figure 5.22). Faire le graphique permettant de comparer valeurs observées et valeurs estimées comme en figure 5.22.

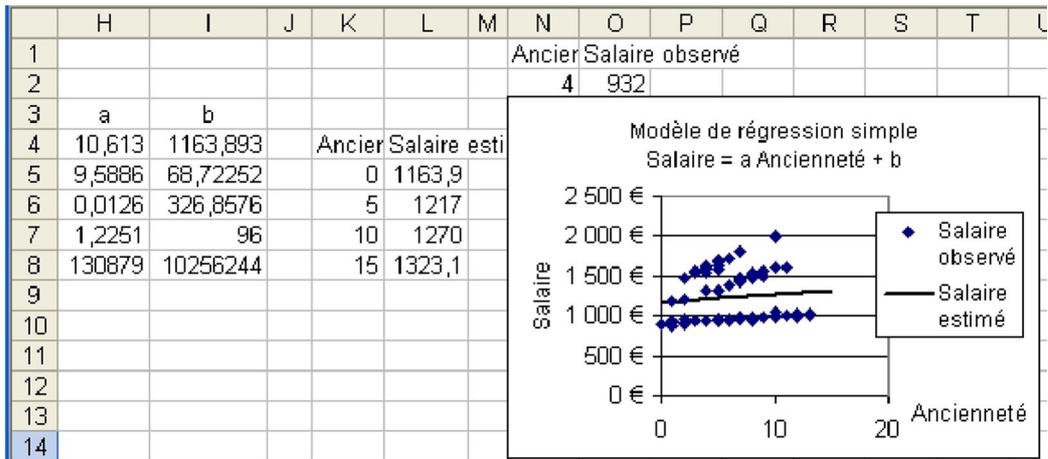


Figure 5.22 : Utilisation du modèle de régression pour dessiner la droite de régression

Question 3

[Solution n°90 p 172]

Q3 : Utilisation de l'outil « ajout d'une courbe de régression linéaire » du tableur

Le tableur permet de dessiner directement les courbes de régression sans faire de calcul. Certains tableurs donnent même l'équation de régression (comme Excel, voir figure 5.23 et OpenOffice3.1). On accède à ces tracés en sélectionnant la série de données, puis en choisissant « Graphique> Ajouter une courbe de tendance » pour Excel et « Insertion>Statistiques » sous OpenOffice. On a alors le choix entre plusieurs courbes de régression. Toutefois cela n'est possible que pour les modèles très simples, comportant une seule variable explicative. Contrôler que la droite que vous avez tracée et la droite de régression dessinée automatiquement par le tableur coïncident, et que l'équation indiquée, ainsi que le coefficient de détermination correspondent bien aux valeurs trouvées avec la fonction DROITEREG.

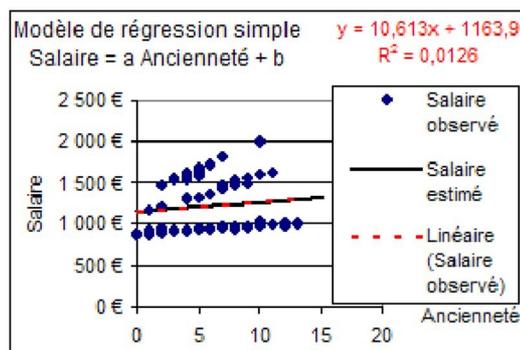


Figure 5.23 : En rouge la droite de régression linéaire figurant dans les options du graphique avec Excel 2003

Question 4

[Solution n°91 p 173]

Q4 : Les résidus au modèle M1

Ajouter une colonne intitulée « écarts » (colonne F de la figure 5.24) contenant pour chaque sujet la différence entre son salaire réel, figurant dans la colonne correspondante, et son salaire estimé d'après l'équation de régression (on utilisera dans la formule, les cellules où figurent a et b, et celles où figurent x et y). Vérifier

qu'on a bien pour moyenne de ces écarts approximativement 0, et pour écart-type, approximativement la valeur figurant dans le tableau à côté de R^2 (cellules K12 et K13 de la figure 5.24).

Si le modèle est bon, les erreurs suivent une loi normale d'espérance (moyenne) 0, et écart-type s , valeur qui a été rendue par le tableur à côté de R^2 dans la matrice de résultats. Dans ce cas, on ne doit trouver que 34% des valeurs au maximum à l'extérieur de la bande où y est compris entre $-s$ et s , et au moins 95% des valeurs entre $2s$ et $2s$, le tout étant réparti de façon équilibrée (symétriquement) autour de la droite $y=0$. Faire un graphique où figurent les points (en utilisant un nuage de points construit sur la seule colonne des écarts) et les droites $y=-2s$, $y=-s$, $y=0$, $y=s$ et $y=2s$, en insérant des nouvelles données au graphique, ces séries étant situées dans un tableau avec 2 valeurs de x (1 et 100), et pour chacune de ces valeurs, les 5 valeurs de y correspondant aux droites. On doit obtenir le graphique à droite de la figure 5.24.

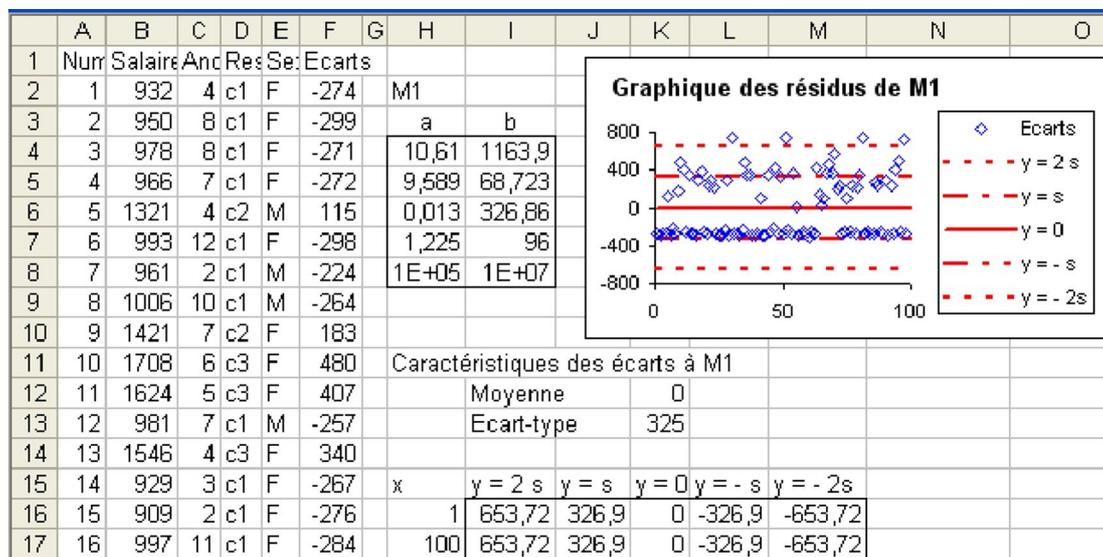


Figure 5.24 : Valeurs des résidus à gauche en colonne F, graphique à droite avec leur répartition en 6 zones devant contenir, en théorie, respectivement 2,5%, 14,5%, 33%, 33%, 14,5%, 2,5% des points.

On voit que les conditions de pourcentage paraissent vérifiées (par exemple, on a 4 points dans la première zone à partir du haut, ce qui n'est pas loin des 2,5% attendus), mais pas les conditions de symétrie (au dessus de la droite d'équation $y=0$, les points s'étalent en hauteur, alors qu'ils sont quasiment alignés sur une droite en dessous de l'axe des x). Pour repérer les individus, trions l'ensemble des colonnes selon le sexe, les femmes se trouvant de la ligne 2 à 58 et des hommes après. On ne voit pas de différences notables entre ces deux parties (figure 5.25 à gauche, F : triangle rouge, M : losange bleu). Faire la même chose en triant selon la responsabilité. On voit (figure 5.25 à gauche, c1 : losange marron, c2 : carré vert, c3 : croix bleue) que pour la responsabilité c1, tous les écarts sont systématiquement négatifs, ce qui signifie que le salaire réel est bien inférieur au salaire estimé, et c'est le contraire pour les 2 autres niveaux de responsabilité. Les conditions de normalité des résidus ne sont plus respectées. On en conclut que le modèle M1 est insuffisant, que la variable « responsabilité » doit intervenir.

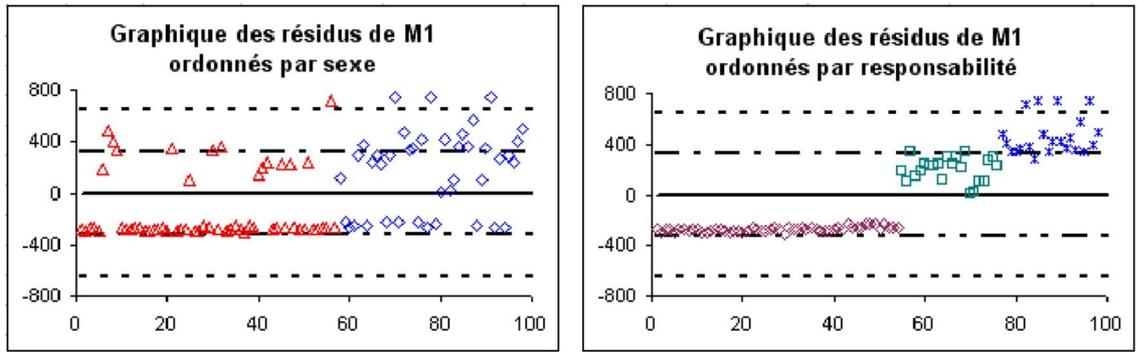


Figure 5.24 après avoir ordonné par sexe (à gauche), par responsabilité (à droite)

Modèle 2

La régression linéaire à variables qualitatives

Nous allons utiliser à nouveau la fonction DROITEREG pour estimer le salaire selon la responsabilité. Pour cela il convient de créer des variables numériques exprimant la valeur de la responsabilité. Quand la variable qualitative a p modalités, on crée $p-1$ nouvelles variables de la façon suivante : une des modalités est la modalité par défaut (par exemple $c1$), les autres sont des indicatrices des autres modalités ($c2$ et $c3$ ici). Il n'y a plus qu'à prendre les colonnes $c2$ et $c3$ comme variables X , et écrire le modèle de régression à 3 paramètres exprimant le salaire en fonction de la responsabilité (voir figure 5.30).

Question 5

[Solution n°92 p 175]

Réaliser cette régression comme indiqué dans la figure 5.30, écrire l'équation de régression correspondante, l'utiliser pour calculer le salaire estimé de 3 personnes, une de chaque responsabilité. Représenter graphiquement les résidus. Trouver la valeur de R^2 . Le modèle M2 est-il meilleur que le modèle M1 ? Pourquoi ? Montrer que toutefois l'ancienneté intervient dans les résidus, ce qui invite à créer un modèle M3 prenant en compte la responsabilité et l'ancienneté.

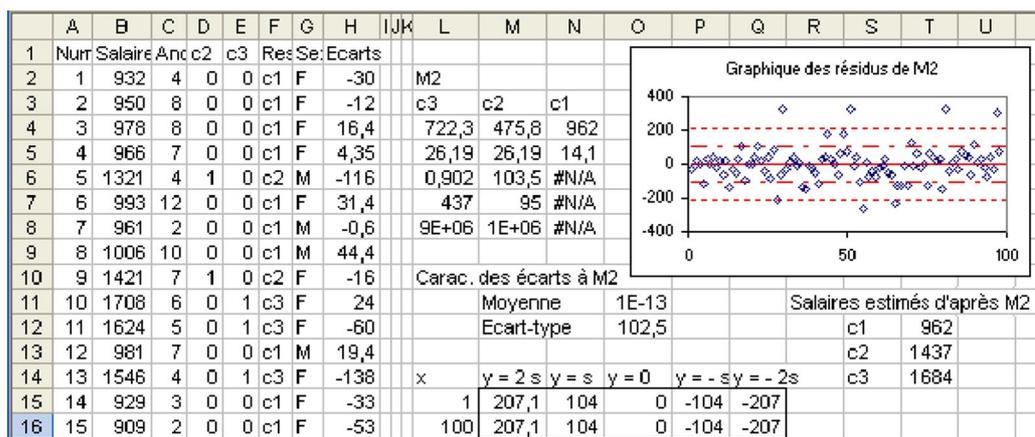


Figure 5.30 : Réalisation du modèle M2 exprimant le salaire en fonction de la responsabilité

Modèle M3

La régression linéaire multiple à deux variables mixtes

Question 6

[Solution n°93 p 177]

Q1. Faire une copie de la feuille de calcul (voir Figure 5.32), puis corriger afin de prendre dans les variables explicatives à la fois l'ancienneté et la responsabilité. Écrire l'équation du modèle M3 trouvé puis faire un tableau donnant les estimations de salaires pour 12 personnes, dont l'ancienneté prend les valeurs 0, 5, 10, 15 et les 3 responsabilités c1, c2 et c3.

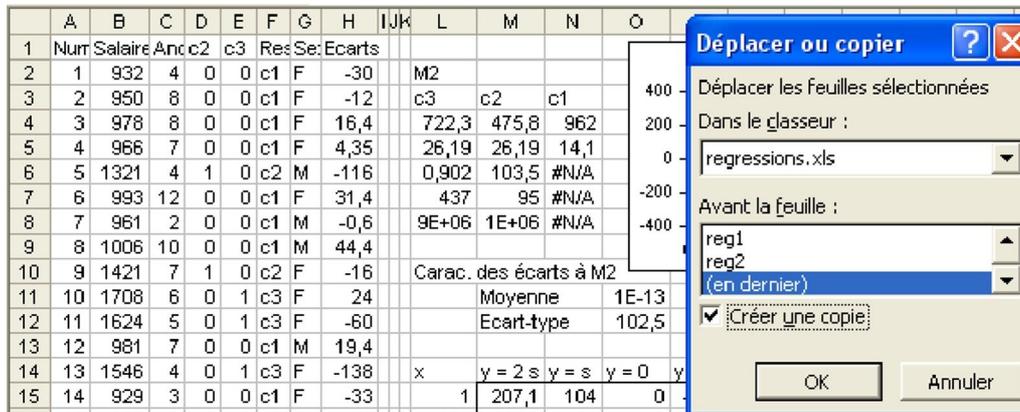


Figure 5.32 : copie de la feuille de calcul contenant le modèle M2 pour la corriger en un modèle M3 sous Excel

Question 7

[Solution n°94 p 177]

Q2. Le modèle M3 est-il bien meilleur que le modèle M2 ? Pour tester la significativité de l'augmentation de R^2 , c'est-à-dire le fait que cette augmentation n'est pas due au hasard, plusieurs méthodes existent dans le cas de modèles emboîtés, la plus ancienne étant de calculer la statistique $R^2_{\text{partiel}} / (1 - R^2_{\text{partiel}}) / ((q_2 - q_1) * (n - q_2))$ avec $R^2_{\text{partiel}} = (R^2_2 - R^2_1) / (1 - R^2_1)$, où R_1 et R_2 sont les coefficients de détermination respectifs de M2 et M3, qui doit suivre la loi de Fisher Snedecor (appelée loi.F dans les fonctions du tableur), à $(q_2 - q_1)$ et $(n - q_2)$ degrés de liberté, q_1 et q_2 étant le nombre de paramètres respectifs de M2 et M3, soit ici 3 et 4. Quelle est votre conclusion ? Examiner les résidus, seuls, puis groupés par sexe. A votre avis la variable sexe doit-elle être mise dans le modèle ?

Modèle 4

La régression linéaire multiple avec 2 variables qualitatives et une quantitative

Question 8

[Solution n°95 p 179]

Q1 : Ajouter la variable sexe dans le modèle. Pour cela insérer une colonne à côté des colonnes des autres variables, en prenant par exemple la variable F par défaut (1 si sexe= 'M', 0 si sexe='F'). Puis écrire le modèle. Et tester son apport par rapport à M3.

Modèle 5

La régression linéaire multiple avec interaction entre 2 variables

On repart de M3, et on essaie de l'améliorer par la prise en compte d'autres effets. On désire prendre en compte une augmentation annuelle de salaire différente selon les responsabilités. Cela s'appelle une interaction entre la variable responsabilité et la variable ancienneté. Pour cela on ajoute la variable produit des deux, donc deux colonnes supplémentaires, Ac2 et Ac3, qu'on obtient par produits respectifs des colonnes A et c2, A et c3), ce qui donne $S = a_1 A + a_2 c_2 + a_3 c_3 + a_4 A c_2 + a_5$

A $c_3 + b + e$ (on dit encore que le modèle est linéaire car il est une combinaison linéaire des paramètres à estimer a_1, a_2, \dots, a_5, b , mais certains auteurs ne disent plus régression linéaire dans ce cas, comme dans le cas où des variables sont élevées à une puissance quelconque).

Question 9

[Solution n°96 p 180]

Q1. Ajouter ces deux variables dans des colonnes et estimer le modèle, et tester son apport par rapport à M3. Donner l'équation générale du modèle, ainsi que les équations par responsabilité. Puis représenter graphiquement le nuage de points ainsi que les 3 droites de régression du salaire sur l'ancienneté pour chaque responsabilité (voir figure 5.38).

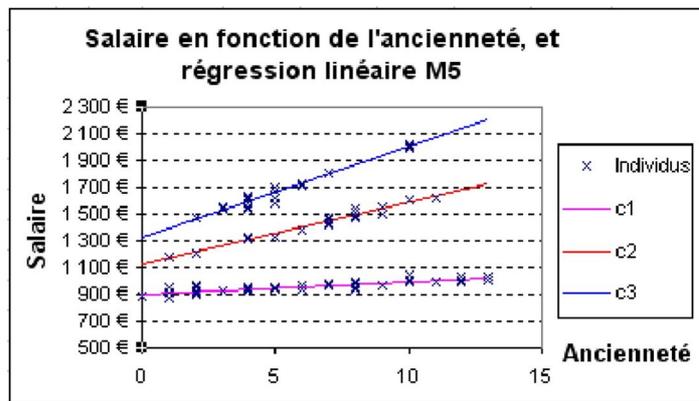


Figure 5.38 : Nuage de points et modèle de régression avec interaction sous la forme de 3 droites

Question 10

[Solution n°97 p 181]

Q2. Tester le modèle M5 par rapport au modèle M4. Ainsi que les paramètres du modèle. Représenter les résidus en fonction du sexe. On voit que cette fois la différence est bien visible entre les résidus des 2 sexes. Il faut donc essayer de nouveau de faire entrer le sexe dans le modèle. Écrire le modèle M6 et le tester.

F. Exercices TD pour la leçon 5

Le but de ce TD est d'écrire différents modèles d'estimation du salaire d'une entreprise par l'ancienneté, la responsabilité et le sexe en utilisant des « régressions linéaires » sur des données artificielles du fichier salaires.txt. Cet exercice correspond à une démarche habituelle de recherche de modèle statistique des données : les relations linéaires sont les plus simples et donc celles qu'on « essaie » en premier, si on ne dispose pas d'autres modèles fournis par des théories des sciences humaines. Suivant cette logique, on commence habituellement par le modèle le plus simple, c'est-à-dire avec une seule variable explicative, et on essaie de l'améliorer. Nous allons réaliser ces différentes étapes à l'aide du tableur. Nous utilisons pour cela principalement la fonction « DROITEREG » du tableur. Bien que nous utilisons les formules matricielles et la loi normale dans ce TP, leur maîtrise n'est pas indispensable car tous les rappels utiles sont faits dans le TP.

1. Modèle 1 : Le salaire en fonction de l'ancienneté

Question 1

[Solution n°98 p 182]

Q1 : Recherche du modèle M1 : $y=ax+b$

En utilisant la fonction DROITEREG du tableur, trouver les valeurs a et b correspondant à l'équation de régression Salaire = a Ancienneté + b + erreur minimisant les erreurs (voir figure 5.21). Le paramètre b représente le salaire à l'embauche (0 année d'ancienneté) et la valeur de a représente l'augmentation annuelle de salaire, si ce modèle correspond à la réalité, bien sûr.

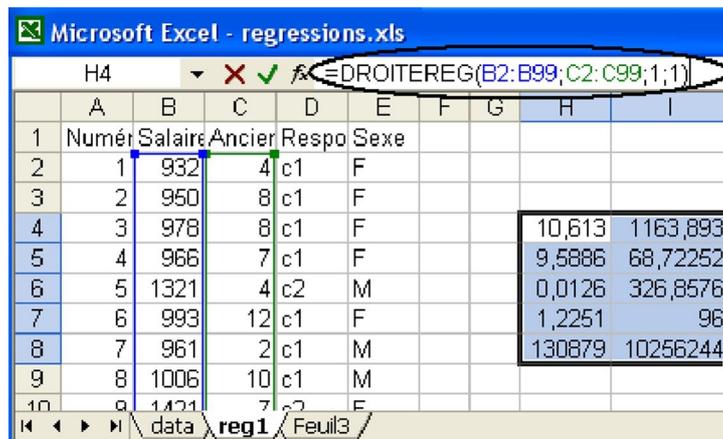


Figure 5.21 : modèle de régression Salaire= a .Ancienneté + b sur les données du fichier « salaires.txt »

Question 2

[Solution n°99 p 182]

Q2 : Utilisation du modèle M1 trouvé pour les prédictions et la visualisation

Si on remplace l'ancienneté par un nombre quelconque, on a une estimation du salaire correspondant à cette ancienneté selon le modèle. Écrire dans le tableur quelques valeurs d'ancienneté puis calculer le salaire estimé correspondant en utilisant les valeurs de a et de b obtenues (voir colonnes K et L de la figure 5.22). Faire le graphique permettant de comparer valeurs observées et valeurs estimées comme en figure 5.22.

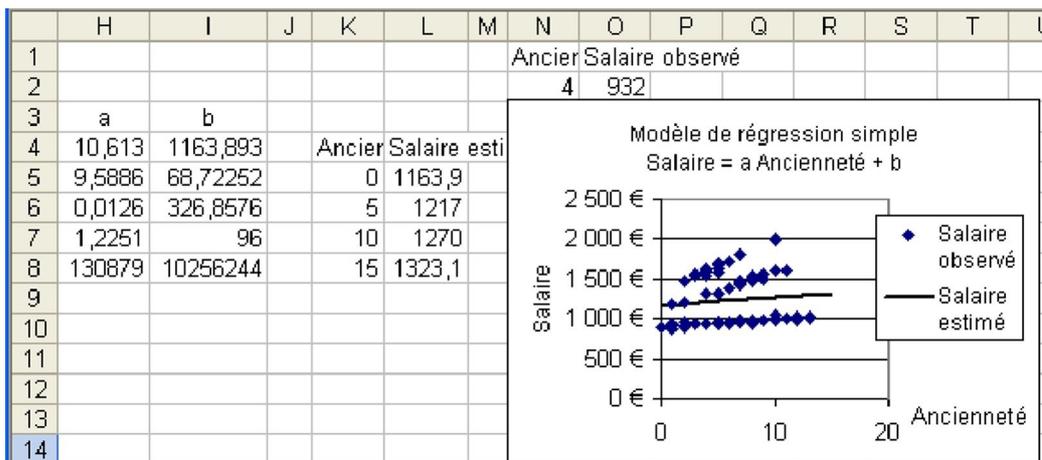


Figure 5.22 : Utilisation du modèle de régression pour dessiner la droite de régression

Question 3

[Solution n°100 p 183]

Q3 : Utilisation de l'outil « ajout d'une courbe de régression linéaire » du tableur

Le tableur permet de dessiner directement les courbes de régression sans faire de calcul. Certains tableurs donnent même l'équation de régression (comme Excel, voir figure 5.23 et OpenOffice3.1). On accède à ces tracés en sélectionnant la série de données, puis en choisissant « Graphique> Ajouter une courbe de tendance » pour Excel et « Insertion>Statistiques » sous OpenOffice. On a alors le choix entre plusieurs courbes de régression. Toutefois cela n'est possible que pour les modèles très simples, comportant une seule variable explicative. Contrôler que la droite que vous avez tracée et la droite de régression dessinée automatiquement par le tableur coïncident, et que l'équation indiquée, ainsi que le coefficient de détermination correspondent bien aux valeurs trouvées avec la fonction DROITEREG.

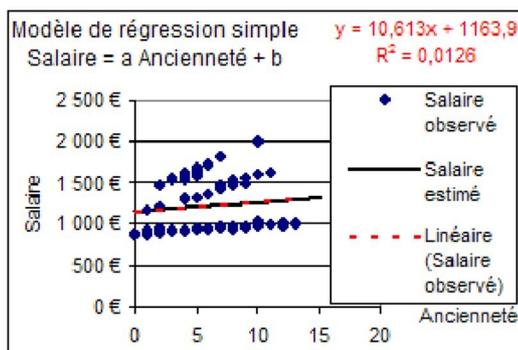


Figure 5.23 : En rouge la droite de régression linéaire figurant dans les options du graphique avec Excel 2003

Question 4

[Solution n°101 p 184]

Q4 : Les résidus au modèle M1

Ajouter une colonne intitulée « écarts » (colonne F de la figure 5.24) contenant pour chaque sujet la différence entre son salaire réel, figurant dans la colonne correspondante, et son salaire estimé d'après l'équation de régression (on utilisera dans la formule, les cellules où figurent a et b, et celles où figurent x et y). Vérifier

qu'on a bien pour moyenne de ces écarts approximativement 0, et pour écart-type, approximativement la valeur figurant dans le tableau à côté de R^2 (cellules K12 et K13 de la figure 5.24).

Si le modèle est bon, les erreurs suivent une loi normale d'espérance (moyenne) 0, et écart-type s , valeur qui a été rendue par le tableur à côté de R^2 dans la matrice de résultats. Dans ce cas, on ne doit trouver que 34% des valeurs au maximum à l'extérieur de la bande où y est compris entre $-s$ et s , et au moins 95% des valeurs entre $2s$ et $2s$, le tout étant réparti de façon équilibrée (symétriquement) autour de la droite $y=0$. Faire un graphique où figurent les points (en utilisant un nuage de points construit sur la seule colonne des écarts) et les droites $y=-2s$, $y=-s$, $y=0$, $y=s$ et $y=2s$, en insérant des nouvelles données au graphique, ces séries étant situées dans un tableau avec 2 valeurs de x (1 et 100), et pour chacune de ces valeurs, les 5 valeurs de y correspondant aux droites. On doit obtenir le graphique à droite de la figure 5.24.

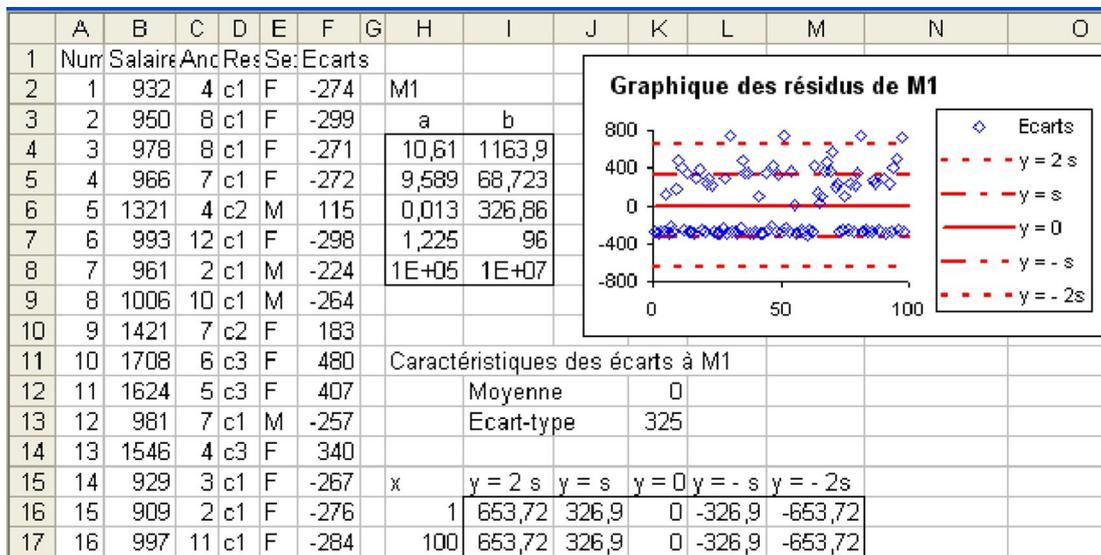


Figure 5.24 : Valeurs des résidus à gauche en colonne F, graphique à droite avec leur répartition en 6 zones devant contenir, en théorie, respectivement 2,5%, 14,5%, 33%, 33%, 14,5%, 2,5% des points.

On voit que les conditions de pourcentage paraissent vérifiées (par exemple, on a 4 points dans la première zone à partir du haut, ce qui n'est pas loin des 2,5% attendus), mais pas les conditions de symétrie (au dessus de la droite d'équation $y=0$, les points s'étalent en hauteur, alors qu'ils sont quasiment alignés sur une droite en dessous de l'axe des x). Pour repérer les individus, trions l'ensemble des colonnes selon le sexe, les femmes se trouvant de la ligne 2 à 58 et des hommes après. On ne voit pas de différences notables entre ces deux parties (figure 5.25 à gauche, F : triangle rouge, M : losange bleu). Faire la même chose en triant selon la responsabilité. On voit (figure 5.25 à gauche, c1 : losange marron, c2 : carré vert, c3 : croix bleue) que pour la responsabilité c1, tous les écarts sont systématiquement négatifs, ce qui signifie que le salaire réel est bien inférieur au salaire estimé, et c'est le contraire pour les 2 autres niveaux de responsabilité. Les conditions de normalité des résidus ne sont plus respectées. On en conclut que le modèle M1 est insuffisant, que la variable « responsabilité » doit intervenir.

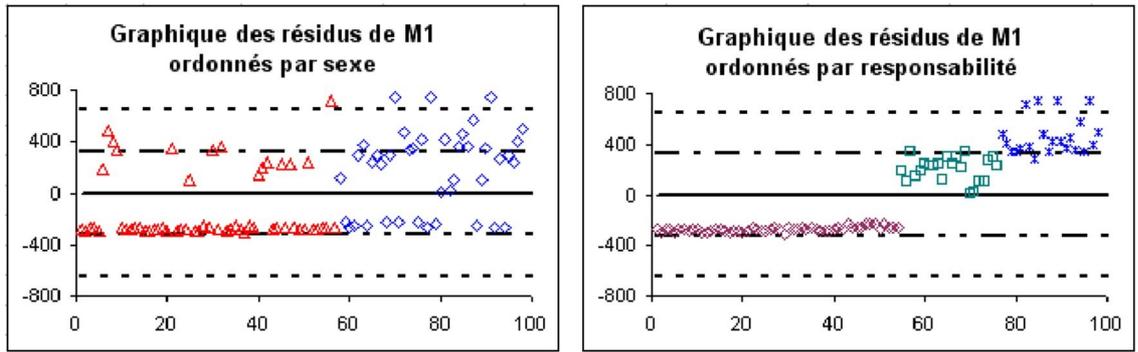


Figure 5.24 après avoir ordonné par sexe (à gauche), par responsabilité (à droite)

2. Modèle 2 : La régression linéaire à variables qualitatives

Nous allons utiliser à nouveau la fonction DROITEREG pour estimer le salaire selon la responsabilité. Pour cela il convient de créer des variables numériques exprimant la valeur de la responsabilité. Quand la variable qualitative a p modalités, on crée $p-1$ nouvelles variables de la façon suivante : une des modalités est la modalité par défaut (par exemple $c1$), les autres sont des indicatrices des autres modalités ($c2$ et $c3$ ici). Il n'y a plus qu'à prendre les colonnes $c2$ et $c3$ comme variables X , et écrire le modèle de régression à 3 paramètres exprimant le salaire en fonction de la responsabilité (voir figure 5.30).

Question

[Solution n°102 p 186]

Réaliser cette régression comme indiqué dans la figure 5.30, écrire l'équation de régression correspondante, l'utiliser pour calculer le salaire estimé de 3 personnes, une de chaque responsabilité. Représenter graphiquement les résidus. Trouver la valeur de R^2 . Le modèle M2 est-il meilleur que le modèle M1 ? Pourquoi ? Montrer que toutefois l'ancienneté intervient dans les résidus, ce qui invite à créer un modèle M3 prenant en compte la responsabilité et l'ancienneté.

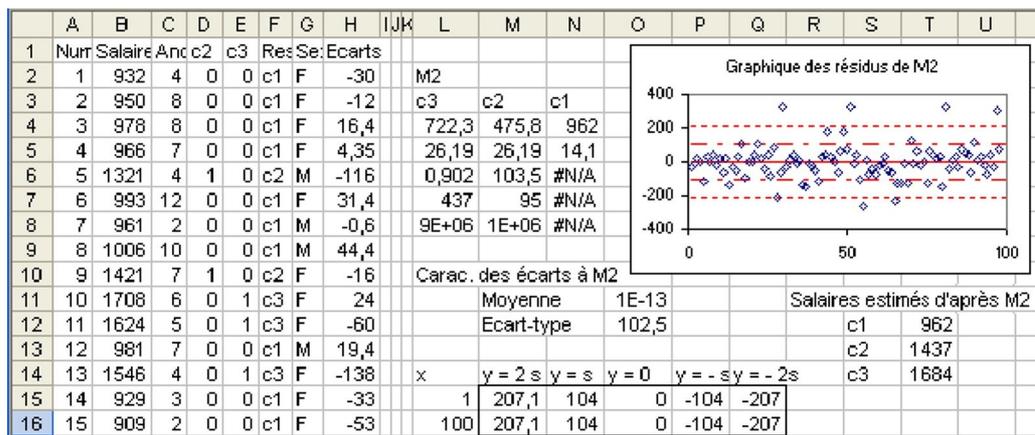


Figure 5.30 : Réalisation du modèle M2 exprimant le salaire en fonction de la responsabilité

3. Modèle 3 : La régression linéaire multiple à deux variables mixtes

Question 1

[Solution n°103 p 188]

Q1. Faire une copie de la feuille de calcul (voir Figure 5.32), puis corriger afin de prendre dans les variables explicatives à la fois l'ancienneté et la responsabilité. Écrire l'équation du modèle M3 trouvé puis faire un tableau donnant les estimations de salaires pour 12 personnes, dont l'ancienneté prend les valeurs 0, 5, 10, 15 et les 3 responsabilités c1, c2 et c3.

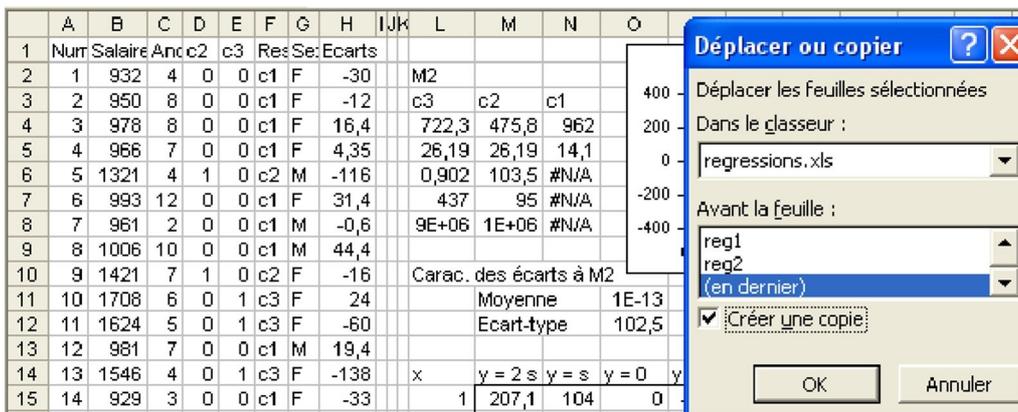


Figure 5.32 : copie de la feuille de calcul contenant le modèle M2 pour la corriger en un modèle M3 sous Excel

Question 2

[Solution n°104 p 188]

Q2. Le modèle M3 est-il bien meilleur que le modèle M2 ? Pour tester la significativité de l'augmentation de R^2 , c'est-à-dire le fait que cette augmentation n'est pas due au hasard, plusieurs méthodes existent dans le cas de modèles emboîtés, la plus ancienne étant de calculer la statistique $R^2_{\text{partiel}} / (1 - R^2_{\text{partiel}}) / ((q_2 - q_1) * (n - q_2))$ avec $R^2_{\text{partiel}} = (R^2_2 - R^2_1) / (1 - R^2_1)$, où R_1 et R_2 sont les coefficients de détermination respectifs de M2 et M3, qui doit suivre la loi de Fisher Snedecor (appelée loi.F dans les fonctions du tableur), à $(q_2 - q_1)$ et $(n - q_2)$ degrés de liberté, q_1 et q_2 étant le nombre de paramètres respectifs de M2 et M3, soit ici 3 et 4. Quelle est votre conclusion ? Examiner les résidus, seuls, puis groupés par sexe. A votre avis la variable sexe doit-elle être mise dans le modèle ?

4. Modèle 4 : La régression linéaire multiple avec 2 variables qualitatives et une quantitative

Question

[Solution n°105 p 190]

Q1 : Ajouter la variable sexe dans le modèle. Pour cela insérer une colonne à côté des colonnes des autres variables, en prenant par exemple la variable F par défaut (1 si sexe= 'M', 0 si sexe='F'). Puis écrire le modèle. Et tester son apport par rapport à M3.

5. Modèle 5 : La régression linéaire multiple avec interaction entre 2 variables

On repart de M3, et on essaie de l'améliorer par la prise en compte d'autres effets. On désire prendre en compte une augmentation annuelle de salaire différente selon les responsabilités. Cela s'appelle une interaction entre la variable responsabilité et la variable ancienneté. Pour cela on ajoute la variable produit des deux, donc deux colonnes supplémentaires, Ac2 et Ac3, qu'on obtient par produits respectifs des colonnes A et c2, A et c3), ce qui donne $S = a_1 A + a_2 c_2 + a_3 c_3 + a_4 A c_2 + a_5 A c_3 + b + e$ (on dit encore que le modèle est linéaire car il est une combinaison linéaire des paramètres à estimer a_1, a_2, \dots, a_5, b , mais certains auteurs ne disent plus régression linéaire dans ce cas, comme dans le cas où des variables sont élevées à une puissance quelconque).

Question 1

[Solution n°106 p 191]

Q1. Ajouter ces deux variables dans des colonnes et estimer le modèle, et tester son apport par rapport à M3. Donner l'équation générale du modèle, ainsi que les équations par responsabilité. Puis représenter graphiquement le nuage de points ainsi que les 3 droites de régression du salaire sur l'ancienneté pour chaque responsabilité (voir figure 5.38).

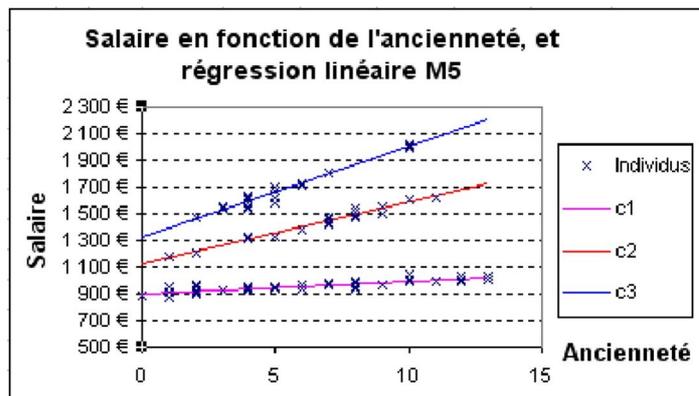


Figure 5.38 : Nuage de points et modèle de régression avec interaction sous la forme de 3 droites

Question 2

[Solution n°107 p 192]

Q2. Tester le modèle M5 par rapport au modèle M4. Ainsi que les paramètres du modèle. Représenter les résidus en fonction du sexe. On voit que cette fois la différence est bien visible entre les résidus des 2 sexes. Il faut donc essayer à nouveau de faire entrer le sexe dans le modèle. Écrire le modèle M6 et le tester.

G. Exercices appliqués à l'économie avec correction

1. Problème 4 : Calcul de coûts

Cet exercice aborde des points clés de la leçon 2 et de la leçon 5.

Cet exercice porte sur le calcul de coûts pour une entreprise qui dispose d'une capacité de production quotidienne de 20 unités, c'est-à-dire qu'elle peut fabriquer

chaque jour une quantité comprise entre 0 et 20 unités. Dans un premier temps on raisonnera en unités entières.

RAPPELS

a) Le coût moyen est égal au coût d'une unité produite, il est donc égal au coût total (cf. p 48) voir rappels sur le coût total plus loin) divisé par q.

$$CM(q) = CT(q) / q$$

Le coût moyen est rarement constant, il dépend de la quantité produite.

Exemple : si le coût total est égal à : $CT(q) = 10 + (15q)^{(1/6)}$,

la fonction de coût moyen est égale à : $CM(q) = [10 + (15q)^{(1/6)}] / q$ $CM(q) = (10/q) + (15^{(1/6)} q^{-5/6})$

b) Le coût marginal est égal au coût de la dernière unité produite.

Remarques :

Certains auteurs choisissent la prochaine unité produite. Lorsque la fonction de coûts est continue et dérivable, on utilisera la dérivée.

$$Cm(q) = CT(q) - CT(q-1)$$

Le coût marginal dépend aussi de la quantité produite.

$Cm(q) = [10 + (15q)^{(1/6)}] - [10 + (15(q-1))^{(1/6)}]$ (**attention** : les crochets doivent être remplacés par des parenthèses dans les formules tableur et ne pas oublier de rajouter le signe de la multiplication et de l'exposant)

LE COUT TOTAL

On appelle coût total (CT) la somme des coûts de tous les facteurs de production utilisés.

$$CT = \sum p_i f_i$$

Avec : p_i = prix d'une unité de facteur i et f_i = quantité de facteur i.

Toutes les dépenses de l'entreprise sont considérées comme servant à produire et sont donc des facteurs de production.
 Σ = somme pour tous les facteurs utilisés

Le coût total dépend de la quantité produite, en effet on utilisera certains facteurs de production en fonction de la quantité à produire. L'entreprise utilisera les facteurs de production de manière optimale, c'est-à-dire de manière à réduire ses coûts.

On appelle fonction de coût total la fonction qui associe le coût total minimal à une quantité produite (q) :

$$CT(q) = f(q)$$

Le coût total dépend donc de la quantité produite, mais aussi des prix des facteurs de production.

$$CT(q) = f(q, p_1, p_2, \dots)$$

Exemple, lorsque les prix de production sont fixes, on peut avoir la fonction de coût total suivante :

$$\text{Coût Total (en milliers d'Euros)} = 10 + (15q)^{1/6} + q^{1.5}$$

Cliquez sur la loupe pour agrandir la fiche

COÛTS FIXES ET COÛTS VARIABLES

Le coût total peut aussi s'exprimer par la somme de coûts fixes et de coûts variables.

$$CT(q) = CF + CV(q)$$

La distinction entre les coûts fixes et les coûts variables est principalement liée au terme envisagé.

Les Coûts Fixes (CF) sont des coûts indépendants des quantités produites : se sont les coûts des facteurs fixes à court terme, tels que les loyers, les charges locatives, les assurances, les frais liés aux remboursements d'emprunts, certains

salaires, ... Les coûts fixes ne dépendent pas des quantités produites.

Dans l'exemple les coûts fixes sont : $CF = 10\ 000$ Euros.

A long terme les coûts fixes deviennent variables : on peut déménager, modifier les contrats d'assurance, ...

Les Coûts Variables (CV) sont les coûts des facteurs variables, c'est-à-dire des facteurs de production dont les quantités varient avec la quantité produite. Les coûts variables sont donc fonction des quantités produites.

Dans l'exemple les coûts variables sont égaux à $CV(q) = (15q)^{(1/6)} + q^{1,5}$

Soit la fonction de coût total suivante :

Coût Total (en milliers d'Euros) = $10 + (15q)^{(1/6)} + q^{1,5}$

Question 1

[Solution n°21 p 159]

1) Calculez les coûts totaux, les coûts moyens et marginaux pour des quantités produites comprises entre 0 et 20.

Question 2

[Solution n°22 p 159]

2) Tracez le graphe correspondant, que constatez-vous ?

Question 3

[Solution n°23 p 159]

3) Comparez le coût moyen et le coût marginal.

Question 4

[Solution n°24 p 159]

4) Déterminez intuitivement la quantité de production qui maximise le profit, pour un prix de vente égal à 5,4€. Calculez le profit correspondant.

Question 5

[Solution n°25 p 159]

5) Même question avec un prix de vente égal à 5,8€

6) Même question pour un prix de vente égal à 3€.

Question 6

[Solution n°26 p 160]

7) Que pensez-vous des résultats obtenus ?

2. Problème 14 : Offre et demande : équilibre du marché

Soit un marché de concurrence parfaite comportant un grand nombre d'entreprises. Dans ce marché les quantités échangées et les prix sont déterminés par des courbes d'offre et de demande, les prix pouvant varier entre 50 et 200 € et les quantités entre 100 et 500.

La fonction déterminant les quantités demandées est de la forme

$$D = 800 - 45P^{0,5}$$

La fonction déterminant la quantité offerte est de la forme

$$O = 30 \cdot P^{0,6} - 210$$

Question 1

[Solution n°108 p 193]

1) Tracez les courbes d'offre et de demande pour des prix compris entre 50 et 200.

Question 2

[Solution n°109 p 193]

2) Déterminez le point d'équilibre.

Question 3

[Solution n°110 p 193]

3) Le prix du marché est actuellement de 155 €, quelle est la situation du marché ?

Question 4

[Solution n°111 p 193]

4) Expliquez comment le marché va revenir à l'équilibre.

Question 5

[Solution n°112 p 193]

5) Sachant que sur ce marché, les quantités offertes ou demandées se corrigent chaque semaine, au bout de combien de semaines parviendra-t-on à l'équilibre (moins de 2 d'écart). Déterminez les étapes intermédiaires.

3. Problème 15 : Données chronologiques

Chiffre d'affaires d'un hypermarché (suite de l'exercice 2 leçon 4)

CA	2006	2007	2008	2009
janvier	22	22	23	23
février	10	10	11	10
mars	11	11	12	12
avril	12	12	13	13
mai	17	19	18	21
juin	20	20	21	21
juillet	17	17	18	20
août	10	10	11	12
septembre	21	20	21	22
octobre	17	17	19	19
novembre	15	15	16	16
décembre	28	30	30	33

Tableau 17 : Données : chiffre d'affaires

Nombre de jours ouvrables

Jours ouvrables	2006	2007	2008	2009
janvier	22	21	22	21
février	19	20	19	20
mars	25	23	25	23
avril	23	23	23	23
mai	17	20	17	20
juin	20	23	20	23
juillet	22	21	22	21
août	21	22	21	22
septembre	23	23	23	23
octobre	23	22	23	24
novembre	24	22	25	22
décembre	23	25	23	24

Tableau 18 : Nombre de jours ouvrables

Dans le traitement de données chronologiques, le temps est une variable discrète qui prend des valeurs de 1 à n.

Question 1

1) Représentez graphiquement la série de chiffre d'affaires

Question 2

[Solution n°113 p 193]

2) Le chef d'entreprise calcule la tendance sur les données brutes, à l'aide d'un ajustement linéaire. Que pensez-vous de ce calcul ? Pour argumenter il conviendra de calculer le coefficient de corrélation linéaire.

Question 3

[Solution n°114 p 193]

3) Le responsable du magasin envisage de faire des prévisions mensuelles, il souhaite commencer par calculer une moyenne mobile.

a. Quel ordre p doit-il utiliser ? pourquoi ?

b. Le responsable souhaite des moyennes mobiles centrées, quel est l'intérêt d'utiliser des moyennes mobiles centrées ? quelle formule utiliser ? pourquoi ?

c. Calculez la série des moyennes mobiles centrées d'ordre 12

Question 4

[Solution n°115 p 194]

4) Tracez le graphe représentant les chiffres d'affaires et les moyennes mobiles. Qu'en pesez-vous ?

Question 5

[Solution n°116 p 194]

5) Le chef d'entreprise souhaite calculer une tendance à partir des moyennes mobiles, comment procéder ?

Question 6

[Solution n°117 p 194]

6) Calculez le coefficient de corrélation linéaire de ce deuxième ajustement et comparez-le au premier, que pouvez-vous en tirer ?

Question 7

[Solution n°118 p 194]

7) Dans l'étape suivante il calcule les coefficients saisonniers, pouvez-vous l'aider ?

Question 8

[Solution n°119 p 194]

8) Désormais le responsable va pouvoir faire des prévisions mensuelles de chiffre d'affaires pour le premier trimestre de l'année 2010.

a. Estimez le chiffre d'affaires de janvier, de février et de Mars 2010.

Question 9

[Solution n°120 p 194]

8) b. On envisage de prévoir avec ce modèle, les estimations de chiffre d'affaires de janvier 2015. Calculez le résultat. Cette estimation vous semble-t-elle fiable ?

4. Problème 16 : Enquête premier salaire

Suite de l'exercice 3 de la leçon 3.

On se propose d'étudier le premier salaire de jeunes sortis, depuis 18 mois, d'un BTS en apprentissages, à partir d'un extrait d'une enquête (voir enquête premier salaire feuille 2).

Question 1

[Solution n°121 p 194]

1) Proposez des représentations graphiques en colonne et en anneaux pour montrer les différences constatées entre hommes et femmes salariés à plein temps ou à temps partiel supérieur à un mi-temps. Commentez.

Question 2

[Solution n°122 p 194]

2) Représentez la répartition des salaires pour tous les secteurs d'activité, et calculez les salaires moyens par secteur. Commentez.

Solution des exercices de TD



> Solution n°1 (exercice p. 21)



Attention

Sous certaines versions de tableurs (comme Excel 2003), il est nécessaire d'écrire les 2 premiers termes de la suite arithmétique et de les sélectionner tous les 2 avant de procéder à un étirage.

> Solution n°2 (exercice p. 21)



Attention

Le collage spécial diffère légèrement sous Excel. Les catégories "chaînes de caractères" et "nombre" ne se retrouvent pas exactement de la même façon.

> Solution n°3 (exercice p. 21)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°4 (exercice p. 21)

L'indice des prix doit être calculé en neutralisant l'effet quantités, c'est-à-dire en choisissant de pondérer les prix les différentes années par les mêmes quantités, celles de la période d'origine (0) ou celles de la période d'arrivée (t) ou encore par des quantités relatives à une autre période.

> Solution n°5 (exercice p. 21)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°6 (exercice p. 21)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°7 (exercice p. 22)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°8 (exercice p. 22)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°9 (exercice p. 22)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°10 (exercice p. 22)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°11 (exercice p. 22)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°12 (exercice p. 34)

Le symbole \$ placé devant le nombre 3 fixe le numéro de ligne à 3, donc lors de la recopie le numéro de ligne ne change pas. Comme il n'y a pas de \$ devant la lettre B indiquant la colonne, si on recopie la formule dans des cellules de la même colonne, on obtient une formule identique. Cela explique que les 5 lignes comportent des nombres identiques. Par contre quand on se déplace d'une colonne, la formule se réfère à la colonne d'après.

- dans la colonne C, la formule reste « $=2*B\$3+1$ », ce qui donne $2*2,6+1$, soit 5,6.
- dans la colonne D, la formule devient « $=2*C\$3+1$ », ce qui donne $2*10+1$, soit 21.
- dans la colonne E, la formule devient « $=2*D\$3+1$ », ce qui donne $2*0+1$, soit 1.

	A	B	C	D	E	F
1						
2						
3		2,3	10			
4		4	2,5			
5		5,3	5,6	21	1	
6			5,6	21	1	
7			5,6	21	1	
8			5,6	21	1	
9			5,6	21	1	
10						

Figure 2.18 : L'effet de la recopie de la cellule C5 vers les cellules C5 à E9

> Solution n°13 (exercice p. 34)

	A	B	C	D	E
2					
3		2.3	10.00		
4		4	2.50		
5		14.80	-3.00		
6					
7					
8					
9					

Figure 2.34 : Écriture d'une formule dans la cellule B5 pour recopie dans les cellules grisées

1) On cherche d'abord les formules.

On peut partir de l'autre forme de la formule « $=\$B\$3+C3+\$C4$ » de B5, qui est « $=L3C2+L(-2)C(+1)+L(-1)C$ » et se recopie ainsi dans les 5 autres cellules. Et on recalcule les nouvelles formes dans chaque cellule.

On peut également procéder par modification de la forme actuelle de la formule, en faisant un seul changement à chaque fois :

- De B5 à B6, on a augmenté le numéro de ligne de 1 unité, donc les numéros de lignes n'ayant pas de \$ sont augmentés de 1. Comme on n'a pas changé de colonne, les numéros de colonnes ne changent pas et « $=\$B\$3+C3+\$C4$ » devient « $=\$B\$3+C4+\$C5$ »
- De B5 à C5, on a augmenté le numéro de colonne de 1 unité, donc les numéros de colonnes n'ayant pas de \$ sont augmentés de 1. Comme on n'a pas changé de ligne, les numéros de lignes ne changent pas et « $=\$B\$3+C3+\$C4$ » devient « $=\$B\$3+D3+\$C4$ »
- De B6 à C6, on a augmenté seulement le numéro de colonne de 1 unité, donc « $=\$B\$3+C4+\$C5$ » devient « $=\$B\$3+D4+\$C5$ »
- De C6 à D6, on a augmenté seulement le numéro de ligne de 1 unité, donc « $=\$B\$3+D4+\$C5$ » devient « $=\$B\$3+D5+\$C6$ »
- De D6 à E6, on a augmenté seulement le numéro de colonne de 1 unité, donc « $=\$B\$3+D5+\$C6$ » devient « $=\$B\$3+E5+\$C6$ »

On obtient ainsi le résultat de la figure 2.35

	A	B	C	D
1				
2				
3		2.3	10	
4		4	2.5	
5		$=\$B\$3+C3+\$C4$	$=\$B\$3+D3+\$C4$	
6		$=\$B\$3+C4+\$C5$	$=\$B\$3+D4+\$C5$	
7			$=\$B\$3+D5+\$C6$	$=\$B\$3+E5+\$C6$
8				
9				

Figure 2.35 : Formules obtenues dans les cellules grisées par copier-coller de B5

2) On cherche maintenant les valeurs qui s'affichent.

Pour cela on choisit un sens de parcours des cellules car les mises à jour vont se

faire en cascade. Selon les divers sens de parcours possibles, les mises à jour intermédiaires peuvent différer, mais le résultat final est le même.

Passage 1 : 6 cellules peuvent changer de valeur, ce sont celles qui contiennent des formules.

- B5 : valeur initiale 5,3 ; formule $=B3+C3+C4$, nouvelle valeur $2,3+10+2,5=14,8$. Ne changera plus car B3, C3, C4 ne contiennent pas de formule, donc ne peuvent pas changer.
- B6 : aucune valeur initiale, formule $=B3+C4+C5$, nouvelle valeur $2,3+2,5+(-3)=1,8$. Pourra encore changer car C5 contient une formule.
- C5 : valeur initiale -3 ; formule $=B3+D3+C4$, nouvelle valeur $2,3+0+2,5=4,8$. Ne changera plus car B3, D3, C4 ne contiennent pas de formule.
- C6 : aucune valeur initiale, formule $=B3+D4+C5$, nouvelle valeur $2,3+0+4,8=7,1$. Ne changera plus car B3 et D4 ne contiennent pas de formule et C5 qui en contient une vient de voir passer sa valeur à 4,8 qui ne changera plus.
- C7 : aucune valeur initiale, formule $=B3+D5+C6$, nouvelle valeur $2,3+0+7,1=9,4$. Ne changera plus car B3 et D5 ne contiennent pas de formule et C6 qui en contient une vient de voir passer sa valeur à 7,1 qui ne changera plus.
- D7 : aucune valeur initiale, formule $=B3+E5+C6$, nouvelle valeur $2,3+0+7,1=9,4$. Ne changera plus car B3 et E5 ne contiennent pas de formule et C6 qui en contient une vient de voir passer sa valeur à 7,1 qui ne changera plus.

Passage 2 : 1 cellule peut encore changer de valeur, c'est la cellule B6

- B6 : valeur initiale 1,8, formule $=B3+C4+C5$, nouvelle valeur $2,3+2,5+4,8=9,6$. Ne changera plus car B3 et C4 ne contiennent pas de formule et C5 ne changera plus de valeur.

	A	B	C	D	E
3		2.3	10.00		
4		4	2.50		
5		14.80	4.80		
6		9.60	7.10		
7			9.40	9.40	
8					

Figure 2.36 : Valeurs affichées obtenues dans les cellules grisées par copier-coller de B5

Sur cet exemple très simple, les modifications de valeurs de cellules s'arrêtent là, il n'y a pas de passage supplémentaire. On obtient l'affichage de la figure 2.36.

Quand les cellules contenant des formules sont plus nombreuses, les passages de mise à jour se succèdent, ce qui peut rendre l'affichage instable pendant un temps plus ou moins long.

> Solution n°14 (exercice p. 44)

pour recopier correctement vers F5, donc une ligne plus bas, il faut fixer la ligne 2 qui contient les valeurs des coefficients afin qu'elle ne devienne pas la ligne 3, mais il ne faut pas fixer la ligne 4 afin qu'elle devienne la ligne 5 qui contient les notes de l'étudiant suivant dont on désire faire la moyenne en F5. Pour pouvoir la recopier vers H4, soit 2 colonnes plus loin, il ne faut pas fixer les colonnes afin que ce soit

bien la moyenne des notes Note3 et Note4 avec les coefficients Coef3 et Coef4 qui soit faite. Ainsi la formule de F4 « $= (B2*B4+C2*C4)/(B2+C2)$ » devient « $= (B\$2*B4+C\$2*C4)/(B\$2+C\$2)$ »

> **Solution n°15** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°16** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°17** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°18** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°19** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°20** (exercice p. 47)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°21** (exercice p. 49,151)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°22** (exercice p. 49,151)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°23** (exercice p. 49,151)

Trois cas de figure peuvent se présenter :

- Le coût marginal est supérieur au coût moyen: dans ce cas le coût moyen est croissant, en effet la dernière unité produite coûtait plus que le coût moyen, elle a donc fait monter la moyenne.
- Le coût marginal est inférieur au coût moyen : dans ce cas le coût moyen est décroissant, en effet la dernière unité produite coûtait moins que le coût moyen, elle a donc fait baisser la moyenne.
- Le coût marginal est égal au coût moyen, dans ce cas la moyenne à atteint un maximum (ou un minimum ou un point d'inflexion) .

> **Solution n°24** (exercice p. 49,151)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°25** (exercice p. 49,151)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°26 (exercice p. 49,151)

Le plus haut profit est réalisé pour un prix de vente de 5,8 € avec une production plus importante (quantité = 15).

> Solution n°27 (exercice p. 49)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°28 (exercice p. 51)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°29 (exercice p. 51)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°30 (exercice p. 51)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°31 (exercice p. 51)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°32 (exercice p. 56)

Les séparateurs exigés dans l'écriture des formules sont les parenthèses et les points-virgules. Il est donc mieux de pas en mettre car il peuvent produire des erreurs dans certains cas : notamment écrire le nombre 1 300 (avec un espace pour les milliers) peut conduire à des erreurs dans certaines versions de tableurs qui l'interpréteront comme un texte et non comme un nombre.

> Solution n°33 (exercice p. 56)

L'aide propose des formules écrites en majuscules, mais les formules ne sont pas « sensibles à la casse », ce qui signifie qu'on peut écrire chacune des lettres les constituant en minuscules comme en majuscules. Conseil : les écrire en minuscules lors de la saisie au clavier, elles seront réécrites en majuscules lorsqu'elles auront été validées par appui sur la touche « entrée ». En cas d'erreur, cela permettra de repérer les fautes de frappe, la partie correspondante de la formule restant en minuscules.

> Solution n°34 (exercice p. 56)

On peut ajouter des fonctions dans un classeur, ou même dans le tableur. On peut les créer soi-même, par programmation (en utilisant le « VBA » et les « macros » qui ne sont pas enseignés dans ce cours) ou les importer (par exemple votre collègue peut vous envoyer son fichier qui les contient). Elles ne sont utilisables que si vous autorisez leur exécution, en enlevant la protection par défaut de votre classeur avant de le sauvegarder. Attention de n'importer que des compléments de toute confiance : les options de protection ont été mises par défaut dans les classeurs afin d'éviter la propagation de virus introduits de cette façon dans le

tableur.

> Solution n°35 (exercice p. 67)

On écrit en C3 la formule « =si(B3>=0 ;racine(B3) ; "Le nombre doit être positif ") »

> Solution n°36 (exercice p. 67)

On écrit en D3 la formule « =si(B3>0;B3*log(B3) ; si(B3=0;0 ;"Le nombre ne doit pas être négatif ")) »

> Solution n°37 (exercice p. 67)

=SI(C4<1;"pas du tout"; SI(C4<2;"très peu"; SI(C4<3; " un peu "; SI(C4<4; " beaucoup"; " énormément")))). Cette formule est recopiée par un copier-coller, dans toute la colonne D. Si on représente cette formule par un organigramme, avec les valeurs dans les rectangles et les conditions dans des losanges, On peut voir dans la figure le calcul de la formule quand C4 contient 3 (en rouge) et quand il contient 1 (en bleu)

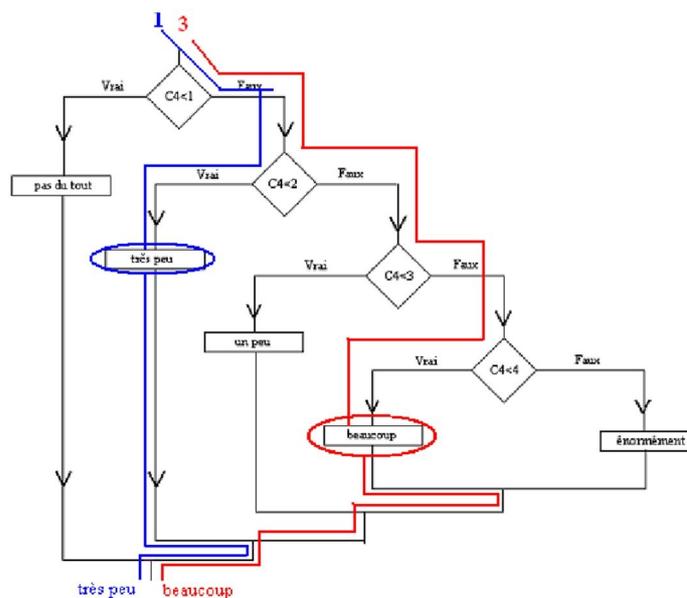


Figure 3.15 : Utilisation d'un organigramme pour trouver les valeurs d'une formule avec SI imbriqués

> Solution n°38 (exercice p. 68)

Dans D8, =NB.SI(D4:D6;5), dans E8, =NB.SI(D4:D6;"<5")-NB.SI(D4:D6;"<2"), on les recopie jusqu'à O8 et O9.

> Solution n°39 (exercice p. 68)

Dans U8, qui se copie de U8 à X8 : =NB.SI(\$D\$8:\$O\$8;U8)

> Solution n°40 (exercice p. 69)

Formule en R8 : =SOMME.SI(D\$3:O\$3;Q8;D8:O8)/NB.SI(D\$3:O\$3;Q8), on la recopie en R9.



Conseil

Pour avoir plus de transparence au niveau du critère, on peut l'écrire dans une cellule, puis remplacer le critère par l'adresse de la cellule (voir fig.3.18). Il peut alors se calculer à l'aide d'une formule. Par exemple si on écrit 4 dans la cellule C20, puis = "<"&C20 dans la cellule D20, on peut écrire dans la cellule E20, =NB.SI(D4:O6;D20)

	C	D	E
17	1	<1	0
18	2	<2	5
19	3	<3	18
20	4	<4	24
21	5	<5	29
22	6	<6	33

Figure 3.18 : Écriture du critère dans la colonne D, pour l'utiliser dans le NB.SI de la colonne E

> Solution n°41 (exercice p. 69)

Dans la cellule E4, on a calculé le nombre d'actions vendues à Dupont Hortense, sachant qu'on ne vend des actions que si la personne est d'accord (D4="oui") et si elle n'a pas dépassé le débit autorisé (B4>-C4). Dans ce cas, on obtient le nombre d'actions vendues à cette personne en divisant la somme du crédit et du débit par le prix de vente unitaire de l'action et on arrondit au nombre entier immédiatement inférieur le résultat de la division. Sinon, on lui en vend 0.

=SI(ET(B4>-C4;D4="oui");ARRONDI.INF((B4+C4)/B\$1;0);0)

Dans la cellule D10, on a compté le nombre de cellules de la plage D4:D8 qui contenaient la valeur "oui"

=NB.SI(D4:D8;"oui")

Dans la cellule D11, on a additionné parmi les crédits (B4:B8), ceux correspondant aux personnes ayant donné leur accord, donc dont la cellule correspondante de la plage D4:D8 est à "oui" . (1800+(-200)+(-200)=1400).

=SOMME.SI(D4:D8;"oui";B4:B8)

Dans la cellule D11, on a compté le nombre de cellules de la plage E4:E8 qui contenaient une valeur ">0"

=NB.SI(E4:E8;">0")

> Solution n°42 (exercice p. 74)

1. Moyennes pondérées :

On écrit dans la cellule G2 « =somme(C2 :F2) », dans la cellule G3 « =SOMMEPROD(C\$2:F\$2;C3:F3) », dans la cellule H2 « =G3/G\$2 », que l'on recopie automatiquement dans les deux cellules du dessous.

2. Distances euclidiennes :

On écrit dans la cellule K3 « =SOMME.XMY2(\$C\$3:\$F\$3;\$C3:\$F3) », que l'on recopie vers le bas, ainsi que dans la cellule L3 en la transformant ensuite en la formule =SOMME.XMY2(\$C\$4:\$F\$4;\$C3:\$F3) » et de la même façon dans M3 « =SOMME.XMY2(\$C\$5:\$F\$5;\$C3:\$F3) ». Formules qu'on copie dans les cellules du dessous.

3. Intitulés du tableau des distances euclidiennes :

Après avoir sélectionné les 3 cellules J3 à J5, on écrit la formule matricielle « =B3 :B5 », et après sélection des cellules K3 à M3 la formule « =transpose(B3 :B5) ». Ces deux formules sont validées par l'appui simultané des touches CTL MAJ ENTREE.

> Solution n°43 (exercice p. 74)

On dispose les valeurs de la matrice A de telle façon que chaque colonne corresponde à une variable, et chaque ligne à une équation, la valeur dans la cellule à l'intersection de la ligne et de la colonne étant le coefficient de la variable dans l'équation. Pour faciliter les calculs, on indique en intitulé de colonne la variable qu'elle représente. Puis on écrit la matrice B en colonne. Ces éléments sont dans la figure 3.24.

	B	C	D	E	F	G	H	I	J	K	L	M
2	A	x	y	z	B			X		AX-B		
3		2	4	0		2		-0,8		0		
4		3	0	1		-1		0,9		0		
5		1	-2	4		3		1,4		0		

Figure 3.24

Pour calculer X, on sélectionne trois cellules (ici I3 à I5) et on tape la formule correspondant à $A \cdot X = B$ dans laquelle le produit s'écrit « produitmat » et l'inverse s'écrit « inversemat », comme on peut le voir dans la ligne d'édition de la figure 3.24. On obtient ainsi $x = -0,8$, $y = 0,9$, $z = 1,4$. Pour la vérification, on écrit la formule $AX - B$ (ici =PRODUITMAT(C3:E5;I3:I5)-G3:G5), qui produit une colonne de zéros comme attendu.



Remarque

Les calculs faits avec les systèmes informatiques sont de plus en plus précis, mais leur précision n'est pas « infinie ». Avec les premières calculatrices, le produit de $1/3$ par 3 produisait plus souvent 0,9999... que 1. On retrouve dans le tableau ce type de problème, particulièrement quand on fait des opérations successives sur des matrices, comportant par exemple des inverses de matrices. Ainsi la vérification de l'exercice-TD 2 pourrait déboucher sur un affichage de valeurs différentes de 0, de l'ordre de 10^{-16} par exemple.

> Solution n°44 (exercice p. 75)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°45 (exercice p. 75)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°46** (*exercice p. 75*)

Médiane = 1400 € (200 au dessus, France et 200 en dessous, Chine.)

> **Solution n°47** (*exercice p. 75*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°48** (*exercice p. 75*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°49** (*exercice p. 75*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°50** (*exercice p. 75*)

C'est l'effet de structure : les salaires ont augmenté dans tous les pays, mais les effectifs ont baissé en France, alors qu'ils augmentaient en Chine où les salaires sont bas. Le nombre de bas salaires est plus important.

> **Solution n°51** (*exercice p. 76*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°52** (*exercice p. 76*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°53** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°54** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°55** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°56** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°57** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°58** (*exercice p. 77*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°59** (*exercice p. 78*)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°60 (exercice p. 78)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°61 (exercice p. 78)

Le bénéfice risque de diminuer.

> Solution n°62 (exercice p. 78)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°63 (exercice p. 78)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°64 (exercice p. 79)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°65 (exercice p. 79)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°66 (exercice p. 102)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°67 (exercice p. 104)

Dans la figure 4.25, on a écrit les paramètres du placement en ligne 27, avec la formule de la somme obtenue dans la cellule H27, qui dépend des 3 cellules précédentes. Puis on a copier les 4 cellules de E27 à H27 vers les lignes « 0, 31 et 32. On a ensuite appelé l'outil « valeur cible » à trois reprises : en sélectionnant la cellule H30 (resp. H31, H32) et en choisissant comme cellule variable la cellule E30 (resp. F31, G32).

	D	E	F	G	H
25					
26		capital initial	annee	taux	capital
27		500	10	2,30%	627,66273
28					
29					
30		557,6243152	10	2,30%	700
31		500	14,796825	2,30%	699,999981
32		500	10	3,42%	699,999987

Figure 4.25 : Recherche des paramètres d'un placement permettant d'obtenir 700 € à terme.

On constate que le nombre d'années en cellule F31 n'est pas très réaliste. On devrait avoir un nombre entier d'années. Aussi on transforme la formule de H31 en conséquence (voir Fig. 4.26 la formule dans la ligne d'édition). Bien sûr la somme de 700 n'est pas atteinte, mais la somme de 703,24... obtenue pour 15 ans (bien que la cellule F31 ne contienne pas la valeur 15, c'est bien 15 la solution correspondant à une somme finale de 703, 24..., comme on peut s'en assurer en

observant que la valeur de H31 ne change pas en tapant 15 en F31).

	D	E	F	G	H	I	J
25							
26		capital initial	annee	taux	capital		
27		500	10	2,30%	627,66273		
28							
29							
30		557,6243152	10	2,30%	700		
31		500	15,0927615	2,30%	703,24153		
32		500	10	3,42%	699,999987		

État de la recherche

Recherche sur la cellule H31
n'a pas trouvé de solution.

Valeur cible : 700
Valeur actuelle : 703,2415298

Figure 4.26 : Recherche des paramètres d'un placement en prenant un nombre entier d'années.

L'outil « valeur cible » ne permet pas d'écrire de façon convenable des contraintes sur les valeurs des cellules variables, comme le fait qu'elles ne prennent que des valeurs entières. Le « solveur », que nous voyons dans la partie suivante, est un outil qui permet d'aller plus loin dans l'écriture d'un problème.

> Solution n°68 (exercice p. 107)

	A	B	C	D	E	F	G
1							
2				Coûts unitaires sur 5 ans			
3		Type	Quantité	Achat	Entretien	Fonctionnement	
4		A	5	600	1900	1000	
5		B	4	1000	1350	1350	
6		C	3	1150	1200	1450	
7		D	2	1800	1100	1100	
8		Totaux	14	14050	20700	16950	51700
9							
10				Contraintes			
11		Borne inférieure	17				
12		Borne supérieure	20	18000	25000	21000	

Figure 4.30 : La préparation des données pour le solveur

Dans la figure 4.30, on voit qu'on a recopié les données dans le tableur (on peut faire un copier-coller de celles-ci), puis on a inséré une colonne pour les quantités (colonne C), qui sont à déterminer, ainsi que leurs bornes 17 et 20 dans les lignes 11 et 12.



Remarque

Les copies d'écran de cette partie correspondent à la résolution de cet exercice avec la version 3.1.0 d'OpenOffice disponible en ligne ainsi que sur le DVD « Vos outils bureautique et internet libres » version 2009-2010 distribuées aux étudiants des universités.

> Solution n°69 (exercice p. 107)

On saisit des quantités quelconques de C4 à C7 afin que les formules écrites

ensuite donnent d'autres valeurs que 0 (heureusement, il n'est pas nécessaire qu'elles soient une solution du problème !). Les formules sont ensuite tapées dans la partie encadrée. En C8 c'est la somme des quantités « =somme(C4 :C7) ». En D8, la somme des achats pondérés par les quantités « =sommeprod(\$C4 :\$C7 ;D4 :D7) », formule recopiée dans les cellules E8 et F8. En G8, c'est le coût total obtenu en sommant les divers coûts « =somme(D8 :F8) ».

> Solution n°70 (exercice p. 107)

On sélectionne la cellule G8 contenant le coût total, puis on lance le solveur par « Outils>Solveur ». On remplit les divers champs de la fenêtre du solveur en cliquant sur les plages de cellules le cas échéant (voir Fig 4.31).

Figure 4.31 : Remplissage des divers champs du solveur d'OpenOffice 3.1

Puis on clique sur « résoudre ». Et si on choisit de conserver le résultat on obtient le nombre optimal de quantités de chaque type qu'il convient d'acheter avec les coûts et les contraintes spécifiées (Voir Fig. 4.32).

	A	B	C	D	E	F	G
3		Type	Quantité	Achat	Entretien	Entretien	
4		A	5	600	1900	1000	
5		B	8	1000	1350	1350	
6		C	2	1150	1200	1450	
7		D	2	1800	1100	1100	Coût total
8		Totaux	17	16900	24900	20900	62700

Résultat de la résolution

La résolution s'est terminée avec succès.

Résultat : 62700

Souhaitez-vous conserver le résultat ou voulez-vous restaurer les valeurs précédentes ?

Figure 4.32 : Solution trouvée par le solveur



Remarque

Il arrive que le solveur ne trouve pas de solution, ou trouve une autre valeur que la solution. Cela peut provenir des paramètres de fonctionnement du tableur, qui sont

modifiables en enfonçant le bouton « Options » de la figure 4.31, comme indiqué dans la figure 4.33 (les options du solveur d'Excel 2003 sont en figure 4.34). Mais cela peut être dû aussi à la convergence vers un minimum local au lieu de global de l'algorithme utilisé. Pour la compréhension de ces problèmes, nous renvoyons le lecteur intéressé aux ouvrages traitant de l'optimisation.

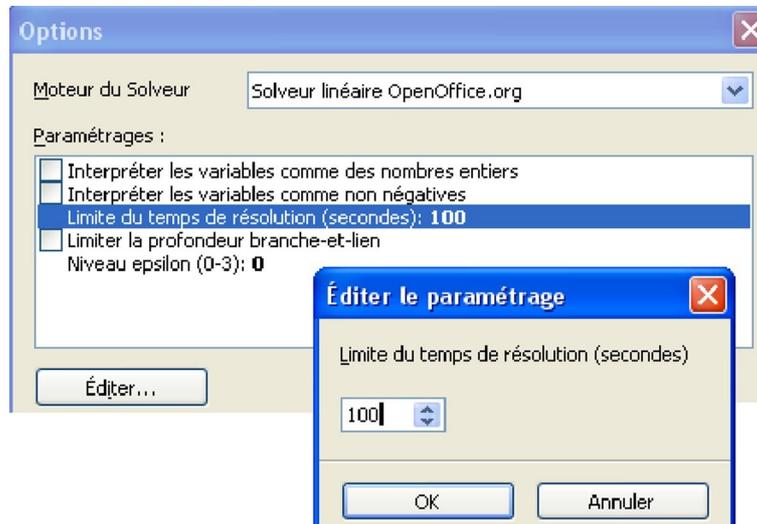


Figure 4.33 : Les paramètres de fonctionnement modifiables avec le solveur d'OpenOffice 3.1

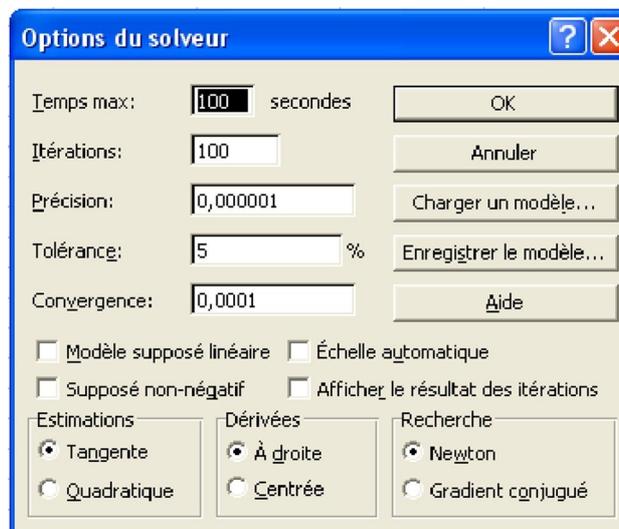


Figure 4.34 : Les paramètres de fonctionnement modifiables avec le solveur d'Excel 2003

> Solution n°71 (exercice p. 107)

Les solutions proposées donnent un coût total de 62 500€, alors qu'on aurait pu avoir un coût total inférieur en prenant les 17 ordinateurs les moins chers. Bien sûr les contraintes sont à l'origine de cette solution. Mais elles ne sont peut-être pas si rigides que cela. Notamment, l'entreprise accepterait certainement un coût total à peine supérieur, pour avoir 18 ordinateurs au lieu de 17. On peut voir sur la figure 4.32 que les plafonds de coûts ne sont pas atteints. Il reste en trésorerie 1100 € de

coûts d'achat, 100 € de coûts d'entretien et autant de fonctionnement, soit 1300€ en tout. Cela ne permet pas d'acheter un nouvel ordinateur (le moins cher est à 3500 €), mais il est peut-être possible d'augmenter les deux derniers plafonds de coûts afin d'avoir d'autres solutions. Une aide à ce type de raisonnement est fournie par les rapports du solveur d'Excel (voir à droite de la figure 4.29 les rubriques « Réponses, sensibilité, limites » du champ « Rapports »). Mais on peut également procéder par essais successifs en changeant quelques cellules et en relançant le solveur sans avoir besoin de remplir à nouveau les champs à condition de se replacer sur la cellule de la valeur à optimiser.

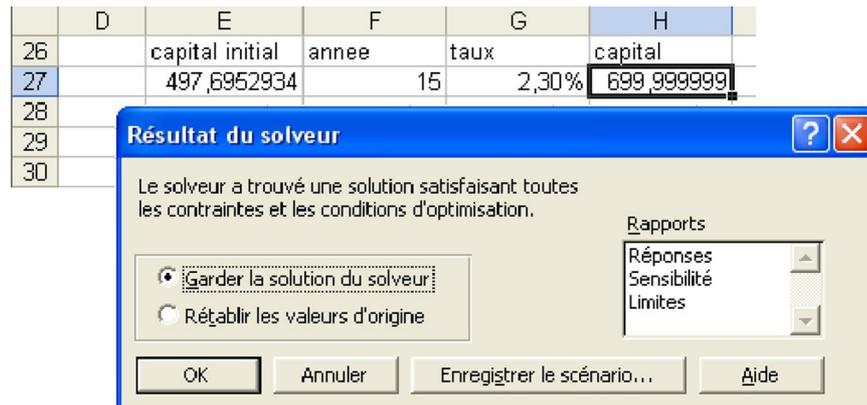


Figure 4.29 : L'utilisation du solveur sous Excel : choix des paramètres

	A	B	C	D	E	F	G
3		Type	Quantité	Achat	Entretien	Fonctionnement	
4		A	5	600	1900	1000	
5		B	8	1000	1350	1350	
6		C	2	1150	1200	1450	
7		D	2	1800	1100	1100	Coût total
B		Totaux	17	16900	24900	20900	62700

Figure 4.32 : Solution trouvée par le solveur

> Solution n°72 (exercice p. 107)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°73 (exercice p. 107)

Consulter la version en ligne du module pour voir la version animée de la solution.

> Solution n°74 (exercice p. 108)

Sans même faire des calculs, on peut en déduire que le bénéfice va baisser, car on gagnera 0,5 € par lot, au lieu de 0,5 € par bouteille et l'augmentation des quantités ne suffira pas à couvrir la baisse de bénéfice : bénéfice par bouteille divisé par trois, quantité multipliée par deux.

> Solution n°75 (exercice p. 108)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°76** (exercice p. 108)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°77** (exercice p. 108)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°78** (exercice p. 108)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°79** (exercice p. 109)

Voir solution de la question 3.

> **Solution n°80** (exercice p. 109)

Voir solution de la question 3.

> **Solution n°81** (exercice p. 109)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°82** (exercice p. 109)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°83** (exercice p. 109)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°84** (exercice p. 110)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°85** (exercice p. 116)



Conseil

En cas de difficultés pour sélectionner un objet du graphique, vérifiez que vous avez bien double-cliqué sur le graphique (il doit alors être entouré d'un trait gris épais). Si la difficulté persiste, agrandir le graphique afin de rendre la sélection de ses objets plus aisée.

> **Solution n°86** (exercice p. 121)

Il convient de sélectionner d'abord les 2 colonnes de valeurs à l'exception du premier et du dernier point, d'insérer le graphique, puis de sélectionner la série de données obtenues, et par un clic droit d'accéder à la plage de données. Il faut alors

ajouter un à un les 2 points comme nouvelle série de donnée (voir figure 5.14), puis de sélectionner chacun afin de lui donner le format adéquat.

Dans l'exemple précédent, on avait en X et en Y des valeurs de même nature (cours d'action à la clôture en euros un jour donné). On peut aussi avoir des valeurs de natures différentes comme dans l'exercice ci-dessous.

> Solution n°87 (exercice p. 122)

On a d'abord importé les données du fichier vers le tableur, puis on a modifié le format des dates figurant dans la première colonne. On a ensuite réalisé le graphique en insérant le diagramme en choisissant le type correspondant à la représentation en nuage de points, avec le sous-type lignes sans points. Ensuite la légende a été déplacée vers le haut du graphique (sous OpenOffice, clic droit sur la légende, puis parmi les propriétés de l'objet, choix de la position « haut »), l'épaisseur de chacune des lignes a été diminuée sous OpenOffice (à 0.00), l'axe des x a été modifié (l'échelle a été corrigée comme indiqué dans la figure 5.16). Quand à l'axe des y, son format monétaire ne vient pas des données, qui sont restées en format standard, mais a été choisi comme propriété de l'objet.

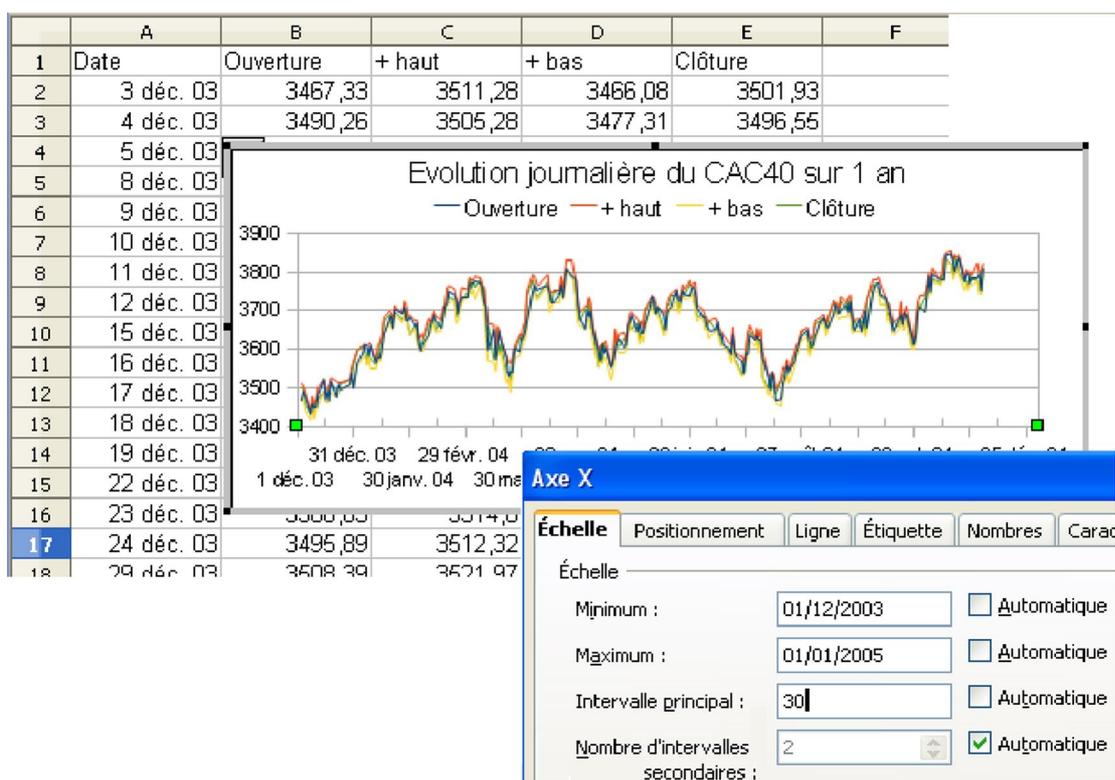


Figure 5.16 : Modification de l'échelle de l'axe des x sous OpenOffice (données de « Cac40_1an.txt »)

> Solution n°88 (exercice p. 138)

Prendre le fichier salaires.txt, le mettre dans deux feuilles d'un classeur. Puis se placer dans la deuxième feuille, la première servant à récupérer des valeurs en cas d'accident. Renommez les feuilles à votre convenance, et sauvegardez le classeur sous le format tableur (.xls pour Excel, .ods pour OpenOffice). Sélectionner la zone

de résultat (dans notre cas 5 lignes et 2 colonnes), et procéder comme pour toute formule matricielle : on tape =droitereg(, on sélectionne la colonne des y (sans les intitulés) , la colonne des x, et on tape vrai et vrai (le premier pour autoriser une valeur de b différente de 0, le second pour demander l'affichage des 4 lignes de statistiques), les 4 arguments étant séparés par des point-virgules, on ferme la parenthèse, on valide en enfonçant simultanément les 3 touches adéquates (CTL, MAJ, Entrée). On obtient ainsi le modèle Salaire = 10,613 Ancienneté + 1163,893

> Solution n°89 (exercice p. 138)

Calcul des estimations : une fois les valeurs saisies dans les cellules K5 à K9, taper dans la cellule L5 la formule « =H\$4*K5+I\$4 », puis tirer vers le bas. Contrôler que le salaire estimé d'embauche (0 années d'ancienneté) est bien égal à b.

Réalisation du graphique : sélectionner les deux colonnes ancienneté et salaire, les représenter par un nuage de points (appelé ainsi sous Excel, et diagramme XY sous OpenOffice). On doit avoir l'ancienneté en abscisse et le salaire en ordonnée. Si vous n'arrivez pas à obtenir cela, vous pouvez changer l'ordre des colonnes de la série sous Excel dans la ligne d'édition une fois la série de points sélectionnés, sinon recopiez les deux colonnes dans le bon ordre (comme fait en figure 5.22 dans les colonnes N et O) et recommencez. Puis sélectionnez le graphique et ajoutez de nouvelles données en indiquant la plage K4 :L8 (« Graphique>Ajouter des données » sous Excel, et sous OpenOffice « Format>Plage de données>Séries de données » comme indiqué sur la figure 4). Changer alors le format de cette nouvelle série pour obtenir la ligne droite attendue.

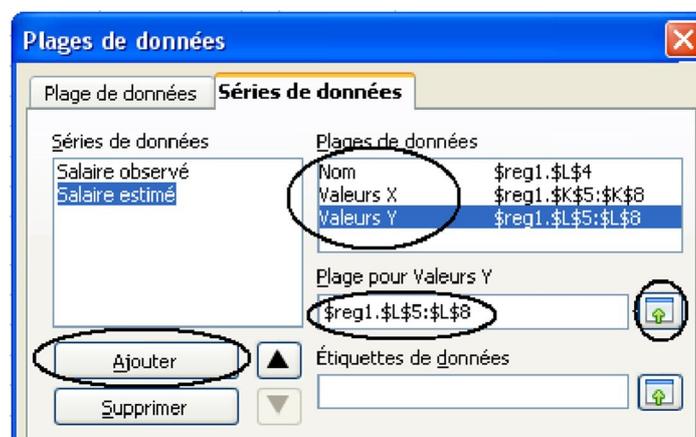


Figure 5.26 : Ajout dans le graphique de la série de données de la plage K4 :L8 sous OpenOffice 3.1

Choisissez le format monétaire pour l'axe des y.

Si on interprète cette équation, on peut dire que le salaire d'embauche estimé d'après le modèle est b (on remplace x par 0 dans l'équation $y=ax+b$), soit 1164 euros environ, et qu'on est augmenté de a, soit 10,61 euros, par année d'ancienneté, ce qui fait bien peu ! D'autres arguments, plus statistiques, vont nous faire abandonner ce modèle pour en trouver un plus approprié.

> Solution n°90 (exercice p. 139)

On y accède par un clic droit après avoir sélectionné la série. Puis on choisit le type linéaire et l'option affichage de l'équation et de R^2 , comme indiqué dans la figure

5.27 pour Excel 2003.

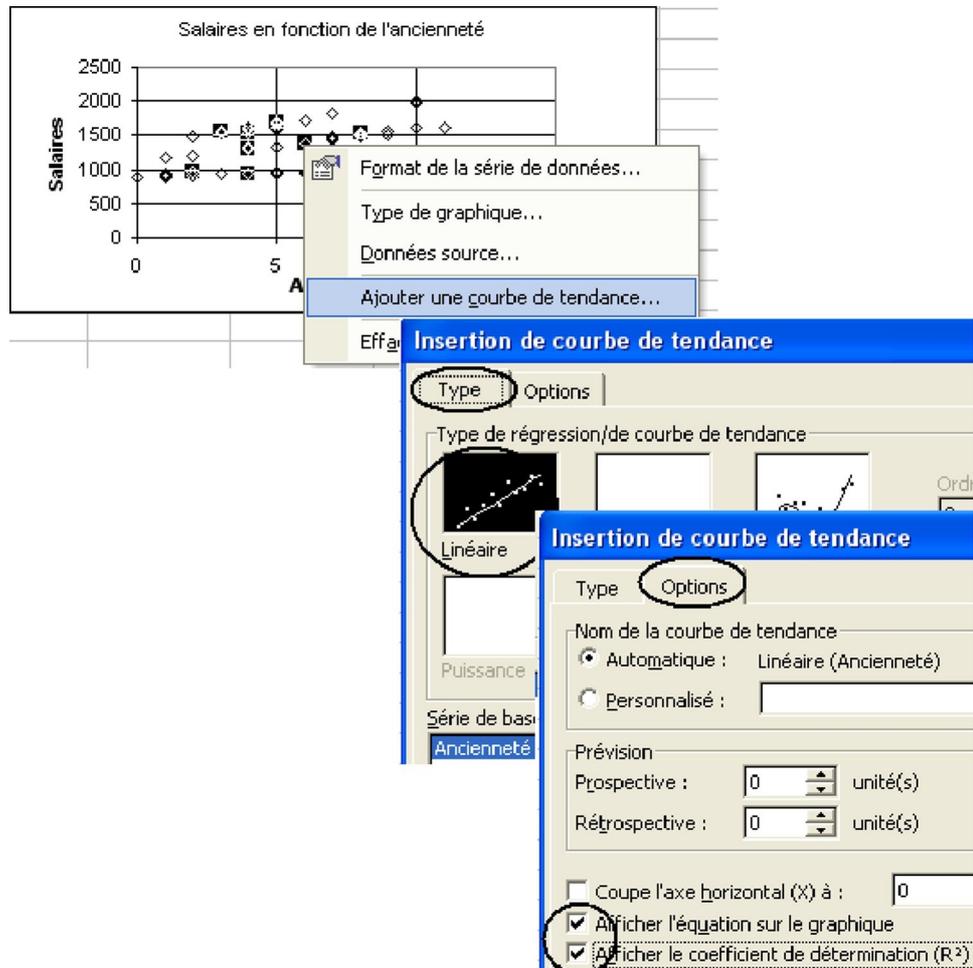


Figure 5.27 : Ajout dans le graphique de la droite de régression linéaire sous Excel 2003

> Solution n°91 (exercice p. 139)

Dans la cellule F2, on tape la formule $=B2-C2*\$H\$4-\$I\4 , et on la recopie vers le bas.

Dans la cellule K12, $=MOYENNE(F2:F99)$ qui doit être 0. On peut toutefois obtenir une valeur très proche de 0 sans lui être égale. En effet les calculs se font en gardant un nombre limité de chiffres après la virgule (la mémoire infinie n'existe pas en informatique), ce qui aboutit à de légers écarts aux valeurs exactes. Ces écarts sont plus importants quand les calculs sont plus lourds, comme pour ceux des coefficients de régression par les moindres carrés qui nécessitent d'inverser une matrice.

Dans la cellule K13, $=ECARTYPE(F2:F99)$, qui donne un résultat proche de celui de la cellule I6, qui est la valeur estimée de l'écart-type des résidus, noté s .

On sélectionne alors la colonne des résidus et on la représente graphiquement par un nuage de points. Les résidus doivent suivre la loi normale $N(0,s)$. Selon cette loi, on connaît les proportions attendues de valeurs dans les intervalles $]-\infty, -2s)$, $(-2s, -s)$, $(-s, 0)$, $(0, s)$, $(s, 2s)$, $(2s, \infty[$, qui sont respectivement de (ces intervalles sont choisis ainsi par tradition, mais on pourrait en choisir d'autres). En

dessinant les 5 droites, cela permet de séparer les points des 6 zones. Pour les tracer, on remplit la zone de H15 à M17 comme indiqué dans la figure 2.24, puis on ajoute ces séries au graphique, en faisant un clic droit sur le graphique sélectionné, puis en choisissant plage de données, et en procédant comme indiqué figure 5.14. Signalons que sous Excel, on peut ajouter en une seule fois la plage correspondant aux 5 droites. On modifie alors le format de chacune pour obtenir la ligne attendue au lieu des 2 points de ses extrémités.

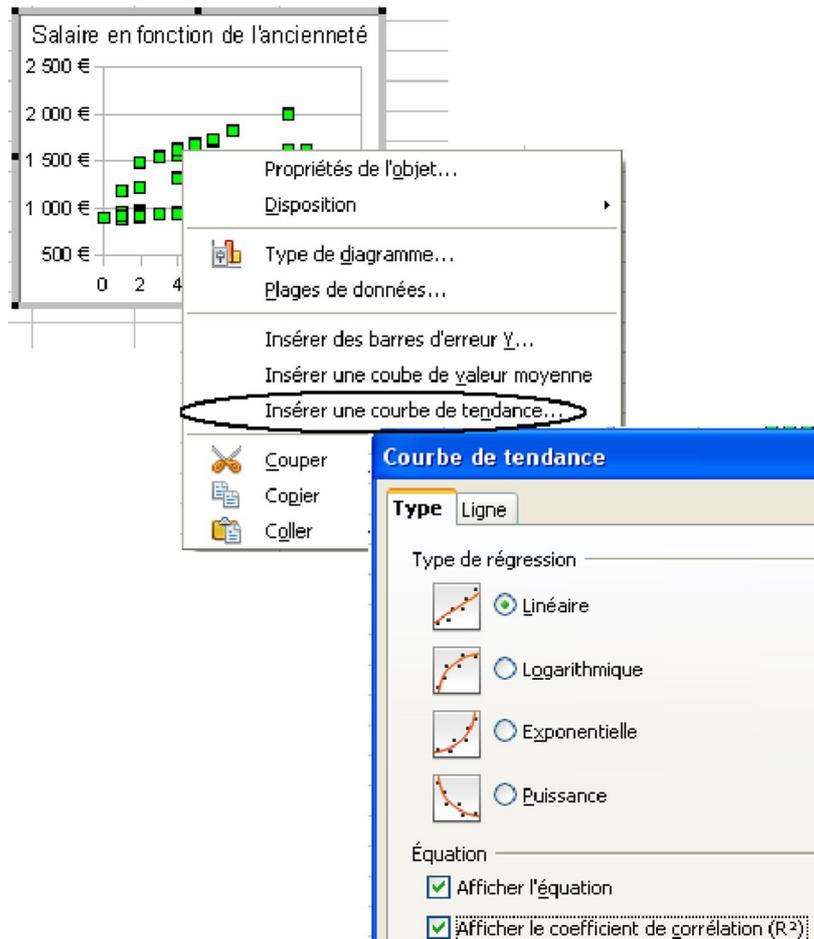


Figure 5.28 : Ajout dans le graphique de la droite de régression linéaire sous OpenOffice 3.1

	A	B	C	D	E	F	G	H	I
55	88	1009	10	c1	M	-261			
56	9	1421	7	c2	F		55	182,8	
57	41	1323	5	c2	F		56	106	
58	49	1615	11	c2	F		57	334,4	
59	64	1375	6	c2	F		58	147,4	
60	71	1431	7	c2	F		59	192,8	
61	72	1501	9	c2	F		60	241,6	
62	77	1476	8	c2	F		61	227,2	
63	79	1472	8	c2	F		62	223,2	
64	87	1487	8	c2	F		63	238,2	
65	5	1321	4	c2	M		64	114,7	
66	17	1543	8	c2	M		65	294,2	
67	21	1480	7	c2	M		66	241,8	
68	22	1544	8	c2	M		67	295,2	
69	23	1459	7	c2	M		68	220,8	
70	44	1613	10	c2	M		69	343	
71	55	1179	1	c2	M		70	4,495	
72	65	1210	2	c2	M		71	24,88	
73	66	1314	4	c2	M		72	107,7	
74	75	1315	4	c2	M		73	108,7	
75	86	1516	8	c2	M		74	267,2	
76	90	1556	9	c2	M		75	296,6	
77	92	1472	7	c2	M		76	233,8	
78	10	1708	6	c3	F		77	480,43	
79	11	1624	5	c3	F		78	407,04	
80	13	1546	4	c3	F		79	339,66	

Figure 5.29 : Les 3 séries décalées de résidus selon la responsabilité, avec ajout des numéros pour c2 et c3

Pour voir la dépendance entre les résidus et le sexe, on sélectionne la plage formée des 6 colonnes et des 99 lignes et on trie selon le sexe. Si on désire différencier les points selon les sexes, comme c'est fait dans le graphique de la figure 5.25, il faut créer 2 séries différentes. Comme la série totale est déjà représentée, on peut la modifier en remplaçant la plage F2 :F99 par F2 :F58, et en insérant une nouvelle série F59 :F99. On peut alors changer le format de chacune. On voit que les résidus ne dépendent apparemment pas du sexe. Inutile donc d'ajouter le sexe dans le modèle M1.

Ce qu'on a fait pour la variable « sexe » est fait de la même façon pour la variable « responsabilité ». On voit cette fois que les résidus dépendent fortement de la responsabilité, qui doit donc être mise dans le modèle.



Remarque

Sous Excel, il est plus pratique d'insérer directement de nouvelles données, mais il faut les décaler par rapport aux anciennes pour en faire une nouvelle série et non de nouveaux points de la même série, ce qui reviendrait à ce qu'on avait avant. Une fois le graphique fait, on peut les remettre en place (voir figure 5.29).

> Solution n°92 (exercice p. 141)

Recodage de la variable responsabilité

Reprendre les données de départ (éviter de prendre celles ordonnées par responsabilité ou par sexe). Insérer 2 colonnes avant la colonne de responsabilité (on peut les créer ailleurs, bien sûr). Ce sont les colonnes D et E de la figure 5.30. En cellule D2, écrire la formule =SI(\$F2=D\$1;1;0), puis la recopier dans la cellule

d'à côté et sélectionner ces 2 cellules et tirer vers le bas pour remplir les 2 colonnes. Contrôler que les 1 s'affichent bien à l'endroit voulu.

Écriture du modèle

Ce modèle s'écrit $S = a_1 \cdot c_2 + a_2 \cdot c_3 + b + e$, et comme il y a 3 paramètres, il faut sélectionner la plage L4 :N8, puis taper la formule =DROITEREG(B2:B99;D2:E99;VRAI;VRAI), et appuyer sur les 3 touches CTRL MAJ Entrée. Taper les lignes d'intitulés du tableau (lignes 2 et 3). L'équation s'écrit $S_{est} = 962 + 475,8 c_2 + 733,3 c_3$. On en déduit que le salaire de base (responsabilité c1 choisie par défaut) est de 962 euros, et qu'il est augmenté de 475 euros environ si on a la responsabilité c2 (soit $962 + 475$), contre 722 euros s'il a la responsabilité c3 (soit $962 + 722$).

Utilisation du modèle

Ce modèle ne prend en compte que la responsabilité. On a écrit dans la colonne S les 3 valeurs de responsabilité c1, c2 et c3. La formule de calcul du salaire écrite dans T12 est =N\$4+M\$4*(S12=M\$3)+L\$4*(S12=L\$3). Puis elle est tirée vers le bas, et on retrouve bien les résultats des sommes indiquées dans le paragraphe précédent.

Évaluation de la qualité du modèle

La valeur de R^2 se trouve dans la cellule L6. Elle est de 0,902, ce qui est très bon. On examine maintenant les résidus. Pour les calculer, on écrit dans H2 la formule =B2-N\$4-M\$4*D2-L\$4*E2, que l'on recopie vers le bas. Les caractéristiques statistiques de ceux-ci sont ensuite calculées. La moyenne fait approximativement 0, comme attendu, et l'écart-type est proche de sa valeur théorique écrite dans la cellule M6. On représente alors le nuage des résidus avec les 6 zones délimitées par les 5 droites. Il a l'air de suivre à peu près la loi normale (en cas de difficulté pour réaliser cette partie, voir les explications détaillées correspondantes du modèle M1).

Indépendance des résidus et des autres variables

Ce modèle est bon, mais on peut peut-être l'améliorer en faisant entrer les autres variables dans le modèle. Pour cela on regarde si les résidus sont liés à d'autres variables, par exemple l'ancienneté. On représente graphiquement les résidus de M2 en fonction de l'ancienneté et le nuage de points montant indique clairement qu'il y a une liaison positive entre les 2 (voir figure 5.31). On ajoute la droite de régression (une fois le graphique sélectionné, par un clic droit, on peut lui ajouter une tendance, qu'on choisit linéaire), son équation et le R^2 correspondant, ce qui confirme clairement la liaison, et invite à faire un modèle de régression avec les deux variables explicatives, ancienneté et responsabilité.

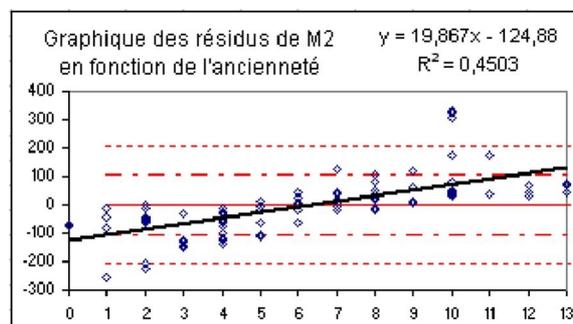


Figure 5.31 : Les résidus de M2 sont une fonction croissante de l'ancienneté

> Solution n°93 (exercice p. 141)

Une fois la copie de feuille faite, on la renomme puis on sélectionne la plage du modèle M2 (plage de cellules L4 :N8) et on sélectionne une zone de 4 colonnes et 5 lignes, comme indiqué en figure 5.33, et on corrige la formule DROITEREG, en remplaçant la plage des variables explicatives étant transformée en C2 : E99 au lieu de D2 :E99. Le modèle obtenu est $Sest.=829+20 A+471,1 c2+764,4 c3$, comme on peut le voir dans la figure 5.34. Ce qui signifie qu'avec 0 années d'ancienneté, on gagne 829 euros si la responsabilité c1, 471 euros environ de plus si la responsabilité est c2, et 746 euros de plus si elle est c3. Si l'ancienneté change, on gagne 20,346 euros de plus par année d'ancienneté, et ceci quelle que soit la responsabilité. On saisit les intitulés des lignes et colonnes de la plage T3 :W15 et on obtient les estimations détaillées de la plage T13 : W15 en écrivant dans la cellule T13 la formule $=\$O\$4+\$M\$4*(\$S13=\$M\$3)+\$L\$4*(\$S13=\$L\$3)+T\$12*\$N\$4$, et en la recopiant vers la droite et vers le bas.

SOMME													=DROITEREG(B2:B99;D2:E99;VRAI;VRAI)			
	A	B	C	D	E	F	G	H	IJK	L	M	N	O			
1	Num	Salair	Anc	c2	c3	Res	Se	Ecarts								
2	1	932	4	0	0	c1	F	-30		M2						
3	2	950	8	0	0	c1	F	-12		c3	c2	Ancienneté				
4	3	978	8	0	0	c1	F	16,4		2:E99;V	475,8	962				
5	4	966	7	0	0	c1	F	4,35		26,19	26,19	14,1				
6	5	1321	4	1	0	c2	M	-116		0,902	103,5	#N/A				
7	6	993	12	0	0	c1	F	31,4		437	95	#N/A				
8	7	961	2	0	0	c1	M	-0,6		9E+06	1E+06	#N/A				

Figure 5.33 : correction du modèle M2 en un modèle M3 sous Excel contenant l'ancienneté en plus

	A	B	C	D	E	F	G	H	IJK	L	M	N	O	P	Q	R	S	T	U	V	W		
1	Num	Salair	Anc	c2	c3	Res	Se	Ecarts															
2	1	932	4	0	0	c1	F	21,2		M3													
3	2	950	8	0	0	c1	F	-42		c3	c2	Ancie	Constan										
4	3	978	8	0	0	c1	F	-14		746,4	471,1	20	829										
5	4	966	7	0	0	c1	F	-5,8		19,51	19,33	2,3	18										
6	5	1321	4	1	0	c2	M	-61		0,947	76,4	###	#N/A										
7	6	993	12	0	0	c1	F	-81		561,8	94	###	#N/A										
8	7	961	2	0	0	c1	M	90,9		1E+07	5E+05	###	#N/A										
9	8	1006	10	0	0	c1	M	-27															
10	9	1421	7	1	0	c2	F	-22		Carac. des écarts à M3													
11	10	1708	6	0	1	c3	F	10,2		Moyenne			-0		Salaires estimés d'après M3								
12	11	1624	5	0	1	c3	F	-53		Ecart-type			75,2					0	5	10	15		
13	12	981	7	0	0	c1	M	9,18										c1	829	931	1033	1135	
14	13	1546	4	0	1	c3	F	-111	x	$y = 2s$	$y = s$	$y = 0$	$y = -s$	$y = -2s$				c2	1301	1402	1504	1606	
15	14	929	3	0	0	c1	F	38,6		1	152,8	76	0	-76	-153				c3	1576	1677	1779	1881
16	15	909	2	0	0	c1	F	38,9		100	152,8	76	0	-76	-153								

Figure 5.34 : Réalisation du modèle M3 exprimant le salaire en fonction de la responsabilité et de l'ancienneté

> Solution n°94 (exercice p. 142)

le R^2 est passé de 0,906 à 0,947, ce qui indique une amélioration, mais elle n'a pas nécessairement de sens. En effet chaque fois qu'on ajoute une variable, la valeur de R^2 augmente généralement, au point que si on met autant ou plus de variables que d'individus, on peut obtenir un R^2 de 1. Il faut contrôler que l'augmentation de R^2 a un sens (du point de vue statistique). Nous avons proposé d'utiliser pour cela la statistique $F=R^2_{partiel}/(1-R^2_{partiel})/(q2-q1)*(n-q2)$ avec $R^2_{partiel}=(R^2_2-R^2_1)/(1-R^2_1)$, on peut voir les résultats du calcul dans la cellule AB10, et la probabilité p d'avoir un $R^2_{partiel}$ aussi grand par hasard est inférieure à 0,01, ce qui nous permet de

conclure que le R^2_{partiel} est très significatif, et que le passage du modèle M2 au modèle M3 est justifié statistiquement.

Remarque



pour trouver exactement les mêmes valeurs que dans la figure 5.35, il faut veiller à écrire dans les formules les adresses des nombres si ceux-ci ne sont pas entiers, et non les valeurs qu'on voit à l'affichage, car ce sont leurs valeurs approchées (par exemple dans la cellule AB3 la formule est =reg2!L6). La valeur de p se trouve par la formule LOI.F(AB10;AB8;AB9) écrite dans AB11.

On peut aussi tester la significativité de chaque paramètre du modèle (c.à.d. qu'ils diffèrent significativement de 0). Le test peut être bilatéral (on ne sait pas s'ils risquent d'avoir un effet positif ou négatif sur le salaire), ou unilatéral. Ici on suppose que l'ancienneté, la responsabilité élevée ont un effet positif sur le salaire, c'est un test unilatéral. Pour calculer la statistique $t = a/s$, on utilise les deux premières lignes du modèle de régression. On sait que si t dépasse largement 2, et que le ddl est élevé, comme ici, la probabilité p d'avoir une valeur si grande par hasard sera très faible. Elle est calculée dans la ligne 7. Dans la cellule AF7 on a la formule =LOI.STUDENT(AF6;\$AI6;1) qu'on recopie vers la droite. On obtient que toutes les valeurs de p sont inférieures à 0,01 et donc que les 3 paramètres sont très significatifs. Il est à noter que le passage de M2 à M3 porte sur une seule variable, la variable ancienneté, et comme elle est unique, on a la relation entre la valeur t ancienneté et FM3/M2 : $t^2 = F$. Et quand on prend un test bilatéral, ce qui n'est pas le cas ici, on trouve la même probabilité avec t ou avec F. Avec un test unilatéral, on trouve une probabilité associée à F double de celle associée à t.

Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1	Test F pour R^2_{partiel} de M3/M2								
2	n	98			Test t pour chaque paramètre a de M3				
3	R^2_1	0,902				c3	c2	Ancien ddl	
4	q1	3			a	746,4	471,1	20,35	
5	R^2_2	0,947			s	19,51	19,33	2,268	
6	q2	4			a/s	38,26	24,37	8,97	94
7	R^2_{partiel}	0,461			p_{un}	2E-59	1E-42	1E-14	
8	ddl1	1							
9	ddl2	94							
10	F	80,46							
11	p	3E-14							

Figure 5.35 : Réalisation du test de significativité du modèle M3/M2 à gauche et de tous les coefficients de l'équation de régression de M3 à droite. Tout est très significatif, les valeurs de p étant $< 0,01$

Le calcul des résidus doit être aussi mis à jour. La formule dans H2 est =B2-O\$4-M\$4*D2-L\$4*E2-C2*N\$4 et on la recopie vers le bas. A ce moment-là, dans les caractéristiques des résidus, la moyenne passe à 0 et l'écart-type devient peu différent du contenu de la cellule. Les équations des 5 droites sont mises à jour et le graphique des résidus également. Si on fait le graphique des résidus en fonction du sexe, on voit que les résidus des femmes (losanges rouges) sont plus groupés autour de l'axe que ceux des hommes (carrés bleus)., ce qui montre une différence de variance.

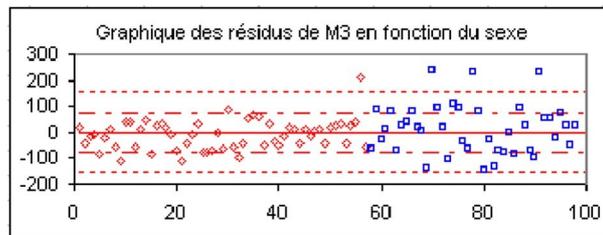


Figure 5.36: Les résidus de M3 diffèrent selon le sexe

On vérifie cela par un tableau croisé dynamique donnant les statistiques des écarts. On constate alors que non seulement les variances (écart-type²) diffèrent mais également les moyennes. Ce qui invite à prendre en compte le sexe dans le modèle.

Données	Sex		Total
	F	M	
Nombre de Ecarts	57	41	98
Moyenne de Ecarts	-9,25	12,86	-0
Écartype de Ecarts	56,7	94,52	75,2

Tableau 19 : Tableau 1

> Solution n°95 (exercice p. 142)

On fait une copie de la feuille de calcul du modèle M3. On insère une colonne parmi les colonnes de variables explicatives, car la fonction droitereg du tableur exige que les colonnes de celles-ci soient contiguës. On a choisit de l'insérer en colonne F. On écrit l'intitulé dans la cellule F1, et en F2 la formule =SI(H2=F\$1;1;0), qu'on recopie vers le bas. Puis on corrige la plage M4 :P8 en M4 :Q8 : après l'avoir sélectionnée, on tape la formule =DROITEREG(B2:B99;C2:F99;VRAI;VRAI) et on enfonce les 3 touches CTRL, MAJ, Entrée. On obtient alors ce qui est dans la figure 5.37 : $S = 820 + 20,8 A + 460,6 c_2 + 732,7 c_3 + 28,23 M$. Avec ce modèle le salaire de l'homme, avec même ancienneté et même responsabilité, est supérieur de 28,23 euros à celui de la femme. La différence entre les deux modèles n'est pas grande. On a repris les tests, et on voit que le test F de passage de M3 à M4 n'est pas significatif, et que le test t de l'augmentation de 28,23€ est à la limite de la significativité ($p \approx 0,05$). On décide de ne pas inclure le sexe dans le modèle pour l'instant et on retourne au modèle M3.

	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	M	Res	Se	Ecarts																		
2	0	c1	F	29	M4									n	98			M	c3	c2	Ancien	ddl
3	0	c1	F	-37	M	c3	c2	Ancie	Constante					R ² ₁	0,947		a	28,23	732,7	461	20,83	
4	0	c1	F	-8,6										q1	4		s	17,53	21,13	20,3	2,269	
5	0	c1	F	0,2										R ² ₂	0,949		a/s	1,61	34,67	22,7	9,18	93
6	1	c2	M	-71										q2	5		p _{uni}	0,055	3E-55	0	6E-15	
7	0	c1	F	-77										R ² _{part}	0,027							
8	1	c1	M	71										ddl1	1							
9	1	c1	M	-51										ddl2	93							
10	0	c2	F	-5,4										F	2,593							
11	0	c3	F	30										p	0,111							
12	0	c3	F	-33																		
13	1	c1	M	-13																		
14	0	c3	F	-90																		
15	0	c1	F	47																		
16	0	c1	F	47																		

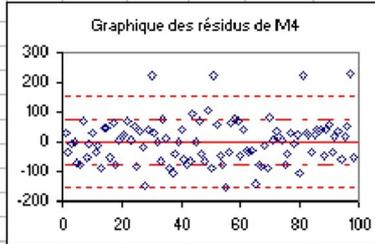


Figure 5.37 : Réalisation du modèle M4 exprimant le salaire en fonction de la responsabilité, de l'ancienneté et du sexe

> Solution n°96 (exercice p. 143)

Pour faire le modèle M5, on ajoute deux colonnes, la colonne F contenant Ac2, obtenue en multipliant la colonne C et la colonne D membre à membre, et la colonne G contenant Ac3, obtenue en multipliant la colonne C et la colonne E membre à membre, puis on procède comme pour les modèles précédents.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Num	Salaire	Anc	c2	c3	Ac2	Ac3	Res	Se	Ecarts												
2	1	932	4	0	0	0	0	c1	F	-6,9	M5											
3	2	950	8	0	0	0	0	c1	F	-25	Ac3	Ac2	c3	c2	Ancien	Constante	R ² ₁	0,947				
4	3	978	8	0	0	0	0	c1	F	2,69							q1	4				
5	4	966	7	0	0	0	0	c1	F	-0,2							R ² ₂	0,995				
6	5	1321	4	1	0	4	0	c2	M	10,8							q2	6				
7	6	993	12	0	0	0	0	c1	F	-19							R ² _{part}	0,909				
8	7	961	2	0	0	0	0	c1	M	40,3							ddl1	2				
9	8	1006	10	0	0	0	0	c1	M	12,5							ddl2	92				
10	9	1421	7	1	0	7	0	c2	F	-29							F	458,9				
11	10	1708	6	0	1	0	6	c3	F	-22							p	1E-48				
12	11	1624	5	0	1	0	5	c3	F	-39							a	58,32	37,54	422,9	221,2	9,106
13	12	981	7	0	0	0	0	c1	M	14,8							s	2,158	2,189	13,24	15,8	0,791
14	13	1546	4	0	1	0	4	c3	F	-49							a/s	27,03	17,15	31,95	13,99	11,52
15	14	929	3	0	0	0	0	c1	F	-0,8							p _{uni}	8E-46	1E-30	7E-52	8E-25	8E-20
16	15	909	2	0	0	0	0	c1	F	-12												
17	16	997	11	0	0	0	0	c1	F	-5,6												
18	17	1543	8	1	0	8	0	c2	M	46,2												
19	18	1585	4	0	1	0	4	c3	M	-10												
20	19	958	5	0	0	0	0	c1	M	10												
21	20	961	6	0	0	0	0	c1	F	3,9												
22	21	1480	7	1	0	7	0	c2	M	29,9												
23	22	1544	8	1	0	8	0	c2	M	47,2												
24	23	1459	7	1	0	7	0	c2	M	8,87												
25	24	916	2	0	0	0	0	c1	F	-4,7												
26	25	1008	13	0	0	0	0	c1	F	-13												
27	26	878	1	0	0	0	0	c1	F	-34												
28	27	1041	10	0	0	0	0	c1	M	47,5												

Figure 5.39 : Réalisation du modèle M5 exprimant le salaire en fonction de la responsabilité, de l'ancienneté avec interaction entre responsabilité et ancienneté

Le modèle s'écrit Sest.=902,5 + 9,1 A+ 221,2 c2+ 422,9 c3 + 37,4 A c2 + 58,32 A c3.

Quand c1 =1, tous les termes contenant c2 et c3 s'annulent, on obtient

Sest.=902,5 + 9,1 A

Quand $c_2 = 1$, tous les termes contenant c_3 s'annulent, on obtient Sest.= (902,5 + 221,2) + (9,1 + 37,4) A

Quand $c_3 = 1$, tous les termes contenant c_2 s'annulent, on obtient Sest.= (902,5 + 422,9) + (9,1 + 58,32) A

C'est ce qui est calculé dans les lignes 25 et 27. La ligne 27 contient les constantes des 3 droites, et la ligne 25 contient les pentes des 3, avec les valeurs exactes car ce sont des adresses qui sont dans les formules, et non les valeurs approchées qu'on vient d'écrire dans les lignes précédentes. On insère alors dans le nuage de points les 3 droites de la plage :R28.

> Solution n°97 (exercice p. 143)

Les résidus sont calculés dans la colonne J.

La formule en J2 est =B2-S\$4-Q\$4*D2-P\$4*E2-C2*R\$4-O\$4*F2-N\$4*G2.

On voit que l'apport de M5 par rapport à M3 est très significatif (cellule V10, $p < 0,01$), et que chaque interaction l'est aussi (N8 et O8, $p < 0,01$). Si on dessine les résidus en fonction du sexe, on réalise que quasiment tous les résidus des hommes sont positifs alors que ceux des femmes sont négatifs (voir figure 5.40).

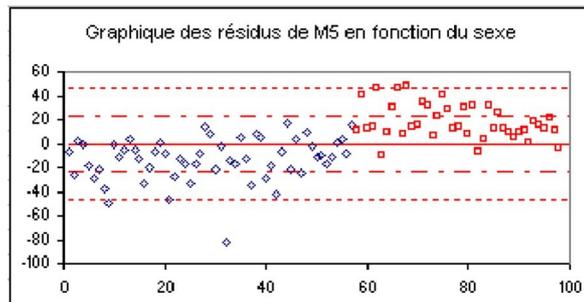


Figure 5.40: Les résidus de M5 diffèrent selon le sexe

On peut voir le modèle M6 dans la figure 5.41. Son apport par rapport à M5 est significatif, et tous les paramètres de la régression le sont. On ne va pas plus loin, car les données ont été construites artificiellement à partir de cette équation, avec ajout d'écarts suivant la loi normale. Et on a retrouvé le modèle progressivement.

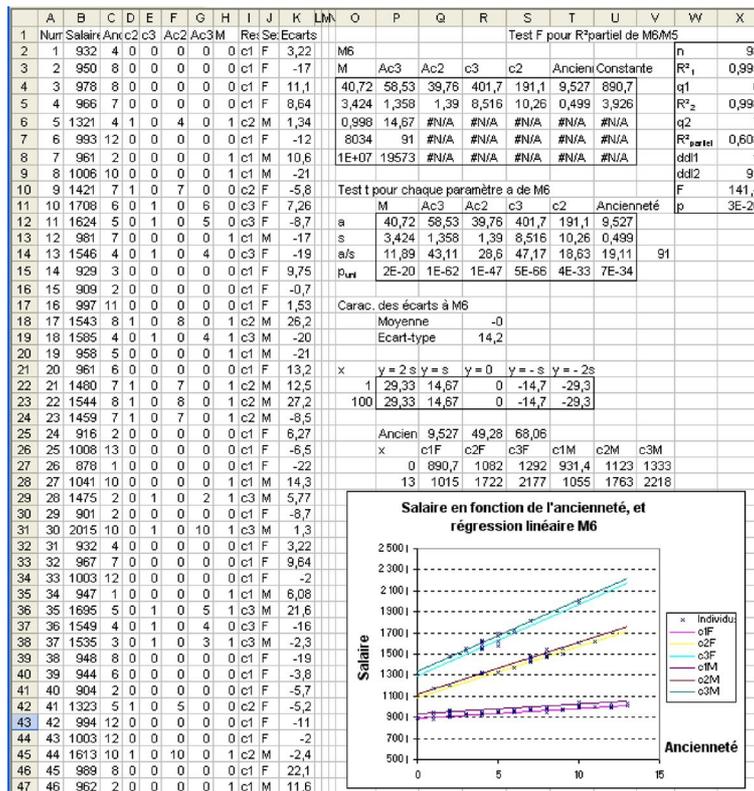


Figure 5.41: Le modèle M6

> **Solution n°98** (exercice p. 144)

Prendre le fichier salaires.txt, le mettre dans deux feuilles d'un classeur. Puis se placer dans la deuxième feuille, la première servant à récupérer des valeurs en cas d'accident. Renommez les feuilles à votre convenance, et sauvegardez le classeur sous le format tableur (.xls pour Excel, .ods pour OpenOffice). Sélectionner la zone de résultat (dans notre cas 5 lignes et 2 colonnes), et procéder comme pour toute formule matricielle : on tape =droitereg(, on sélectionne la colonne des y (sans les intitulés) , la colonne des x, et on tape vrai et vrai (le premier pour autoriser une valeur de b différente de 0, le second pour demander l'affichage des 4 lignes de statistiques), les 4 arguments étant séparés par des point-virgules, on ferme la parenthèse, on valide en enfonçant simultanément les 3 touches adéquates (CTL, MAJ, Entrée). On obtient ainsi le modèle Salaire = 10,613 Ancienneté + 1163,893

> **Solution n°99** (exercice p. 144)

Calcul des estimations : une fois les valeurs saisies dans les cellules K5 à K9, taper dans la cellule L5 la formule « =H\$4*K5+I\$4 », puis tirer vers le bas. Contrôler que le salaire estimé d'embauche (0 années d'ancienneté) est bien égal à b.

Réalisation du graphique : sélectionner les deux colonnes ancienneté et salaire, les représenter par un nuage de points (appelé ainsi sous Excel, et diagramme XY sous OpenOffice). On doit avoir l'ancienneté en abscisse et le salaire en ordonnée. Si vous n'arrivez pas à obtenir cela, vous pouvez changer l'ordre des colonnes de la série sous Excel dans la ligne d'édition une fois la série de points sélectionnés, sinon recopiez les deux colonnes dans le bon ordre (comme fait en figure 5.22 dans les colonnes N et O) et recommencez. Puis sélectionnez le graphique et ajouter de

nouvelles données en indiquant la plage K4 :L8 (« Graphique>Ajouter des données » sous Excel, et sous OpenOffice « Format>Plage de données>Séries de données » comme indiqué sur la figure 4). Changer alors le format de cette nouvelle série pour obtenir la ligne droite attendue.



Figure 5.26 : Ajout dans le graphique de la série de données de la plage K4 :L8 sous OpenOffice 3.1

Choisissez le format monétaire pour l'axe des y.

Si on interprète cette équation, on peut dire que le salaire d'embauche estimé d'après le modèle est b (on remplace x par 0 dans l'équation $y=ax+b$), soit 1164 euros environ, et qu'on est augmenté de a, soit 10,61 euros, par année d'ancienneté, ce qui fait bien peu ! D'autres arguments, plus statistiques, vont nous faire abandonner ce modèle pour en trouver un plus approprié.

> Solution n°100 (exercice p. 145)

On y accède par un clic droit après avoir sélectionné la série. Puis on choisit le type linéaire et l'option affichage de l'équation et de R^2 , comme indiqué dans la figure 5.27 pour Excel 2003.

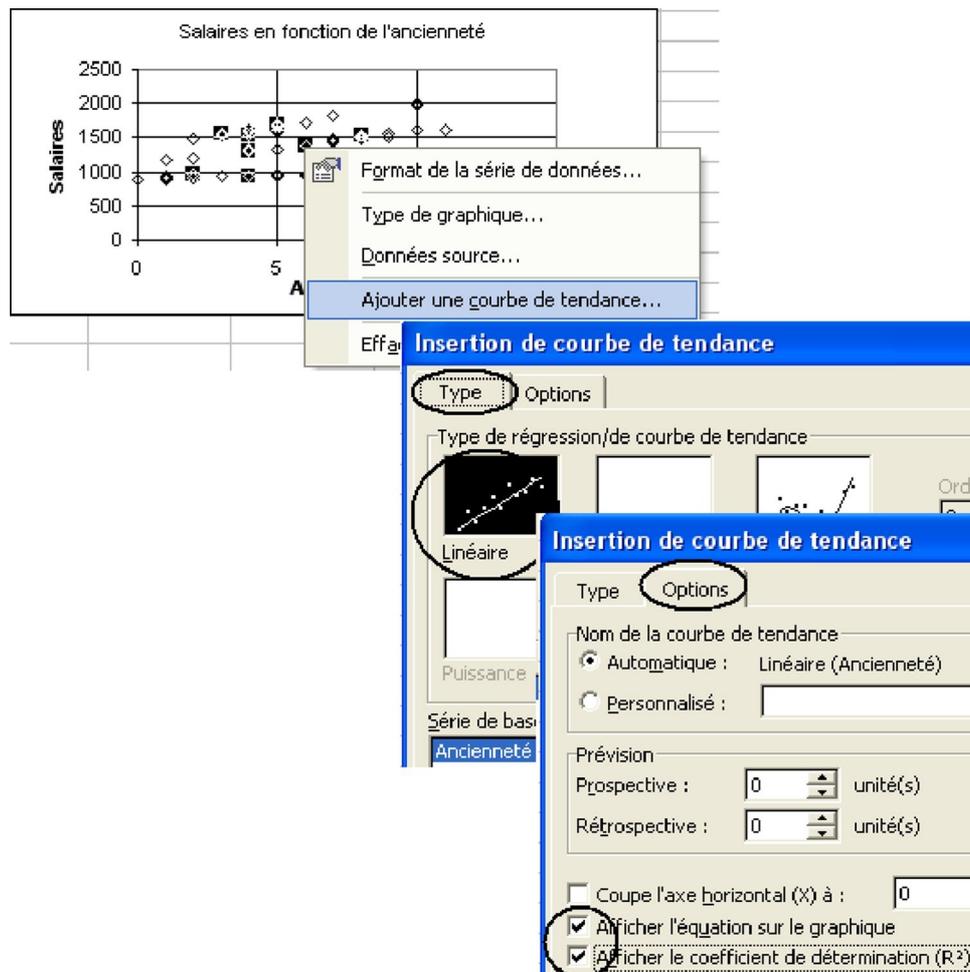


Figure 5.27 : Ajout dans le graphique de la droite de régression linéaire sous Excel 200

> Solution n°101 (exercice p. 145)

Dans la cellule F2, on tape la formule $=B2-C2*\$H\$4-\$I\4 , et on la recopie vers le bas.

Dans la cellule K12, $=MOYENNE(F2:F99)$ qui doit être 0. On peut toutefois obtenir une valeur très proche de 0 sans lui être égale. En effet les calculs se font en gardant un nombre limité de chiffres après la virgule (la mémoire infinie n'existe pas en informatique), ce qui aboutit à de légers écarts aux valeurs exactes. Ces écarts sont plus importants quand les calculs sont plus lourds, comme pour ceux des coefficients de régression par les moindres carrés qui nécessitent d'inverser une matrice.

Dans la cellule K13, $=ECARTYPE(F2:F99)$, qui donne un résultat proche de celui de la cellule I6, qui est la valeur estimée de l'écart-type des résidus, noté s .

On sélectionne alors la colonne des résidus et on la représente graphiquement par un nuage de points. Les résidus doivent suivre la loi normale $N(0,s)$. Selon cette loi, on connaît les proportions attendues de valeurs dans les intervalles $]-\infty, -2s)$, $(-2s, -s)$, $(-s, 0)$, $(0, s)$, $(s, 2s)$, $(2s, \infty[$, qui sont respectivement de (ces intervalles sont choisis ainsi par tradition, mais on pourrait en choisir d'autres). En dessinant les 5 droites, cela permet de séparer les points des 6 zones. Pour les

tracer, on remplit la zone de H15 à M17 comme indiqué dans la figure 2.24, puis on ajoute ces séries au graphique, en faisant un clic droit sur le graphique sélectionné, puis en choisissant plage de données, et en procédant comme indiqué figure 5.14. Signalons que sous Excel, on peut ajouter en une seule fois la plage correspondant aux 5 droites. On modifie alors le format de chacune pour obtenir la ligne attendue au lieu des 2 points de ses extrémités.

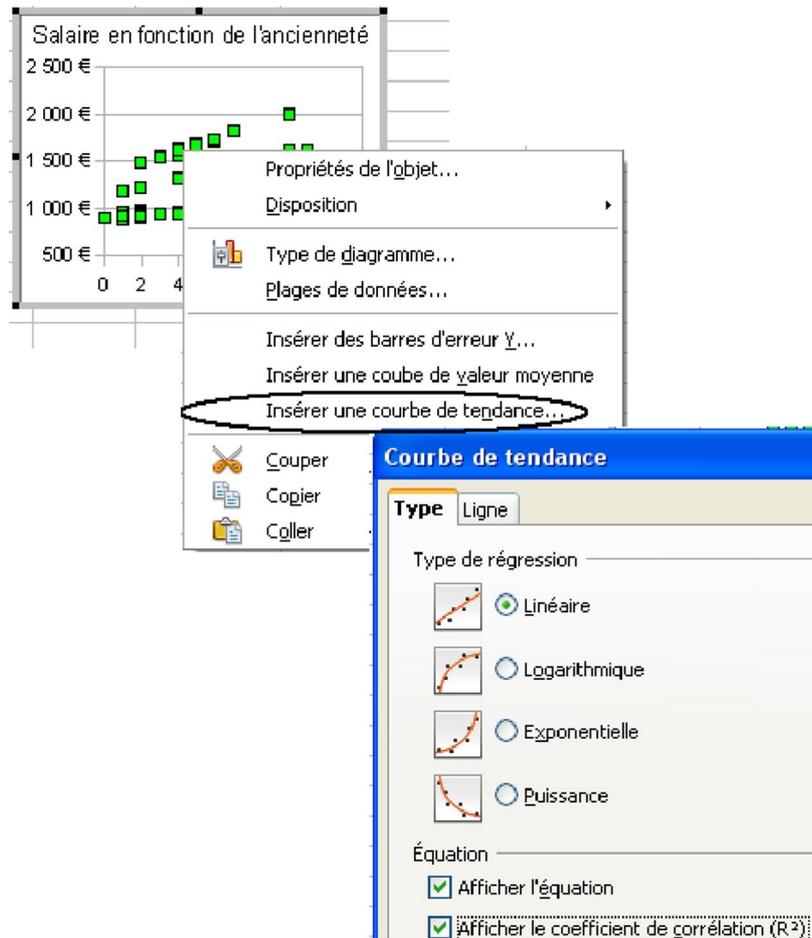


Figure 5.28 : Ajout dans le graphique de la droite de régression linéaire sous OpenOffice 3.1

	A	B	C	D	E	F	G	H	I
55	88	1009	10	c1	M	-261			
56	9	1421	7	c2	F		55	182,8	
57	41	1323	5	c2	F		56	106	
58	49	1615	11	c2	F		57	334,4	
59	64	1375	6	c2	F		58	147,4	
60	71	1431	7	c2	F		59	192,8	
61	72	1501	9	c2	F		60	241,6	
62	77	1476	8	c2	F		61	227,2	
63	79	1472	8	c2	F		62	223,2	
64	87	1487	8	c2	F		63	238,2	
65	5	1321	4	c2	M		64	114,7	
66	17	1543	8	c2	M		65	294,2	
67	21	1480	7	c2	M		66	241,8	
68	22	1544	8	c2	M		67	295,2	
69	23	1459	7	c2	M		68	220,8	
70	44	1613	10	c2	M		69	343	
71	55	1179	1	c2	M		70	4,495	
72	65	1210	2	c2	M		71	24,88	
73	66	1314	4	c2	M		72	107,7	
74	75	1315	4	c2	M		73	108,7	
75	86	1516	8	c2	M		74	267,2	
76	90	1556	9	c2	M		75	296,6	
77	92	1472	7	c2	M		76	233,8	
78	10	1708	6	c3	F			77	480,43
79	11	1624	5	c3	F			78	407,04
80	13	1546	4	c3	F			79	339,66

Figure 5.29 : Les 3 séries décalées de résidus selon la responsabilité, avec ajout des numéros pour c2 et c3

Pour voir la dépendance entre les résidus et le sexe, on sélectionne la plage formée des 6 colonnes et des 99 lignes et on trie selon le sexe. Si on désire différencier les points selon les sexes, comme c'est fait dans le graphique de la figure 5.25, il faut créer 2 séries différentes. Comme la série totale est déjà représentée, on peut la modifier en remplaçant la plage F2 :F99 par F2 :F58, et en insérant une nouvelle série F59 :F99. On peut alors changer le format de chacune. On voit que les résidus ne dépendent apparemment pas du sexe. Inutile donc d'ajouter le sexe dans le modèle M1.

Ce qu'on a fait pour la variable « sexe » est fait de la même façon pour la variable « responsabilité ». On voit cette fois que les résidus dépendent fortement de la responsabilité, qui doit donc être mise dans le modèle.

Remarque



Sous Excel, il est plus pratique d'insérer directement de nouvelles données, mais il faut les décaler par rapport aux anciennes pour en faire une nouvelle série et non de nouveaux points de la même série, ce qui reviendrait à ce qu'on avait avant. Une fois le graphique fait, on peut les remettre en place (voir figure 5.29).

> Solution n°102 (exercice p. 147)

Recodage de la variable responsabilité

Reprendre les données de départ (éviter de prendre celles ordonnées par responsabilité ou par sexe). Insérer 2 colonnes avant la colonne de responsabilité (on peut les créer ailleurs, bien sûr). Ce sont les colonnes D et E de la figure 5.30. En cellule D2, écrire la formule =SI(\$F2=D\$1;1;0), puis la recopier dans la cellule

d'à côté et sélectionner ces 2 cellules et tirer vers le bas pour remplir les 2 colonnes. Contrôler que les 1 s'affichent bien à l'endroit voulu.

Écriture du modèle

Ce modèle s'écrit $S = a_1 * c_2 + a_2 * c_3 + b + e$, et comme il y a 3 paramètres, il faut sélectionner la plage L4 :N8, puis taper la formule =DROITEREG(B2:B99;D2:E99;VRAI;VRAI), et appuyer sur les 3 touches CTRL MAJ Entrée. Taper les lignes d'intitulés du tableau (lignes 2 et 3). L'équation s'écrit $S_{est.} = 962 + 475,8 c_2 + 733,3 c_3$. On en déduit que le salaire de base (responsabilité c_1 choisie par défaut) est de 962 euros, et qu'il est augmenté de 475 euros environ si on a la responsabilité c_2 (soit $962 + 475$), contre 722 euros s'il a la responsabilité c_3 (soit $962 + 722$).

Utilisation du modèle

Ce modèle ne prend en compte que la responsabilité. On a écrit dans la colonne S les 3 valeurs de responsabilité c_1 , c_2 et c_3 . La formule de calcul du salaire écrite dans T12 est =N\$4+M\$4*(S12=M\$3)+L\$4*(S12=L\$3). Puis elle est tirée vers le bas, et on retrouve bien les résultats des sommes indiquées dans le paragraphe précédent.

Évaluation de la qualité du modèle

La valeur de R^2 se trouve dans la cellule L6. Elle est de 0,902, ce qui est très bon. On examine maintenant les résidus. Pour les calculer, on écrit dans H2 la formule =B2-N\$4-M\$4*D2-L\$4*E2, que l'on recopie vers le bas. Les caractéristiques statistiques de ceux-ci sont ensuite calculées. La moyenne fait approximativement 0, comme attendu, et l'écart-type est proche de sa valeur théorique écrite dans la cellule M6. On représente alors le nuage des résidus avec les 6 zones délimitées par les 5 droites. Il a l'air de suivre à peu près la loi normale (en cas de difficulté pour réaliser cette partie, voir les explications détaillées correspondantes du modèle M1).

Indépendance des résidus et des autres variables

Ce modèle est bon, mais on peut peut-être l'améliorer en faisant entrer les autres variables dans le modèle. Pour cela on regarde si les résidus sont liées à d'autres variables, par exemple l'ancienneté. On représente graphiquement les résidus de M2 en fonction de l'ancienneté et le nuage de points montant indique clairement qu'il y a une liaison positive entre les 2 (voir figure 5.31). On ajoute la droite de régression (une fois le graphique sélectionné, par un clic droit, on peut lui ajouter une tendance, qu'on choisit linéaire), son équation et le R^2 correspondant, ce qui confirme clairement la liaison, et invite à faire un modèle de régression avec les deux variables explicatives, ancienneté et responsabilité.

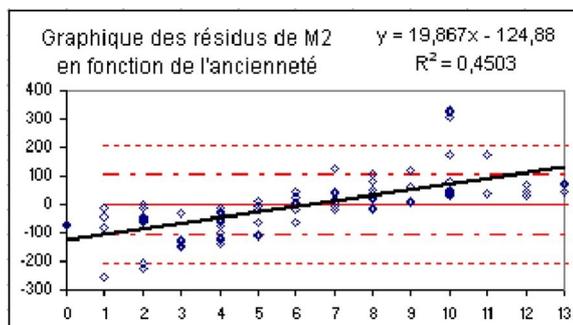


Figure 5.31 : Les résidus de M2 sont une fonction croissante de l'ancienneté

> Solution n°103 (exercice p. 148)

Une fois la copie de feuille faite, on la renomme puis on sélectionne la plage du modèle M2 (plage de cellules L4 :N8) et on sélectionne une zone de 4 colonnes et 5 lignes, comme indiqué en figure 5.33, et on corrige la formule DROITEREG, en remplaçant la plage des variables explicatives étant transformée en C2 : E99 au lieu de D2 :E99. Le modèle obtenu est $Sest.=829+20 A+471,1 c2+764,4 c3$, comme on peut le voir dans la figure 5.34. Ce qui signifie qu'avec 0 années d'ancienneté, on gagne 829 euros si la responsabilité c1, 471 euros environ de plus si la responsabilité est c2, et 746 euros de plus si elle est c3. Si l'ancienneté change, on gagne 20,346 euros de plus par année d'ancienneté, et ceci quelle que soit la responsabilité. On saisit les intitulés des lignes et colonnes de la plage T3 :W15 et on obtient les estimations détaillées de la plage T13 : W15 en écrivant dans la cellule T13 la formule $=\$O\$4+\$M\$4*(\$S13=\$M\$3)+\$L\$4*(\$S13=\$L\$3)+T\$12*\$N\$4$, et en la recopiant vers la droite et vers le bas.

SOMME ✕ ✓ ✎ =DROITEREG(B2:B99;D2:E99;VRAI;VRAI)													
	A	B	C	D	E	F	G	H	IJK	L	M	N	O
1	Num	Salair	Anc	c2	c3	Res	Se:	Ecarts					
2	1	932	4	0	0	c1	F	-30		M2			
3	2	950	8	0	0	c1	F	-12		c3	c2	Ancienneté	
4	3	978	8	0	0	c1	F	16,4		2:E99;v	475,8	962	
5	4	966	7	0	0	c1	F	4,35		26,19	26,19	14,1	
6	5	1321	4	1	0	c2	M	-116		0,902	103,5	#N/A	
7	6	993	12	0	0	c1	F	31,4		437	95	#N/A	
8	7	961	2	0	0	c1	M	-0,6		9E+06	1E+06	#N/A	

Figure 5.33 : correction du modèle M2 en un modèle M3 sous Excel contenant l'ancienneté en plus

	A	B	C	D	E	F	G	H	IJK	L	M	N	O	P	Q	R	S	T	U	V	W
1	Num	Salair	Anc	c2	c3	Res	Se:	Ecarts													
2	1	932	4	0	0	c1	F	21,2		M3											
3	2	950	8	0	0	c1	F	-42		c3	c2	Anci	Constan								
4	3	978	8	0	0	c1	F	-14		746,4	471,1	20	829								
5	4	966	7	0	0	c1	F	-5,8		19,51	19,33	2,3	18								
6	5	1321	4	1	0	c2	M	-61		0,947	76,4	###	#N/A								
7	6	993	12	0	0	c1	F	-81		561,8	94	###	#N/A								
8	7	961	2	0	0	c1	M	90,9		1E+07	5E+05	###	#N/A								
9	8	1006	10	0	0	c1	M	-27													
10	9	1421	7	1	0	c2	F	-22		Carac. des écarts à M3											
11	10	1708	6	0	1	c3	F	10,2		Moyenne			-0								
12	11	1624	5	0	1	c3	F	-53		Ecart-type			75,2								
13	12	981	7	0	0	c1	M	9,18													
14	13	1546	4	0	1	c3	F	-111		x	y = 2s	y = s	y = 0	y = -s	y = -2s						
15	14	929	3	0	0	c1	F	38,6		1	152,8	76	0	-76	-153						
16	15	909	2	0	0	c1	F	38,9		100	152,8	76	0	-76	-153						

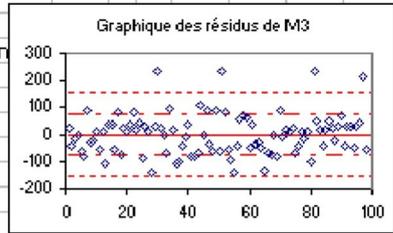


Figure 5.34 : Réalisation du modèle M3 exprimant le salaire en fonction de la responsabilité et de l'ancienneté

> **Solution n°104** (exercice p. 148)

le R² est passé de 0,906 à 0,947, ce qui indique une amélioration, mais elle n'a pas nécessairement de sens. En effet chaque fois qu'on ajoute une variable, la valeur de R² augmente généralement, au point que si on met autant ou plus de variables que d'individus, on peut obtenir un R² de 1. Il faut contrôler que l'augmentation de R² a un sens (du point de vue statistique). Nous avons proposé d'utiliser pour cela la statistique $F=R^2_{partiel}/(1-R^2_{partiel})/(q2-q1)*(n-q2)$ avec $R^2_{partiel}=(R^22-R^21)/(1-R^21)$, on peut voir les résultats du calcul dans la cellule AB10, et la probabilité p d'avoir un R²partiel aussi grand par hasard est inférieure à 0,01, ce qui nous permet de

conclure que le R^2_{partiel} est très significatif, et que le passage du modèle M2 au modèle M3 est justifié statistiquement.



Remarque

pour trouver exactement les mêmes valeurs que dans la figure 5.35, il faut veiller à écrire dans les formules les adresses des nombres si ceux-ci ne sont pas entiers, et non les valeurs qu'on voit à l'affichage, car ce sont leurs valeurs approchées (par exemple dans la cellule AB3 la formule est =reg2!L6). La valeur de p se trouve par la formule LOI.F(AB10;AB8;AB9) écrite dans AB11.

On peut aussi tester la significativité de chaque paramètre du modèle (c.à.d. qu'ils diffèrent significativement de 0). Le test peut être bilatéral (on ne sait pas s'ils risquent d'avoir un effet positif ou négatif sur le salaire), ou unilatéral. Ici on suppose que l'ancienneté, la responsabilité élevée ont un effet positif sur le salaire, c'est un test unilatéral. Pour calculer la statistique $t = a/s$, on utilise les deux premières lignes du modèle de régression. On sait que si t dépasse largement 2, et que le ddl est élevé, comme ici, la probabilité p d'avoir une valeur si grande par hasard sera très faible. Elle est calculée dans la ligne 7. Dans la cellule AF7 on a la formule =LOI.STUDENT(AF6;\$AI6;1) qu'on recopie vers la droite. On obtient que toutes les valeurs de p sont inférieures à 0,01 et donc que les 3 paramètres sont très significatifs. Il est à noter que le passage de M2 à M3 porte sur une seule variable, la variable ancienneté, et comme elle est unique, on a la relation entre la valeur t ancienneté et FM3/M2 : $t^2 = F$. Et quand on prend un test bilatéral, ce qui n'est pas le cas ici, on trouve la même probabilité avec t ou avec F . Avec un test unilatéral, on trouve une probabilité associée à F double de celle associée à t .

	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
1		Test F pour R^2_{partiel} de M3/M2								
2	n	98								
3	R^2_1	0,902				Test t pour chaque paramètre a de M3				
4	q1	3			a	746,4	471,1	20,35		
5	R^2_2	0,947			s	19,51	19,33	2,268		
6	q2	4			a/s	38,26	24,37	8,97	94	
7	R^2_{partiel}	0,461			p_{uni}	2E-59	1E-42	1E-14		
8	ddl1	1								
9	ddl2	94								
10	F	80,46								
11	p	3E-14								

Figure 5.35 : Réalisation du test de significativité du modèle M3/M2 à gauche et de tous les coefficients de l'équation de régression de M3 à droite. Tout est très significatif, les valeurs de p étant $< 0,01$

Le calcul des résidus doit être aussi mis à jour. La formule dans H2 est =B2-O\$4-M\$4*D2-L\$4*E2-C2*N\$4 et on la recopie vers le bas. A ce moment-là, dans les caractéristiques des résidus, la moyenne passe à 0 et l'écart-type devient peu différent du contenu de la cellule. Les équations des 5 droites sont mises à jour et le graphique des résidus également. Si on fait le graphique des résidus en fonction du sexe, on voit que les résidus des femmes (losanges rouges) sont plus groupés autour de l'axe que ceux des hommes (carrés bleus)., ce qui montre une différence de variance.

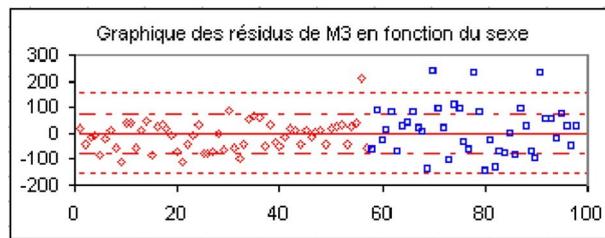


Figure 5.36: Les résidus de M3 diffèrent selon le sexe

On vérifie cela par un tableau croisé dynamique donnant les statistiques des écarts. On constate alors que non seulement les variances (ecart-type²) diffèrent mais également les moyennes. Ce qui invite à prendre en compte le sexe dans le modèle.

	Sex		Total
	F	M	
Données			
Nombre de Ecarts	57	41	98
Moyenne de Ecarts	-9,25	12,86	-0
Écartype de Ecarts	56,7	94,52	75,2

Tableau 20 : Tableau 1

> Solution n°105 (exercice p. 148)

On fait une copie de la feuille de calcul du modèle M3. On insère une colonne parmi les colonnes de variables explicatives, car la fonction droitereg du tableur exige que les colonnes de celles-ci soient contiguës. On a choisit de l'insérer en colonne F. On écrit l'intitulé dans la cellule F1, et en F2 la formule =SI(H2=F\$1;1;0), qu'on recopie vers le bas. Puis on corrige la plage M4 :P8 en M4 :Q8 : après l'avoir sélectionnée, on tape la formule =DROITEREG(B2:B99;C2:F99;VRAI;VRAI) et on enfonce les 3 touches CTRL, MAJ, Entrée. On obtient alors ce qui est dans la figure 5.37 : $S=820+20,8 A+460,6 c_2 + 732,7 c_3 + 28,23 M$. Avec ce modèle le salaire de l'homme, avec même ancienneté et même responsabilité, est supérieur de 28,23 euros à celui de la femme. La différence entre les deux modèles n'est pas grande. On a repris les tests, et on voit que le test F de passage de M3 à M4 n'est pas significatif, et que le test t de l'augmentation de 28,23€ est à la limite de la significativité ($p \approx 0,05$). On décide de ne pas inclure le sexe dans le modèle pour l'instant et on retourne au modèle M3.

	F	G	H	I	JKL	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	M	Res	Se:	Ecarts																
2	0	c1	F	29	M4							n	98							
3	0	c1	F	-37	M	c3	c2	Ancie	Constante			R ² ₁	0,947		a	28,23	732,7	461	20,83	
4	0	c1	F	-8,6		28,23	732,7	460,6	20,8	820		q1	4		s	17,53	21,13	20,3	2,269	
5	0	c1	F	0,2		17,53	21,13	20,25	2,27	18,82		R ² ₂	0,949		a/s	1,61	34,67	22,7	9,18	93
6	1	c2	M	-71		0,949	75,76	#N/A	#N/A	#N/A		q2	5		p _{uni}	0,055	3E-55	0	6E-15	
7	0	c1	F	-77		429,1	93	#N/A	#N/A	#N/A		R ² _{part}	0,027							
8	1	c1	M	71		1E+07	5E+05	#N/A	#N/A	#N/A		ddl1	1							
9	1	c1	M	-51								ddl2	93							
10	0	c2	F	-5,4								F	2,593							
11	0	c3	F	30								p	0,111							
12	0	c3	F	-33																
13	1	c1	M	-13																
14	0	c3	F	-90		x	y = 2 s	y = s	y = 0	y = - s	y = - 2s									
15	0	c1	F	47		1	151,5	75,76	0	-75,8	-152									
16	0	c1	F	47		100	151,5	75,76	0	-75,8	-152									

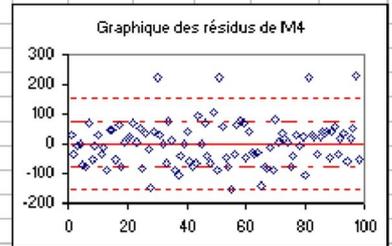


Figure 5.37 : Réalisation du modèle M4 exprimant le salaire en fonction de la responsabilité, de l'ancienneté et du sexe

> **Solution n°106** (exercice p. 149)

Pour faire le modèle M5, on ajoute deux colonnes, la colonne F contenant Ac2, obtenue en multipliant la colonne C et la colonne D membre à membre, et la colonne G contenant Ac3, obtenue en multipliant la colonne C et la colonne E membre à membre, puis on procède comme pour les modèles précédents.

	A	B	C	D	E	F	G	H	I	J	KL	M	N	O	P	Q	R	S	T	U	V
1	Nurr	Salaire	Anc	c2	c3	Ac2	Ac3	Res	Se:	Ecarts											
2	1	932	4	0	0	0	0	c1	F	-6,9	M5										
3	2	950	8	0	0	0	0	c1	F	-25	Ac3	Ac2	c3	c2	Ancien	Constante				n	98
4	3	978	8	0	0	0	0	c1	F	2,69	58,32	37,54	422,9	221,2	9,106	902,5				R ² ₁	0,947
5	4	966	7	0	0	0	0	c1	F	-0,2	2,158	2,189	13,24	15,8	0,791	6,039				R ² ₂	0,995
6	5	1321	4	1	0	4	0	c2	M	10,8	0,995	23,31	#N/A	#N/A	#N/A	#N/A				q2	6
7	6	993	12	0	0	0	0	c1	F	-19	3805	92	#N/A	#N/A	#N/A	#N/A				R ² _{part}	0,909
8	7	961	2	0	0	0	0	c1	M	40,3	1E+07	49989	#N/A	#N/A	#N/A	#N/A				ddl1	2
9	8	1006	10	0	0	0	0	c1	M	12,5										ddl2	92
10	9	1421	7	1	0	7	0	c2	F	-29										F	458,9
11	10	1708	6	0	1	0	6	c3	F	-22		Ac3	Ac2	c3	c2	Ancienneté				p	1E-48
12	11	1624	5	0	1	0	5	c3	F	-39	a	58,32	37,54	422,9	221,2	9,106					
13	12	981	7	0	0	0	0	c1	M	14,8	s	2,158	2,189	13,24	15,8	0,791					
14	13	1546	4	0	1	0	4	c3	F	-49	a/s	27,03	17,15	31,95	13,99	11,52				92	
15	14	929	3	0	0	0	0	c1	F	-0,8	p _{uni}	8E-46	1E-30	7E-52	8E-25	8E-20					
16	15	909	2	0	0	0	0	c1	F	-12											
17	16	997	11	0	0	0	0	c1	F	-5,6											
18	17	1543	8	1	0	8	0	c2	M	46,2											
19	18	1585	4	0	1	0	4	c3	M	-10											
20	19	958	5	0	0	0	0	c1	M	10											
21	20	961	6	0	0	0	0	c1	F	3,9	x	y = 2 s	y = s	y = 0	y = - s	y = - 2s					
22	21	1480	7	1	0	7	0	c2	M	29,9	1	46,62	23,31	0	-23,31	-46,62					
23	22	1544	8	1	0	8	0	c2	M	47,2	100	46,62	23,31	0	-23,31	-46,62					
24	23	1459	7	1	0	7	0	c2	M	8,87											
25	24	916	2	0	0	0	0	c1	F	-4,7											
26	25	1008	13	0	0	0	0	c1	F	-13		Ancien	9,106	46,64	67,43						
27	26	878	1	0	0	0	0	c1	F	-34	x	c1	c2	c3							
28	27	1041	10	0	0	0	0	c1	M	47,5		0	902,5	1124	1325						

Figure 5.39 : Réalisation du modèle M5 exprimant le salaire en fonction de la responsabilité, de l'ancienneté avec interaction entre responsabilité et ancienneté

Le modèle s'écrit $S_{est.} = 902,5 + 9,1 A + 221,2 c_2 + 422,9 c_3 + 37,4 A c_2 + 58,32 A c_3$.

Quand $c_1 = 1$, tous les termes contenant c_2 et c_3 s'annulent, on obtient

Sest.=902,5 + 9,1 A

Quand $c_2 = 1$, tous les termes contenant c_3 s'annulent, on obtient Sest.= (902,5 + 221,2) + (9,1 + 37,4) A

Quand $c_3 = 1$, tous les termes contenant c_2 s'annulent, on obtient Sest.= (902,5 + 422,9) + (9,1 + 58,32) A

C'est ce qui est calculé dans les lignes 25 et 27. La ligne 27 contient les constantes des 3 droites, et la ligne 25 contient les pentes des 3, avec les valeurs exactes car ce sont des adresses qui sont dans les formules, et non les valeurs approchées qu'on vient d'écrire dans les lignes précédentes. On insère alors dans le nuage de points les 3 droites de la page 026 :R28.

> Solution n°107 (exercice p. 149)

Les résidus sont calculés dans la colonne J.

La formule en J2 est =B2-S\$4-Q\$4*D2-P\$4*E2-C2*R\$4-O\$4*F2-N\$4*G2.

On voit que l'apport de M5 par rapport à M3 est très significatif (cellule V10, $p < 0,01$), et que chaque interaction l'est aussi (N8 et O8, $p < 0,01$). Si on dessine les résidus en fonction du sexe, on réalise que quasiment tous les résidus des hommes sont positifs alors que ceux des femmes sont négatifs (voir figure 5.40).

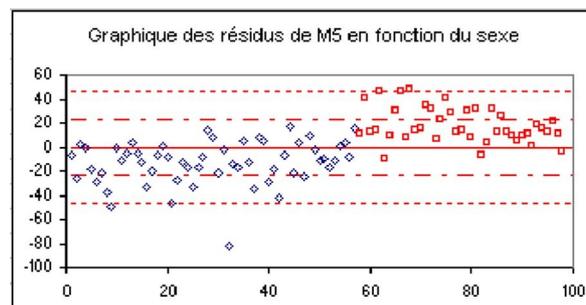


Figure 5.40: Les résidus de M5 diffèrent selon le sexe

On peut voir le modèle M6 dans la figure 5.41. Son apport par rapport à M5 est significatif, et tous les paramètres de la régression le sont. On ne va pas plus loin, car les données ont été construites artificiellement à partir de cette équation, avec ajout d'écart suivant la loi normale. Et on a retrouvé le modèle progressivement.

> **Solution n°115** (exercice p. 153)

Les moyennes mobiles donnent une série beaucoup plus alignée. Un ajustement linéaire peut être envisagé.

> **Solution n°116** (exercice p. 153)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°117** (exercice p. 153)

Le coefficient de corrélation linéaire est plus proche de 1, l'ajustement est meilleur. Les données peuvent plus facilement être assimilées à une droite.

> **Solution n°118** (exercice p. 154)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°119** (exercice p. 154)

a. Il s'agit d'appliquer la formule précédente $Y_t = f_{t+1} + S_t$, avec $t = 49, 50$ et 51 et S_t égal respectivement S_1, S_2, S_3

> **Solution n°120** (exercice p. 154)

b. A plus long terme (au-delà d'une année), les estimations ne sont pas fiables car les erreurs se cumuleront.

> **Solution n°121** (exercice p. 154)

Consulter la version en ligne du module pour voir la version animée de la solution.

> **Solution n°122** (exercice p. 154)

Consulter la version en ligne du module pour voir la version animée de la solution.

Solution des Quiz



> Solution n°1 (quiz p. 22)

- .xsl
- .txt
- .bmp
- aucun de ceux-là

> Solution n°2 (quiz p. 51)

Après validation, quelle est la valeur obtenue dans cette cellule : **4**
On recopie alors cette formule vers le bas. Qu'obtient-on dans G4 : **3**
On recopie ensuite cette formule à droite. Qu'obtient-on dans H3 : **5**

> Solution n°3 (quiz p. 51)

- =A4+\$B4+C4
- =A\$4+B\$4+C\$4
- =somme(A4:C4)
- =A\$4+\$B4+\$C\$4
- aucune de celles-là

> Solution n°4 (quiz p. 52)

- Vous tapez 2 dans les cellules B2, B3 et B4 puis vous sélectionnez à l'aide de la souris la plage A2 :A4 à B2 :B4 et vous tirez vers la gauche, l'ordinateur reconnaîtra la suite de nombre et inscrira le nombre 3 dans les cellules C2 à C4.
- Vous sélectionnez la plage de A2 à A4 et vous tirez vers la gauche, l'ordinateur écrira une suite de nombre 1,2,3 dans les cellules B2 à B4 et C2 à C4.
- Vous écrivez le nombre 2 dans la colonne C2, puis à l'aide de souris, vous sélectionnez la cellule C2 puis vous tirez vers le bas. Ensuite vous sélectionnez la plage A2 :A4 à B2 :B4 et vous tirez vers la gauche pour remplir les cases de C2 à C4.
(la meilleure réponse est celle qui ne fait remplir qu'une cellule, que ce soit par un nombre, ou une formule, le reste étant obtenu par des manipulations de souris ou des raccourcis clavier, qu'on prendra soin de détailler).
- Vous rentrez toutes les données à la main.

> **Solution n°5** (quiz p. 52)

cellule | formule | nombre
B3 | \$A4-B2+C\$3 | **2**
B4 | \$A5-B3+C\$3 | **1**
C3 | \$A4-C2+D\$3 | **-2**
C4 | \$A5-C3+D\$3 | **-3**
D3 | \$A4-D2+E\$3 | **1**
D4 | \$A5-D3+E\$3 | **0**

cellule | formule | nombre
B3 | \$A4-B2+C\$3 | **2**
B4 | \$A4-B2+C\$3 | **2**
C3 | \$A4-B2+C\$3 | **2**
C4 | \$A4-B2+C\$3 | **2**
D3 | \$A4-B2+C\$3 | **2**
D4 | \$A4-B2+C\$3 | **2**

cellule formule nombre
B3 | \$A4-B2+C\$3 | **3**
B4 | \$A5-B3+C\$3 | **2**
C3 | \$A4-C2+D\$3 | **-3**
C4 | \$A5-C3+D\$3 | **5**
D3 | \$A4-D2+E\$3 | **1**
D4 | \$A5-D3+E\$3 | **0**

cellule formule nombre
B3 | \$A4-B2+C\$3 | **3**
B4 | \$A4-B2+C\$3 | **3**
C3 | \$A4-B2+C\$3 | **3**
C4 | \$A4-B2+C\$3 | **3**
D3 | \$A4-B2+C\$3 | **3**
D4 | \$A4-B2+C\$3 | **3**

> **Solution n°6** (quiz p. 79)

<input type="checkbox"/>	0
<input type="checkbox"/>	1
<input type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	10
<input type="checkbox"/>	30
<input type="checkbox"/>	lettre
<input checked="" type="checkbox"/>	autre <i>Une erreur se produit car il n'y a pas de signe</i>

> Solution n°7 (quiz p. 80)

<input type="checkbox"/>	0
<input type="checkbox"/>	1
<input type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input checked="" type="checkbox"/>	10 <i>la solution est, en fait 9, mais 10 est plus proche</i>
<input type="checkbox"/>	30
<input type="checkbox"/>	lettre
<input type="checkbox"/>	autre

> Solution n°8 (quiz p. 81)

<input type="checkbox"/>	0
<input type="checkbox"/>	1
<input checked="" type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	10
<input type="checkbox"/>	30
<input type="checkbox"/>	lettre
<input type="checkbox"/>	autre

> **Solution n°9** (quiz p. 82)

<input type="checkbox"/>	0
<input type="checkbox"/>	1
<input checked="" type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	10
<input type="checkbox"/>	30
<input type="checkbox"/>	lettre
<input type="checkbox"/>	autre

> **Solution n°10** (quiz p. 83)

- | | |
|-------------------------------------|--------|
| <input type="checkbox"/> | 0 |
| <input type="checkbox"/> | 1 |
| <input checked="" type="checkbox"/> | 2 |
| <input type="checkbox"/> | 3 |
| <input type="checkbox"/> | 4 |
| <input type="checkbox"/> | 5 |
| <input type="checkbox"/> | 10 |
| <input type="checkbox"/> | 30 |
| <input type="checkbox"/> | lettre |
| <input type="checkbox"/> | autre |

> Solution n°11 (quiz p. 84)

- | | |
|-------------------------------------|---|
| <input type="checkbox"/> | système n° 1 : $y + 2t = 1$ $y + z = 0$ $z + t = 2$ |
| <input type="checkbox"/> | système n° 2 : $y + t = 1$ $z + t = 0$ $2y + t = 2$ |
| <input type="checkbox"/> | système n° 3 : $y + z = 1$ $z + t = 0$ $y + 2t = 2$ |
| <input checked="" type="checkbox"/> | système n° 4 : $y + t = 1$ $y + z = 0$ $z + 2t = 2$
<i>Les équations sont dans l'ordre, mais pas les variables qui sont dans l'ordre t, y et z</i> |

> Solution n°12 (quiz p. 84)

- | | |
|-------------------------------------|---|
| <input type="checkbox"/> | méthode n°1 : taper la formule dans une cellule, puis recopier vers le bas pour avoir les 3 nombres. |
| <input checked="" type="checkbox"/> | méthode n°2 : sélectionner une zone de 3 cellules, puis taper la formule. |
| <input type="checkbox"/> | méthode n°3 : taper la formule dans une cellule, et utiliser le "copier-coller" du menu "édition ", ou les combinaisons de touches Ctrl C, Ctrl V |

Glossaire



Butineur

C'est la traduction du mot anglais « browser »

Copier/coller

L'opération « copier-coller » consiste en une duplication d'un élément qui en produit un deuxième exemplaire sans modifier quoi que ce soit au premier exemplaire. Elle se décompose en quatre étapes :

- 1) **Sélection de l'élément** dont on veut faire une copie (souvent balayage avec la souris de la zone visée, bouton gauche enfoncé, ou successions de « raccourcis-clavier »)
- 2) **Action de copie** (à l'aide d'un clic droit, ou en allant dans le menu édition, ou raccourcis-clavier)
- 3) **Sélection de l'endroit** où on veut le recopier (un simple clic gauche indiquant l'endroit où il se mettra)
- 4) **Action de collage** (à l'aide d'un clic droit, ou allant dans le menu édition, ou raccourcis-clavier). On peut aussi faire des « couper-coller » qui détruisent le premier exemplaire ou des « copier-collage spécial » qui ne copient que certains éléments du premier exemplaire.

Double-clic

L'équivalent du double-clic sous Windows, est en général le simple clic sous Linux.

Extension de fichiers

Cette habitude de notation des fichiers date de DOS (Disk Operating System), système d'exploitation prédécesseur de Windows. On ne pouvait pas avoir d'espace dans un nom de fichier, le seul caractère autre que les lettres (non accentuées) et les chiffres étant le soulignement « _ ». Avec les dernières versions de Windows, ces contraintes n'existent plus, mais il est mieux de les respecter si on veut pouvoir réutiliser ses fichiers partout sans problème.

Glisser/déposer

Cliquer sur un élément et, sans relâcher le bouton de la souris, le déplacer. Relâcher ensuite le bouton de la souris pour "déposer" l'élément en question.

Le PIB

Le PIB est le total de la valeur créée dans le pays, c'est aussi l'ensemble de la demande, c'est aussi l'ensemble des revenus versés.

Produit intérieur brut (approche production) = Valeur ajoutée au prix de base + Impôts sur les produits - Subventions sur les produits

Produit intérieur brut (approche demande) = Dépense de consommation finale

+ Formation brute de capital + Exportations de biens et services - Importations de biens et services

Produit intérieur brut (approche revenus) = Rémunérations des salariés + Excédent brut d'exploitation et revenu mixte brut + Impôts sur la production et les importations - Subventions