

Utiliser un tableur dans l'enseignement de la statistique. Pourquoi et comment ?

*Guy Mélard **

- ECARES & Faculté des Sciences sociales et politiques - Solvay Brussels School of Economics and Management, U.L.B., CP114 Avenue Franklin Roosevelt 50, B-1050 Bruxelles, Belgique, gmelard@ulb.ac.be

Préalable

- ⇒ L'auteur a l'avantage de relativement bien connaître les tableurs ainsi qu'un certain nombre de logiciels statistiques
- ⇒ Il a enseigné l'informatique et de statistique (y compris la statistique informatique) à différents niveaux universitaires sauf en première année
- ⇒ Il a employé Excel dans un livre avec CD-Rom (Mélard G., **Méthodes de prévision à court terme**, 2^e éd. (2007) après avoir participé à une communication à ce sujet (Cohen et al. 2003)

Plan de l'exposé

- ⇒ Synthèse des **critiques** relatives aux tableurs
- ⇒ En particulier différentes versions d'Excel de Microsoft critiquées — presque en vain — dans une série d'articles depuis une **quinzaine d'années**
- ⇒ Le plus récent (**Office 2007 & OpenOffice 3.0**) :
 - ✓ section de Computational and Data Analysis
 - ✓ site tenu par Heiser (2010)
- ⇒ Ces critiques seront **actualisées** pour les sorties récentes d'**Office 2010** et d'**OpenOffice.org 3.2**
- ⇒ Application au classeurs du CD-Rom de **MPCT2**

Critiques d'Excel (1)

- ⇒ Les critiques des différentes versions d'Excel (depuis 1998 pour la version de 1995 !) portaient sur
- ✓ la précision des **fonctions statistiques**
 - ✓ les **outils complémentaires**
 - ✓ le module **Solver**
 - ✓ le générateur de **nombres aléatoires**
 - ✓ les **graphiques**
- ⇒ Ce qui a déplu à la communauté statistique : pendant plusieurs années Microsoft a **ignoré** ces critiques
- ⇒ Premières améliorations (**timides**) dans Office 2003

Critiques d'Excel (2)

⇒ Le plus récent à propos d'Office 2007 et OpenOffice.org 3.0 :

- ✓ les articles dans une section d'un numéro de Computational and Data Analysis édité par McCullough (2008), en particulier l'article de Tahla Yalta (2008)
- ✓ un site très complet tenu par Heiser (2010)

⇒ Entretemps : Office 2010 et OpenOffice 3.2

Exemple : calcul de la variance

⇒ Excel 97, 2000 et 2002 calculent la variance de 3 entiers consécutifs comme suit ...

	A	B	C	D	E	F	G	H	I	J	K
1		1 000	10 000	100 000	1 000 000	10 000 000	77 490 000	77 500 000	100 000 000	770 000 000	771 000 000
2		+	+	+	+	+	+	+	+	+	+
3	0	1000	10000	100000	1000000	10000000	77490000	77500000	100000000	770000000	771000000
4	1	1001	10001	100001	1000001	10000001	77490001	77500001	100000001	770000001	771000001
5	2	1002	10002	100002	1000002	10000002	77490002	77500002	100000002	770000002	771000002
6											
7	Variance	0,666667	0,666667	0,666667	0,666667	0,666667	0,666667	0,000000	0,000000	0,000000	85,333333
8	Ecart-type	0,816497	0,816497	0,816497	0,816497	0,816497	0,816497	0,000000	0,000000	0,000000	9,237604

⇒ En cause : l'algorithme « **calculatrice** » basé sur la formule théoriquement correcte mais imprécise

$$s_n^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$$

⇒ Il a fallu attendre Excel **2003** pour que le calcul soit correct (en employant la définition)

Autres exemples

- ⇒ De façon générale, jusqu'à Excel 2002, la plupart des fonctions et procédures statistiques étaient **numériquement déficientes** en présence de grandes données
- ⇒ Optimisation non linéaire : le module Solver fournit des messages de convergence atteinte souvent erronés (McCullough & Heiser, 2008)

Conséquence

- ⇒ Une **retombée positive** du manque de réaction de Microsoft a été le développement par les statisticiens (en commençant par McCullough, 1999) d'une technologie pour juger de la précision des résultats, de manière similaire à ce qui existait pour juger de la qualité des logiciels statistiques
- ⇒ Egalement, d'**autres développeurs de tableurs** ont mis l'accent sur la qualité des fonctions statistiques, des outils complémentaires, du générateur de nombres pseudo-aléatoires et des graphiques

Exemple de résultat de T. Yalta

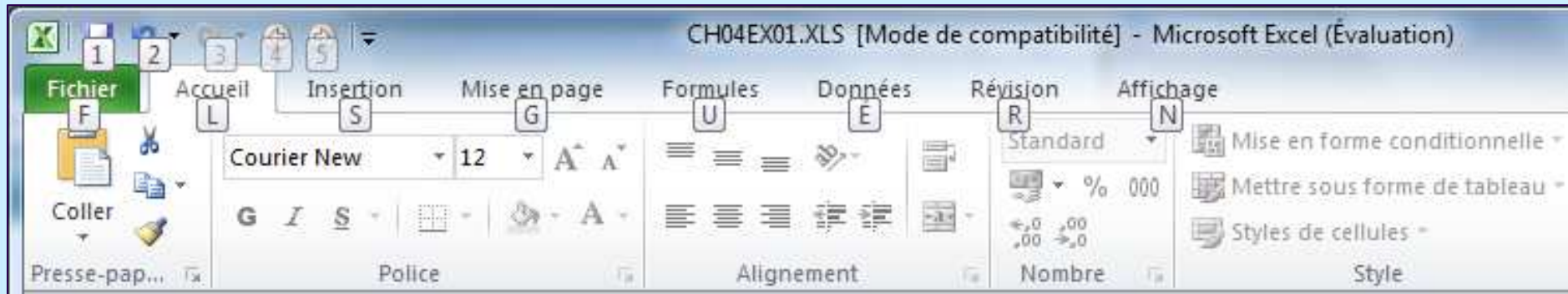
⇒ Probabilités binomiales ($n = 1030, p = 0.5, \Sigma = 1$)

k	EXACT	ELV Ed.2	EXCEL 97/2K/XP	EXCEL 2003/2007
1	8.96114E-308	0	8.95245E-308	0
2	4.61499E-305	0	Exact	0
100	1.39413E-169	0	Exact	0
200	5.45781E-92	Exact	Exact	0
300	2.91621E-42	Exact	Exact	0
390	3.18196E-15	Exact	Exact	0
391	5.24099E-15	Exact	Exact	2.05902E-15
400	3.89735E-13	Exact	Exact	3.86553E-13
410	3.19438E-11	Exact	Exact	3.19406E-11
420	1.76037E-09	Exact	Exact	Exact
500	1.83106E-01	Exact	#NUM!	Exact
550	9.86550E-01	Exact	#NUM!	Exact
575	9.99920E-01	Exact	#NUM!	Exact
589	9.99998E-01	Exact	#NUM!	Exact

⇒ Question : y a-t-il amélioration dans Office 2010 ?

Office 2010

⇒ Paru le 12 mai 2010 (en version beta depuis fin 2009)



⇒ Quelques caractéristiques

- ✓ Conserve le ruban de la version 2007 avec un onglet Fichier (appelé « Backstage »)
- ✓ Complément Solver amélioré
- ✓ Précision des fonctions améliorée + nouveaux noms
- ✓ Graphiques améliorés (nombre de points, mise en forme plus rapide, enregistrement de macros)

⇒ Qu'en est-il ?

Objet de l'étude

- ⇒ **Juger l'amélioration d'Excel 2010** du point de vue
 - ✓ de la précision des fonctions statistiques
 - ✓ de la précision des procédures statistiques (compléments)
 - ✓ des problèmes liés au générateur de nombres aléatoires
 - ✓ de la possibilités des graphiques statistiques
- ⇒ Parallèlement **examiner OpenOffice.org Calc 3.2**
- ⇒ A titre d'exemple, étudier l'utilisation des **classeurs de MPCT2 (2007)**
- ⇒ On n'examinera pas ici les améliorations du Solver

Précision des fonctions statistiques

Loi binomiale (T. Yalta, table 2)

Table 2	=LOI.BINOMIALE.N(k;1030;0,5;VRAI)		Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
k							
1			0.00000E+00	0.00000E+00	0.00000E+00	8.96114E-308	8.96114E-308
2			0.00000E+00	0.00000E+00	0.00000E+00	4.61499E-305	4.61499E-305
100			0.00000E+00	0.00000E+00	1.39413E-169	1.39413E-169	1.39413E-169
200			0.00000E+00	0.00000E+00	5.45781E-92	5.45781E-92	5.45781E-92
300			0.00000E+00	0.00000E+00	2.91621E-42	2.91621E-42	2.91621E-42
390			0.00000E+00	0.00000E+00	3.18196E-15	3.18196E-15	3.18196E-15
391			2.05902E-15	2.05902E-15	5.24099E-15	5.24099E-15	5.24099E-15
400			3.86553E-13	3.86553E-13	3.89735E-13	3.89735E-13	3.89735E-13
410			3.19406E-11	3.19406E-11	3.19438E-11	3.19438E-11	3.19438E-11
420			1.76037E-09	1.76037E-09	1.76037E-09	1.76037E-09	1.76037E-09
500			1.83106E-01	1.83106E-01	1.83106E-01	1.83106E-01	1.83106E-01
550			9.86550E-01	9.86550E-01	9.86550E-01	9.86550E-01	9.86550E-01
575			9.99920E-01	9.99920E-01	9.99920E-01	9.99920E-01	9.99920E-01
589			9.99998E-01	9.99998E-01	9.99998E-01	9.99998E-01	9.99998E-01

Notes :

1. On compare Excel 2003, 2007, 2010 beta, OpenOffice.org Calc 3.0 et 3.2
2. Le fond vert indique un résultat exact (à la précision affichée)
3. Ici Excel s'améliore

Loi de Poisson (T. Yalta, table 4)

Table 4 =LOI.POISSON.N(k;lambda;sigma)			Microsoft Office			OpenOffice.org	
k	lambda	sigma	Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
0	200	FALSE	0.00000E+00	0.00000E+00	1.38390E-87	1.38390E-87	1.38390E-87
103	200	FALSE	0.00000E+00	0.00000E+00	1.41720E-14	1.41720E-14	1.41720E-14
104	200	FALSE	2.72538E-14	2.72538E-14	2.72538E-14	2.72538E-14	2.72538E-14
133	200	FALSE	1.01322E-07	1.01322E-07	1.01322E-07	1.01322E-07	1.01322E-07
134	200	FALSE	1.51227E-07	1.51227E-07	1.51227E-07	1.51227E-07	1.51227E-07
200	200	FALSE	2.81977E-02	2.81977E-02	2.81977E-02	2.81977E-02	2.81977E-02
314	200	FALSE	2.23568E-14	2.23568E-14	2.23568E-14	2.23568E-14	2.23568E-14
315	200	FALSE	0.00000E+00	0.00000E+00	1.41948E-14	1.41948E-14	1.41948E-14
400	200	FALSE	0.00000E+00	0.00000E+00	5.58069E-36	5.58069E-36	5.58069E-36
900	200	FALSE	0.00000E+00	0.00000E+00	1.73230E-286	1.73230E-286	1.73230E-286
1.E+03	1.E+03	TRUE	5.08409E-01	5.08409E-01	5.08409E-01	#VALUE!	5.08409E-01
1.E+05	1.E+05	TRUE	6.79499E-01	6.79499E-01	5.00841E-01	#VALUE!	5.00841E-01
1.E+07	1.E+07	TRUE	9.52000E-01	9.52000E-01	5.00084E-01	#VALUE!	Err:523
1.E+09	1.E+09	TRUE	9.94979E-01	9.94979E-01	5.00008E-01	#VALUE!	Err:523

Notes :

1. Excel 2010 fait mieux que ses prédécesseurs en donnant le résultat exact
2. Il fait même mieux qu'OpenOffice.org Calc 3.0 et 3.2
3. OpenOffice.org Calc 3.2 est légèrement meilleurs que la version 3.0

Loi gamma (T. Yalta, table 5)



Table 5		=LOI.GAMMA.N(x;alpha;beta;VRAI)				
x	alpha	Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
0.1	0.1	#NUM!	#NUM!	0.827552	0.827552	0.827552
0.2	0.1	0.879419	0.879419	0.879420	0.879420	0.879420
0.2	0.2	0.764434	0.764434	0.764435	0.764435	0.764435
0.3	0.2	0.816527	0.816527	0.816527	0.816527	0.816527
0.3	0.3	0.726957	0.726957	0.726957	0.726957	0.726957
0.4	0.3	0.776380	0.776380	0.776381	0.776381	0.776381
0.4	0.4	0.701441	0.701441	0.701441	0.701441	0.701441
0.5	0.4	0.748018	0.748018	0.748019	0.748019	0.748019
0.5	0.5	0.682689	0.682689	0.682689	0.682689	0.682689
0.6	0.5	0.726678	0.726678	0.726678	0.726678	0.726678

Notes :

1. Excel 2010 fait mieux que ses prédécesseurs en donnant le résultat exact
2. Il atteint la précision d'OpenOffice.org Calc 3.0 et 3.2

Loi normale (T. Yalta, table 6)

Table 6	=LOI.NORMALE.INVERSE.N(p;0;1)		Microsoft Office			OpenOffice.org	
p			Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
5.00E-01			-1.39E-16	-1.39E-16	0.000000	0.000000	0.000000
1.00E-01			-1.28155	-1.28155	-1.28155	-1.28155	-1.28155
1.00E-02			-2.32635	-2.32635	-2.32635	-2.32635	-2.32635
1.00E-03			-3.09023	-3.09023	-3.09023	-3.09023	-3.09023
1.00E-04			-3.71902	-3.71902	-3.71902	-3.71902	-3.71902
1.00E-05			-4.26489	-4.26489	-4.26489	-4.26489	-4.26489
1.00E-06			-4.75342	-4.75342	-4.75342	-4.75342	-4.75342
1.00E-07			-5.19934	-5.19934	-5.19934	-5.19934	-5.19934
1.00E-15			-7.94135	-7.94135	-7.94135	-7.94135	-7.94135
1.00E-16			-8.22208	-8.22208	-8.22208	-8.22208	-8.22208
1.00E-100			-21.2735	-21.2735	-21.2735	-21.2735	-21.2735
1.00E-197			-29.9763	-29.9763	-29.9763	-29.9763	-29.9763
1.00E-198			-30.0000	-30.0000	-30.0529	-30.0529	-30.0529
1.00E-300			-30.0000	-30.0000	-37.0471	-37.0471	-37.0471

Notes :

1. Excel 2010 fait mieux que ses prédécesseurs en donnant le résultat exact
2. Il atteint la précision d'OpenOffice.org Calc 3.0 et 3.2

Loi χ^2 (T. Yalta, table 7)

Table 7		=LOI.KHIDEUX.INVERSE(p;n)		Microsoft Office		OpenOffice.org	
p	n	Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2	
2.00E-01	1	1.64238	1.64238	1.64237	1.64237	1.64237	
2.00E-01	5	7.28928	7.28928	7.28928	7.28928	7.28928	
1.00E-01	1	2.70554	2.70554	2.70554	2.70554	2.70554	
1.00E-01	5	9.23636	9.23636	9.23636	9.23636	9.23636	
1.00E-05	1	19.5114	19.5114	19.5114	19.5114	19.5114	
1.00E-05	5	30.8562	30.8562	30.8562	30.8562	30.8562	
1.00E-06	1	23.9281	23.9281	23.9281	23.9281	23.9281	
1.00E-06	5	35.8882	35.8882	35.8882	35.8882	35.8882	
1.00E-07	1	#NUM!	#NUM!	28.3740	28.3740	28.3740	
1.00E-07	5	#NUM!	#NUM!	40.8630	40.8630	40.8630	
1.00E-12	1	#NUM!	#NUM!	50.8442	50.8441	50.8441	
1.00E-12	5	#NUM!	#NUM!	65.2387	65.2386	65.2386	
4.80E-01	778	#NUM!	#NUM!	779.312	779.312	779.312	
5.00E-01	780	779.333	779.333	779.333	779.333	779.333	
5.20E-01	782	#NUM!	#NUM!	779.353	779.353	779.353	

Notes :

1. Excel 2010 fait beaucoup mieux que ses prédécesseurs
2. Il se rapproche de la perfection et d'OpenOffice.org Calc 3.0 et 3.2

Loi beta (T. Yalta, table 8)



Table 8	=BETA.INVERSE.N(p;5;2;0;1)					
	p	Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
	1.00E-01	4.89684E-01	4.89684E-01	4.89684E-01	4.89684E-01	4.89684E-01
	1.00E-02	2.94315E-01	2.94315E-01	2.94314E-01	2.94314E-01	2.94314E-01
	1.00E-03	1.81396E-01	1.81396E-01	1.81386E-01	1.81386E-01	1.81386E-01
	1.00E-04	1.13037E-01	1.13037E-01	1.12969E-01	1.12969E-01	1.12969E-01
	1.00E-05	7.03125E-02	7.03125E-02	7.07371E-02	7.07371E-02	7.07371E-02
	1.00E-06	4.29688E-02	4.29688E-02	4.44270E-02	4.44270E-02	4.44270E-02
	1.00E-07	3.12500E-02	3.12500E-02	2.79523E-02	2.79523E-02	2.79523E-02
	1.00E-08	3.12500E-02	3.12500E-02	1.76057E-02	1.76057E-02	1.76057E-02
	1.00E-09	3.12500E-02	3.12500E-02	1.10963E-02	1.10963E-02	1.10963E-02
	1.00E-10	3.12500E-02	3.12500E-02	6.99645E-03	6.99645E-03	6.99645E-03
	1.00E-11	3.12500E-02	3.12500E-02	4.41255E-03	4.41255E-03	4.41255E-03
	1.00E-12	3.12500E-02	3.12500E-02	2.78337E-03	2.78337E-03	2.78337E-03
	1.00E-13	3.12500E-02	3.12500E-02	1.75589E-03	1.75589E-03	1.75589E-03
	1.00E-100	3.12500E-02	3.12500E-02	6.98827E-21	1.00000E-08	6.98827E-21

Notes :

1. Excel 2010 fait vraiment beaucoup mieux que ses prédécesseurs
2. Il fait jeu égal avec OpenOffice.org Calc 3.2 (qui a corrigé une imprécision de la version 3.0)

Loi de Student (T. Yalta, table 9)

Table 9		=LOI.STUDENT.INVERSE.N(2*p;1)			Microsoft Office		OpenOffice.org	
	p	Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2		
	2.00E-01	1.37638E+00	1.37638E+00	1.37638E+00	1.37638E+00	1.37638E+00		
	1.00E-01	3.07768E+00	3.07768E+00	3.07768E+00	3.07768E+00	3.07768E+00		
	1.00E-02	3.18205E+01	3.18205E+01	3.18205E+01	3.18205E+01	3.18205E+01		
	1.00E-03	3.18309E+02	3.18309E+02	3.18309E+02	3.18309E+02	3.18309E+02		
	1.00E-04	3.18310E+03	3.18310E+03	3.18310E+03	3.18310E+03	3.18310E+03		
	1.00E-05	3.18310E+04	3.18310E+04	3.18310E+04	1.00000E+04	3.18310E+04		
	1.00E-06	3.18310E+05	3.18310E+05	3.18310E+05	1.00000E+04	3.18310E+05		
	1.00E-07	3.18310E+06	3.18310E+06	3.18310E+06	1.00000E+04	3.18310E+06		
	1.00E-08	1.00000E+07	1.00000E+07	3.18310E+07	1.00000E+04	3.18310E+07		
	1.00E-09	1.00000E+07	1.00000E+07	3.18310E+08	1.00000E+04	3.18310E+08		
	1.00E-10	1.00000E+07	1.00000E+07	3.18310E+09	1.00000E+04	3.18310E+09		
	1.00E-11	1.00000E+07	1.00000E+07	3.18310E+10	1.00000E+04	3.18310E+10		
	1.00E-12	1.00000E+07	1.00000E+07	3.18310E+11	1.00000E+04	3.18310E+11		
	1.00E-13	1.00000E+07	1.00000E+07	3.18310E+12	1.00000E+04	3.18310E+12		
	1.00E-100	1.00000E+07	1.00000E+07	3.18310E+99	1.00000E+04	3.18310E+99		

Notes :

1. Excel 2010 fait vraiment mieux que ses prédécesseurs
2. Il fait jeu égal avec OpenOffice.org Calc 3.2 (qui a corrigé des erreurs de la version 3.0)

Loi de Fisher-Snedecor (T. Yalta, table 10)

Table 10	=INVERSE.LOI.F.N(p;1;1)	Microsoft Office			OpenOffice.org	
p		Excel 2003	Excel 2007	Excel 2010	OOo Calc 3.0	OOo Calc 3.2
5.00E-01		1	1	1	0.999999997	1
4.00E-01		1.89443E+00	1.89443E+00	1.89443E+00	1.89443E+00	1.89443E+00
2.00E-01		9.47214E+00	9.47214E+00	9.47214E+00	9.47214E+00	9.47214E+00
1.00E-01		3.98635E+01	3.98635E+01	3.98635E+01	3.98635E+01	3.98635E+01
1.00E-02		4.05218E+03	4.05218E+03	4.05218E+03	4.05218E+03	4.05218E+03
1.00E-03		4.05284E+05	4.05284E+05	4.05284E+05	4.05284E+05	4.05284E+05
1.00E-04		4.05285E+07	4.05285E+07	4.05285E+07	4.05285E+07	4.05285E+07
1.00E-05		1.00000E+09	1.00000E+09	4.05285E+09	1.00000E+08	4.05285E+09
1.00E-06		1.00000E+09	1.00000E+09	4.05292E+11	1.00000E+08	4.05285E+11
1.00E-12		1.00000E+09	1.00000E+09	#NUM!	1.00000E+08	4.05285E+23
1.00E-13		1.00000E+09	1.00000E+09	#NUM!	1.00000E+08	4.05285E+25
1.00E-100		1.00000E+09	1.00000E+09	#NUM!	1.00000E+08	4.05285E+199

Notes :

1. Excel 2010 fait un peu mieux que ses prédécesseurs
2. Il n'atteint pas la précision d'OpenOffice.org Calc 3.2 (qui a corrigé des erreurs de la version 3.0)

Premières conclusions

- ⇒ **Amélioration significative** d'Excel 2010, presque aussi bon qu'OpenOffice.org Calc 3.2
- ⇒ Ajout de nouvelles fonctions avec **noms plus explicites** mais problèmes d'incompatibilité quand on revient à une ancienne version
- ⇒ Noms français différents des noms anglais mais toujours **conversion automatique** quand on passe d'une version linguistique à une autre
- ⇒ Pour des raisons de compatibilité, maintien des anciens noms

Générateur de nombres aléatoires

Générateur(s) de nombres aléatoires



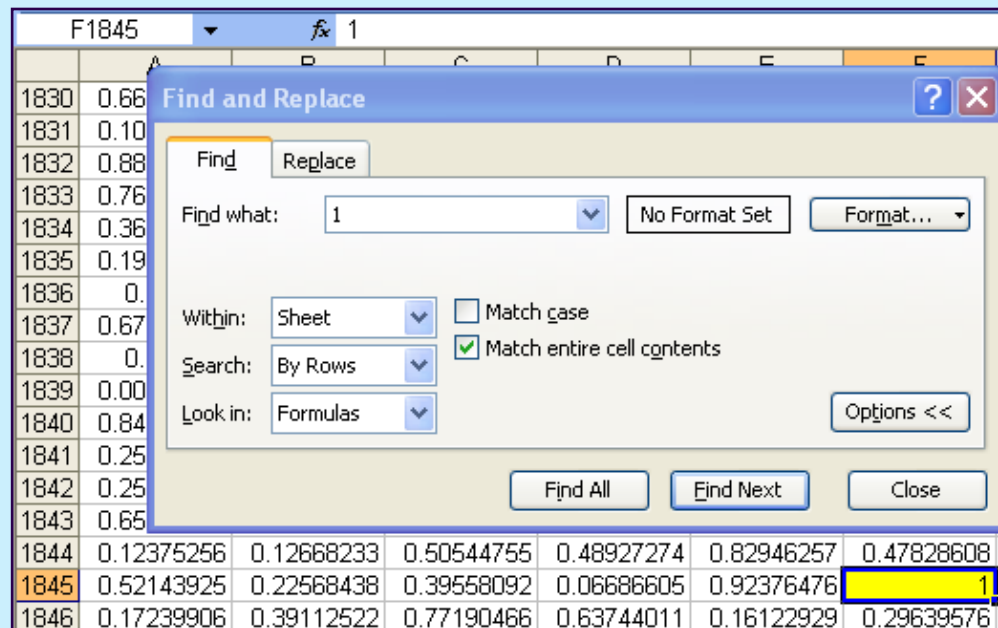
- ⇒ Il y a 2 (ou 3) générateurs de nombres pseudo-aléatoires dans Excel 2007
 - A. Le générateur invoqué par la fonction ALEA() et les autres fonctions de générateur de lois statistiques
 - B. Le générateur fourni par l'outil complémentaire Outils > Analyse de données > Génération de nombres aléatoires
 - C. Le générateur du Visual Basic for Application
- ⇒ Les algorithmes de A et B sont celui de Wichmann-Hill
- ⇒ McCullough (2008) a montré que B peut fournir facilement 0 (un peu surprenant) ou 1 (très étonnant)* ce qui implique une **erreur d'implantation de l'algorithme**
- ⇒ L'algorithme C est mauvais (L'Ecuyer & Simard, 2007)

* Voir page suivante

Et Excel 2010 ?



- ⇒ Microsoft a déclaré avoir amélioré la fonction ALEA dans Excel 2010
- ⇒ Il faudrait une étude fine pour examiner les qualités du nouveau générateur A
- ⇒ Le générateur B de l'outil complémentaire est **toujours inchangé**, donc avec les mêmes défauts
- ⇒ Pour le générateur C du VBA, la question est ouverte



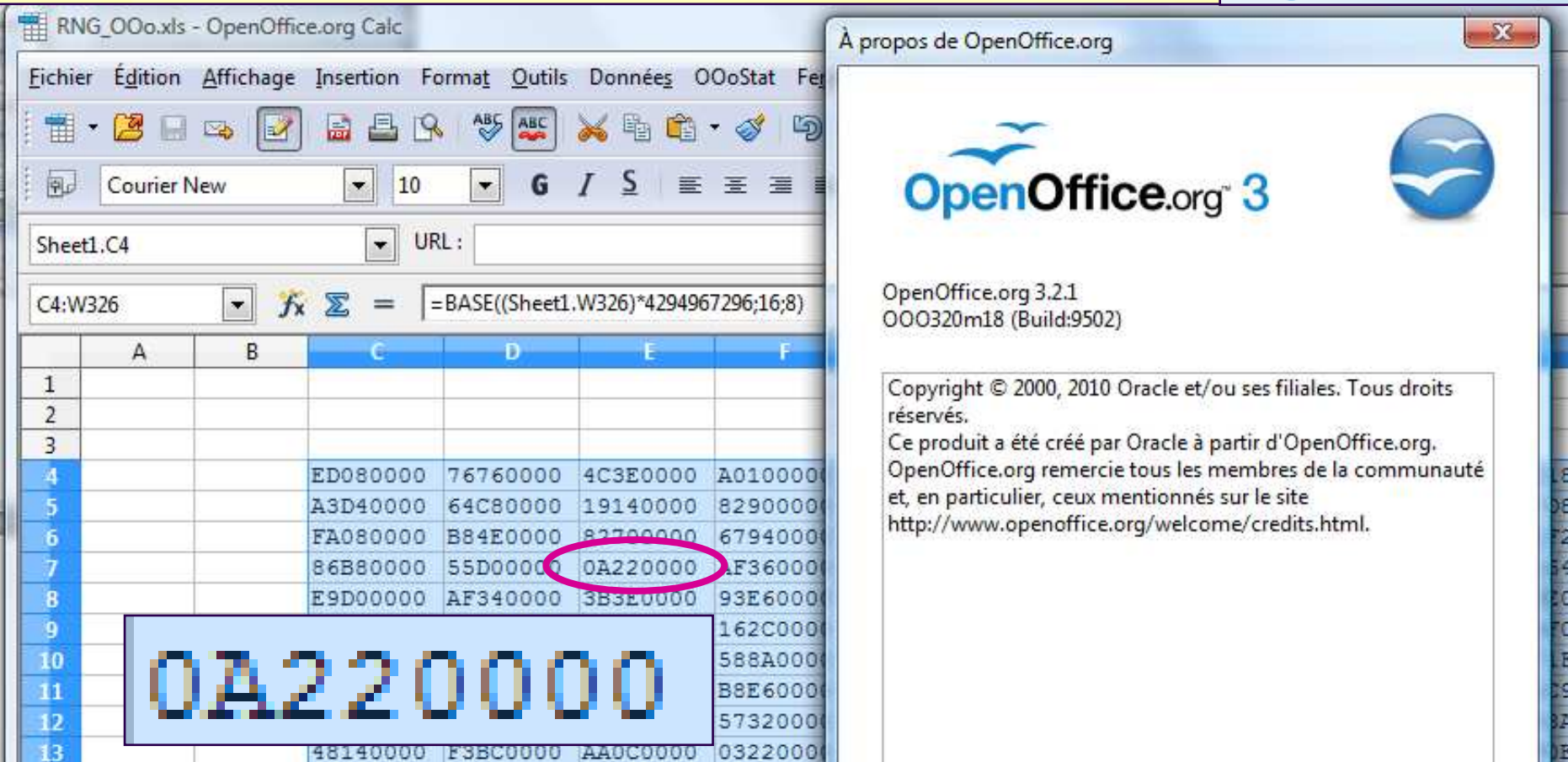
Et OpenOffice.org Calc 3.2 ?



- ⇒ Il n'y a pas eu d'étude du générateur de nombres pseudo-aléatoires d'OpenOffice.org Calc
- ⇒ Cependant les messages dans les forums semblent très critiques
- ⇒ J'ai repris l'expérience de Cliff Crabtree sur OOo 3.2 avec les mêmes constatations. La période du générateur ne serait que de 2^{15} (contre 2^{43} pour Wichmann-Hill bien implémenté) (à vérifier)

[http://www.cliffcrabtree.com/index.php?option=com_content
&view=article&id=14:how-random-is-random&catid=2:joomla-notes](http://www.cliffcrabtree.com/index.php?option=com_content&view=article&id=14:how-random-is-random&catid=2:joomla-notes)

Résultat (!)



The screenshot shows the OpenOffice Calc application window with the following details:

- File name: RNG_OOo.xls - OpenOffice.org Calc
- Formula bar: `=BASE((Sheet1.W326)*4294967296;16;8)`
- Spreadsheet grid (rows 1-13, columns A-F):

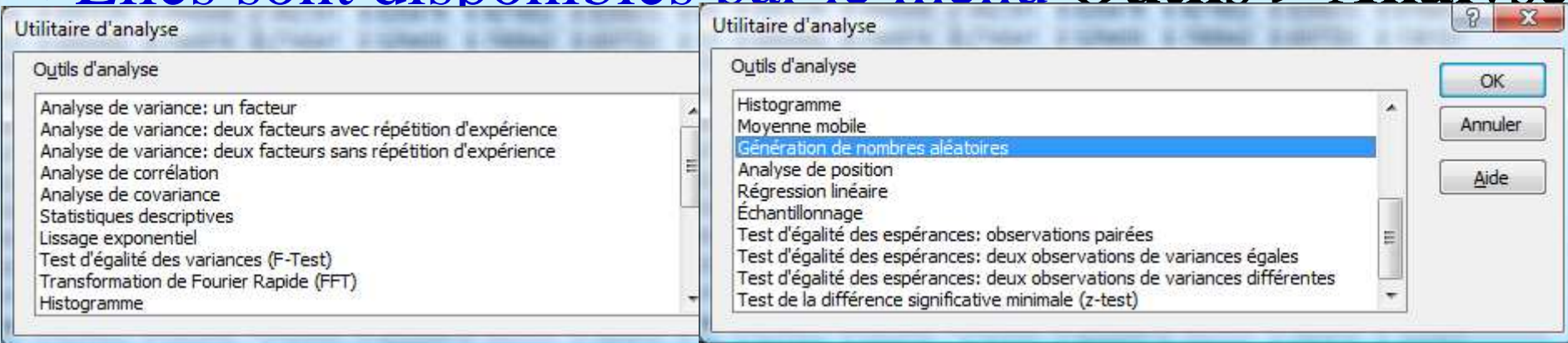
	A	B	C	D	E	F
1						
2						
3						
4			ED080000	76760000	4C3E0000	A0100000
5			A3D40000	64C80000	19140000	82900000
6			FA080000	B84E0000	82700000	67940000
7			86B80000	55D00000	0A220000	AF360000
8			E9D00000	AF340000	3B3E0000	93E60000
9						162C0000
10						588A0000
11						B8E60000
12						57320000
13			48140000	F3BC0000	AA0C0000	03220000
- Dialog box: "À propos de OpenOffice.org" (About OpenOffice.org) showing version 3.2.1 (Build:9502) and copyright information.

⇒ Après multiplication par 2^{32} , les nombres en hexadécimal se terminent par « 0000 » d'où précision de 32 bits seulement.

Procédures statistiques

Procédures statistiques en Excel 2007

⇒ Elles sont disponibles par le menu **Outils > Analyse**



⇒ Pas de test sur une moyenne ou une variance mais **tests d'égalité de moyennes** dans différents cas ou d'égalité de variance

⇒ Peu d'utilisateurs les emploient heureusement sauf

Tests d'égalité des moyennes & variances

⇒ Les fonctions TTEST (3 cas) et FTEST fonctionnent correctement (comme dans la version 2003)



⇒ Le test d'égalité des moyennes, cas des observations appariées, est toujours erroné dans le cas de données manquantes (cellules vides)

	Set 1		
	A	B	
1	3	2	
2	4	2	
3	3	2	
4	2	3	
5	2	3	
6	4	3	
19	4	2	

TTEST(A,B,2,1) 0.036878

	Variable 1	Variable 2
Moyenne	3.21052632	2.57894737
Variance	0.61988304	0.47953216
Observations	19	19
Coefficient de corrélation de Pearson	-0.17699808	
Différence hypothétique des moyenne	0	
Degré de liberté	18	
Statistique t	1.71428571	
P(T<=t) bilatéral	0.10364302	

N.B. Cette constatation^(*) due à Gary Simon date de 2000 !

⇒ Le test d'égalité des moyennes, cas d'observations indépendantes et variances inégales, est parfois erroné parce que le nombre de degrés de liberté de la méthode de Welsch est arrondi (contrairement à la fonction TTEST où il est fractionnaire)

^(*) <https://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0012&L=ASSUME&F=&S=&X=31BDFD6E4E1E0E4F11&P=4325>

Outil complémentaire de régression



- ⇒ Toujours pas plus de 16 variables explicatives !
- ⇒ Pas d'amélioration dans les graphiques
- ⇒ Mais la fonction (mal-nommée) DROITEREG l'accepte
- ⇒ Sa documentation est toujours très pauvre

Et OpenOffice.org Calc 3.2 ?



⇒ Il n'y a pas d'outils complémentaires dans OpenOffice.org Calc

⇒ On peut employer ...

- ✓ ODStatistics de Walter Garcia-Fontes

- ✓ OOoStat de David Hitchcock

- ✓ Projet «R and Calc» en développement. Nécessite

 - ✓ OpenOffice.org 2.1

 - ✓ Java 5.0

 - ✓ RServe 0.4.7

ODStatistics



odstats.ods - OpenOffice.org Calc

Fichier Édition Affichage Insertion Format Outils Données OOoStat Fenêtre Aide

ODStatistics 0.3.13 - URL :

A1 ODStatistics 0.3.13 -

ODStatistics 0.3.13 - Data Analysis with OpenOffice Calc - © Walter García-Fontes 2009-2010

Help English License

Data Type						
Variables	Cross Section				Time Series	
Numerical	1 Numerical	1Num	2 Numerical	2Num	Time Series	TimeSer
Categorical	1 Categorical	1Cat	2 Categorical	2Cat	Index Numbers	Indices
Combinations	1 Numerical 1 Categorical	1Num1Cat	2 Numerical 1 Categorical	2Num1Cat		

Other worksheets

Normal Probability NormProb

Transformation Transform

ODStatistics / Localization

Feuille 1 / 2 Standard INS STD * Somme=0 100%

OOoStat



Fichier Édition Affichage Insertion Format Outils Données **OOoStat** Fenêtre Aide

- Applications
 - Histogram
 - Patterned
 - Latin Square
 - Stack
- Basic Stats
- Multivariate Statistics

OOoStat Statistics Macros for OpenOffice - Version 0.5

Feuille 1 / 16 Standard STD Somme=0 100%

Fichier Édition Affichage Insertion Format Outils Données **OOoStat** Fenêtre Aide

- Applications
- Basic Stats
 - 1-Way ANOVA
 - 2-Way ANOVA
 - Chi Squared
 - Correlation_Regression
 - Multiple Regression
- Multivariate Statistics

OOoStat Statistics Macros for OpenOffice - Version 0.5

Fichier Édition Affichage Insertion Format Outils Données **OOoStat** Fenêtre Aide

- Applications
- Basic Stats
- Multivariate Statistics
 - Principal Components
 - Principal Components [Covariance]
 - Correspondence Analysis
 - Multidimensional Scaling
 - MANOVA

OOoStat Statistics Macros for OpenOffice - Version 0.5

Graphiques statistiques

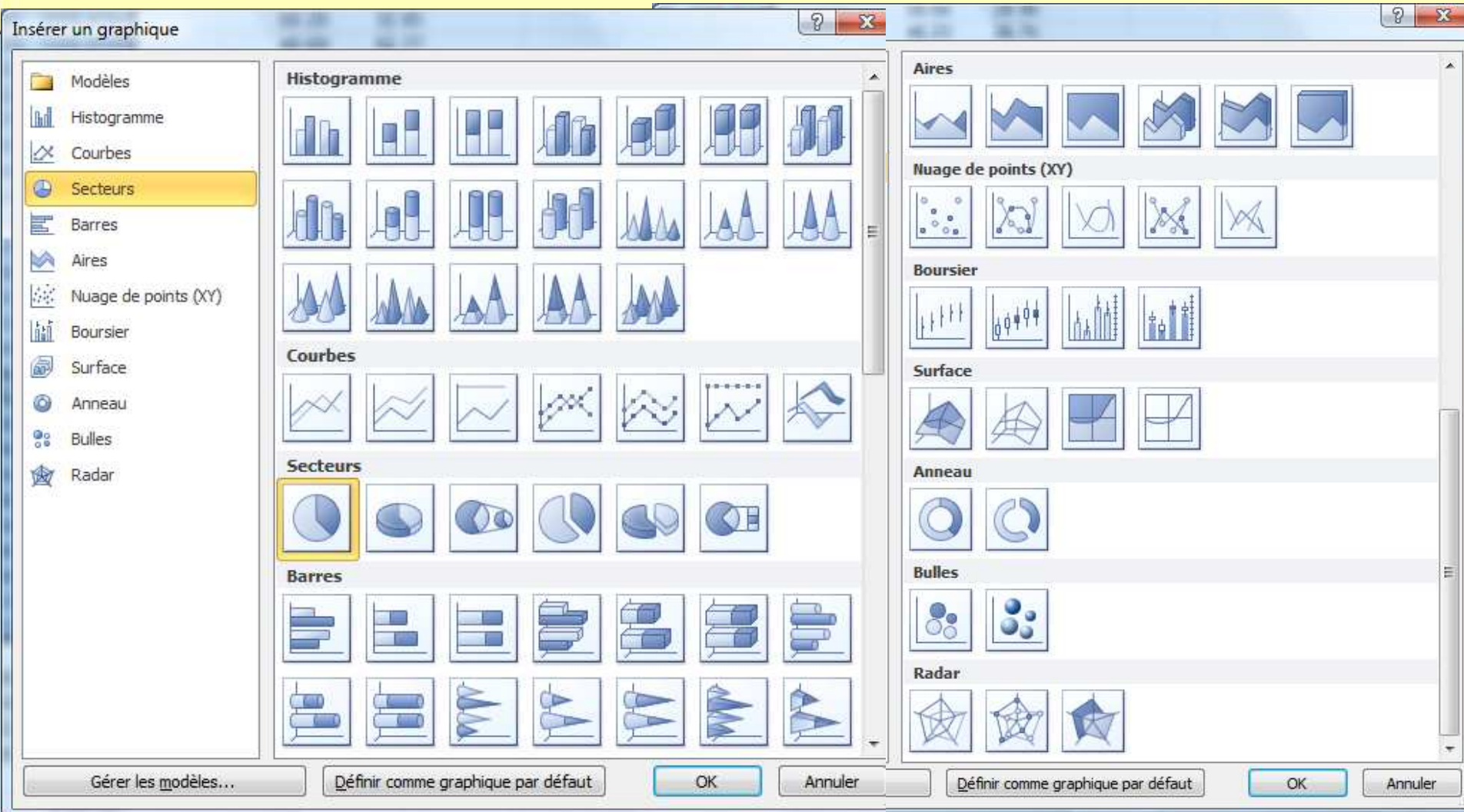
Graphiques statistiques

⇒ L'extrait suivant de la section 20 du site de Heiser résume bien le problème :

THE SITUATION

- 1. THE DEFAULT EXCEL 2007 GRAPHIC CHARTS ARE ALL BAD.**
- 2. EXCEL 2007 CHARTS ARE LOADED WITH CHARTJUNK.**
- 3. EXCEL 2007 LACKS THE CAPABILITY TO CREATE SOME STANDARD STATISTICAL CHARTS.**
- 4. CHARTS AND GRAPHICS BUILT UNDER EXCEL 2000 OR 2003 MAY NOT WORK IN EXCEL 2007.**
- 5. BUILDING GOOD STATISTICAL CHARTS IN EXCEL 2007 IS MORE COMPLEX AND DIFFICULT THAN IN PREVIOUS VERSIONS.**

Et Excel 2010 ?



⇒ Toujours pas de graphiques statistiques

Et OpenOffice.org Calc 3.2 ?



Types de diagrammes disponibles

Choisissez l'un des types de diagramme en fonction du type de données



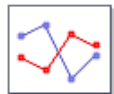
Colonne ou barre



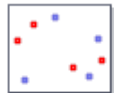
Secteur



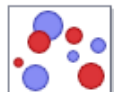
Remplissage



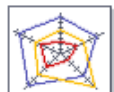
Ligne



XY (dispersé)



Bulle



Toile



Cours



Colonne et ligne

⇒ Ce n'est pas mieux du point de vue statistique

⇒ Moins chargé toutefois

Conclusions de cette partie

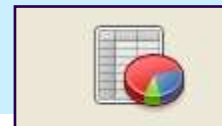
- ⇒ Les différentes versions d'Excel (sauf celles antérieures à 2003) ne sont **pas parfaites** mais sont suffisantes **sauf pour les graphiques**
- ⇒ Excel 2010 s'avère meilleur sur certains aspects
- ⇒ Un tableur est l'outil de calcul le plus logique quand le téléphone portable ne suffit pas
- ⇒ La statistique exige un **esprit critique** donc un outil imparfait peut suffire
- ⇒ **Calculettes** programmables et à fonction statistique = objets du passé
- ⇒ Evidemment un **logiciel statistique** est mieux mais l'enseignement peut rarement accepter son coût (sauf peut-être WPS = clone de SAS, moins cher)

<http://www.teamwpc.co.uk/products/wps>

Autres solutions

Autres possibilités qu'Office 2010

- ✓ OpenOffice.org 3.2 qui
 - ✓ permet maintenant d'ouvrir les fichiers .***x d'Office 2007 et 2010
 - ✓ accepte beaucoup des macros VBA d'Office
 - ✓ possède autant d'imperfections qu'Excel mais ...
- ✓ Excel avec **XLStat** (commercial) et autres solutions similaires (WinStat, Analyze-it, ...)
- ✓ R avec l'interface **RCommander** (RComdr) *
- ✓ **GNumeric** (tableur libre compatible avec Excel et OpenOffice.org Calc), mais peu répandu



* Mais sorties épouvantables ...

Graphiques de Gnumeric



CH08EX04.XLS - Gnumeric

Fichier Édition Affichage Insertion Format Outils Statistics Données Aide

Étape 1 sur 2 : choix du type de graphique

Type de tracé

- Aire
- Anneau
- Barre
- Bulle
- Camembert
- Colonne
- DropBar
- Ligne
- Ligne de niveau
- Min-Max
- Polaire
- Radar
- Statistiques
- Surface
- XY
- XY coloré

Variante

Suite telle que : Auto

Utiliser la première série comme abscisse partagée

New graph sheet

Aide Insérer Suivant Annuler

	A	B	C
833	397	-0.243	-0.335
834	398	-0.172	-0.265
835	399	-0.690	-0.783
836	400	-0.742	-0.835
837			
838	Moy. =	0.09278	9.8E-015
839	Var. =	0.98073	
840			Autocorre
841			
842	On a généré cent échantil		
843	SAMPL		Ecl
844			
845			
846			
847			
848			
849			
850			
851			

Main LAG1 LAG2 LAG3 DISTRIB langues

Somme=0

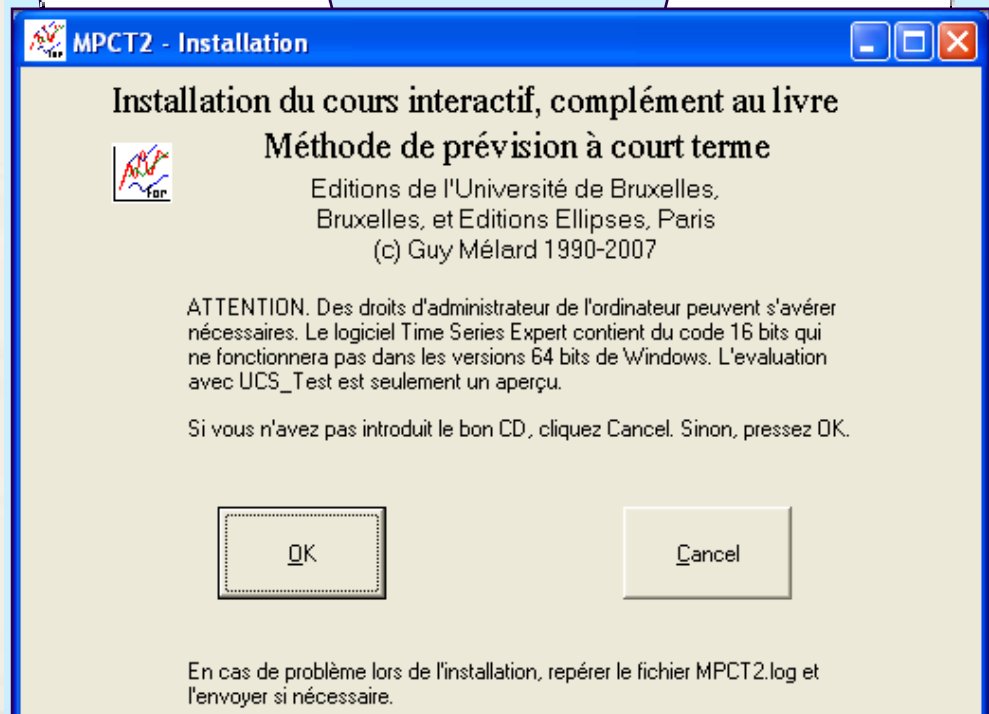
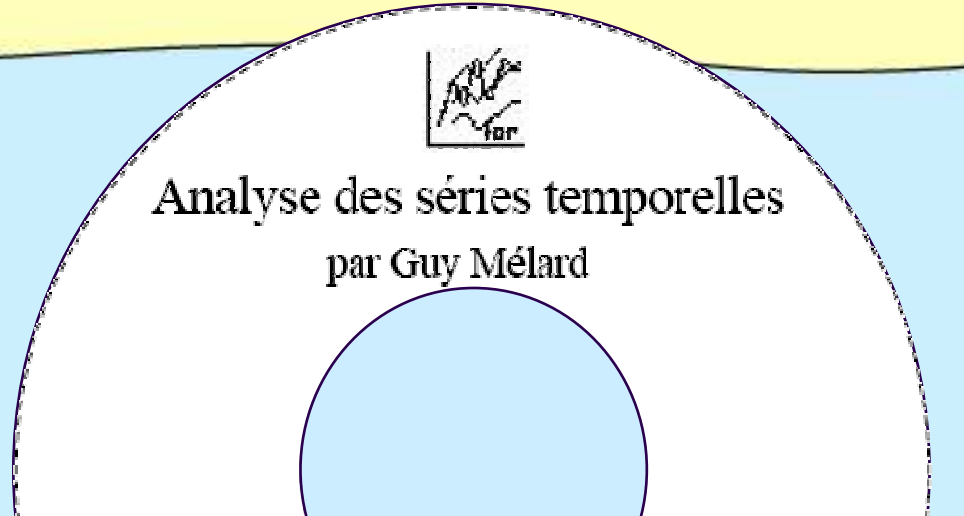
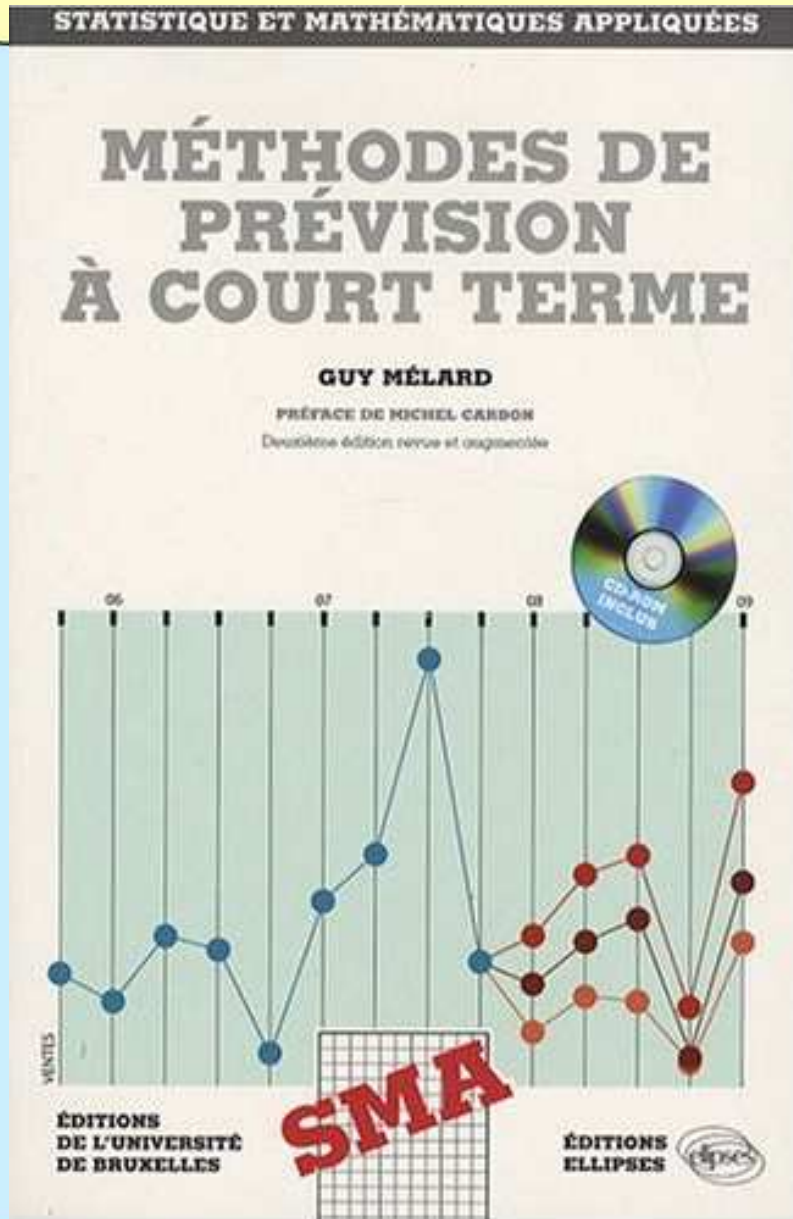
Menu statistique de Gnumeric



The screenshot shows the Gnumeric application window titled "Book1.gnumeric - Gnumeric". The menu bar includes "Fichier", "Édition", "Affichage", "Insertion", "Format", "Outils", "Statistics", "Données", and "Aide". The "Statistics" menu is open, showing a list of options: "Descriptive Statistics", "Échantillonnage...", "Dependent Observations", "One Sample Tests", "Two Sample Tests", "Multiple Sample Tests", and "Table de contingences". The "Multiple Sample Tests" option is selected, and its sub-menu is open, showing "ANOVA" (selected), "Table de contingences", "Facteur unique...", "Deux facteurs...", "k Test...", and "Wilcoxon-Mann-Whitney Test...". The spreadsheet area shows columns A, B, and C, and rows 1 through 13. The status bar at the bottom indicates "Somme=0".

***Application : examen des
classseurs du CD-Rom
joint au livre MPCT2***

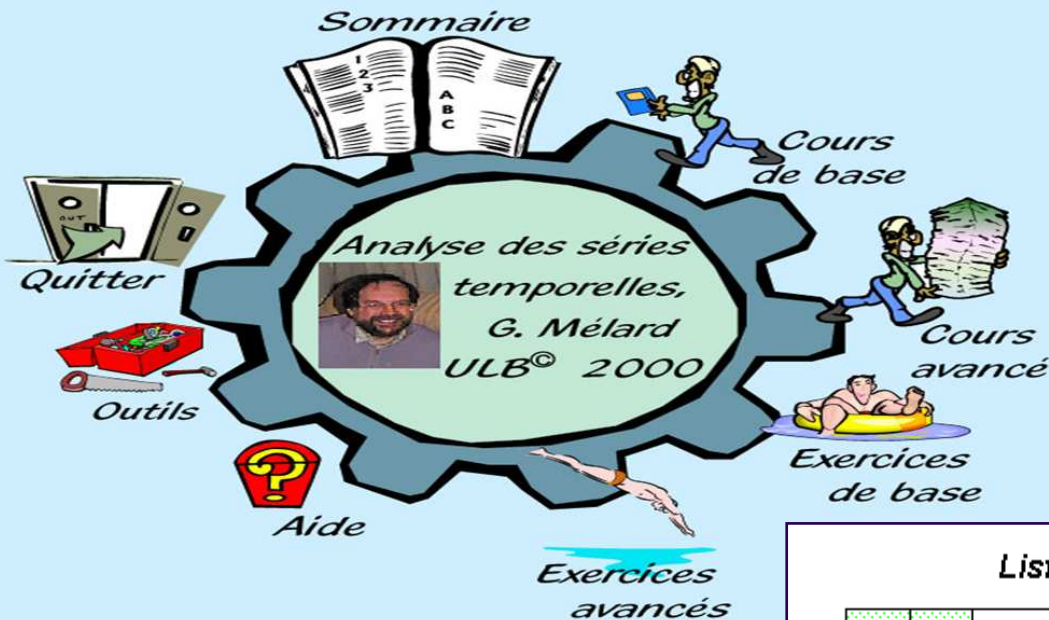
Et les classeurs de MPCT 2e édition ?



Les chapitres du livre MPCT et du CD-Rom (*)

	Nombre de classeurs
1. Concepts et définitions	3
2. Régression simple	4
3. Courbes de croissance	5
4. Moyennes mobiles	3
5. Décomposition saisonnière	10
6. Lissage exponentiel	8
7. Régression multiple	9
8. Autocorrélation et stationnarité	3
9. Modèles ARMA	0
10/11. Méthode de Box et Jenkins	0
12/13. Méthodes X12-ARIMA/TRAMO-SEATS (*)	2
Total	47

MPCT2 : accueil et exercices



Accueil

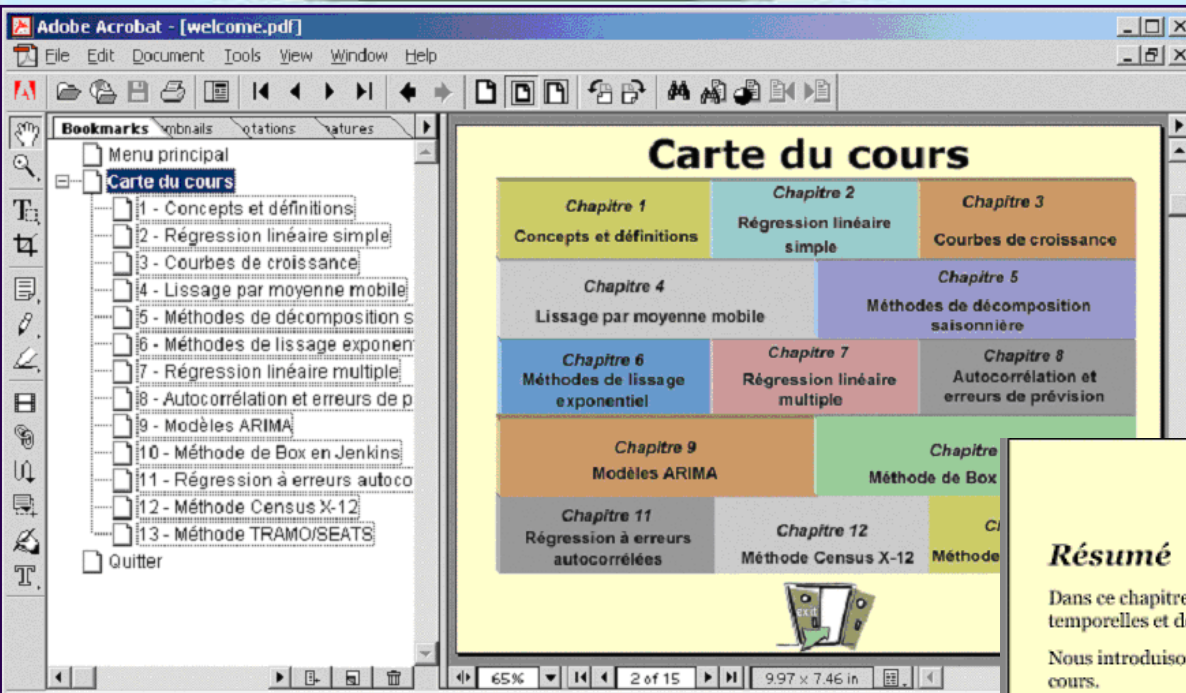
Exercices

Liste des exercices classée par chapitre

CH	EX	Données	Sujet	Accès lecteur PDF	Accès aux fichiers X: Excel TSE: Time Series Expert Demetra néant
01	01	4 données artificielles	Comparaison des critères pour deux méthodes de prévision	Cliquez ici	X: 1,2,A
01	02	Nombre de contrats d'assurance-vie	Comparaison de méthodes de prévision sur des données mensuelles (notamment lissage exponentiel)	Cliquez ici	X: 1,2,A,B



Accès aux chapitres du cours



Accès
à la présentation
et à l'évaluation



1 - Concepts et définitions

Résumé

Dans ce chapitre nous vous présentons les différents aspects de l'analyse des données temporelles et de la prévision.

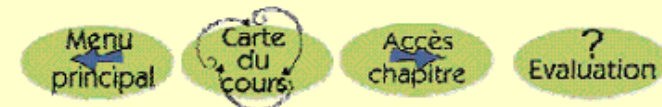
Nous introduisons un certain nombre de concepts et de définitions qui serviront durant tout le cours.

Nous montrons d'emblée les résultats d'analyse et la manière de les comparer à l'aide de différents critères.

Nous mettons l'accent sur la prévision et justifions cette décision.

Matériel

Pour étudier ce chapitre, vous disposez d'un support papier fourni dans la malette pédagogique et d'exercices pour Excel.



Des présentations aux exercices

Annotation placée ici par l'apprenant

Adobe Acrobat - [chap01.pdf]

File Edit Document Tools View Window Help

Bookmarks Thumbnails Annotations Signatures

- Début du chapitre : Concepts et défin
- Objectifs du chapitre
- Plan du chapitre
- 1.1 Concepts de base
 - Graphiques et tableaux
 - Qu'est ce qu'une donnée temp
 - Introduire les différents exemp
 - Les graphiques
 - Les tableaux
 - Question : Les prix du cuivre (
 - Question : Les salaires en Ital
 - Question : Le chômage en We
 - Question : Les taux d'intérêt a
 - Question : La production d'ass
 - Question : La popularité des p
 - Qualité des tableaux
 - Types de données temporelles

Exercice sur les critères

Données : *fictives (4 données)*

But de l'exercice : *introduire les critères (et donc les fonctions de coût) sur un exemple simple où les calculs sont faciles*

⇒ Instructions (parties 1 et 2)

⇒ Exemple (fichier CH01EX01.XLS)

Chapitre 1 Analyse des séries temporelles, G. Méland 100 Concepts et définitions

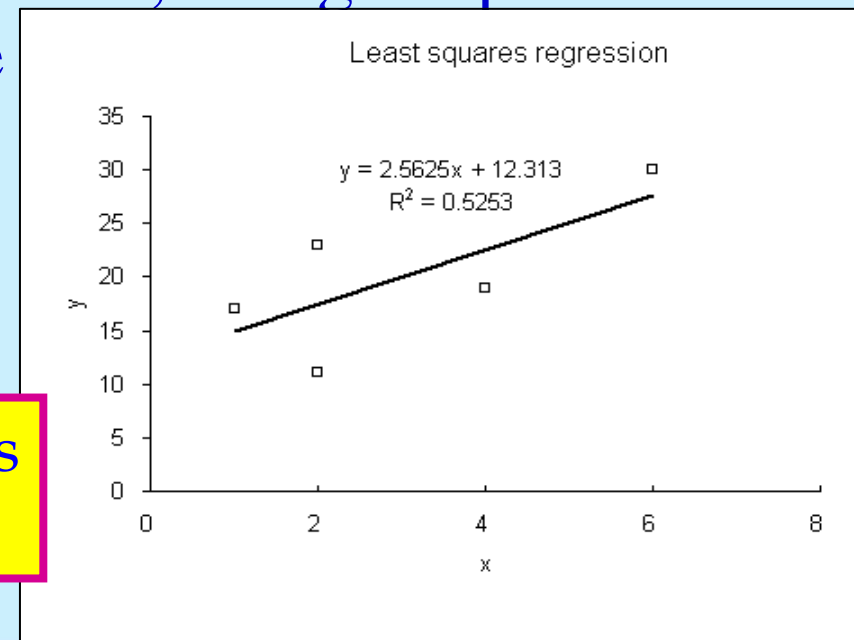
51% 100 of 142 11.68 x 8.25 in

Accès à l'énoncé

On lance directement l'exercice, ici dans Excel.

Que fait-on avec Excel ? (1/7)

- ⇒ Examens des formules par rapport à la théorie
- ⇒ Effet immédiat de changement d'une cellule (p.ex. données extrême)
- ⇒ Beaucoup de formules peuvent être obtenues par copier & coller
- ⇒ Les récurrences (moyennes mobiles, lissage exponentiel simple) sont mises en évidence
- ⇒ Ajustement statistique (linéaire, exponentielle, ...) de graphiques standards



Nous indiquerons les problèmes liés à Excel 2010 ou OpenOffice.org 3.2

Que fait-on avec Excel ? (2/7)

⇒ Fonctions statistiques et autres incorporées

- ✓ moyenne, variance, médiane
- ✓ probabilités et quantiles de lois classiques
- ✓ estimation par moindres carrés
- ✓ génération de nombres pseudo-aléatoires
- ✓ nombres complexes pour analyse spectrale

-0.1162					
1.5532	-0.1216				
-0.8938	1.5561	-0.1075			
-0.0878638	0.0587452				
15	16	Somme	1 - Somme	1 - Somme ^2	Spectre
0.94693012	0.93969262				
-0.0879	0.0587	0.98776299	1.223700012	0.000149744	6678.056
8.32008636	5.52023993	0.98728262	1.271737777	0.000357563	2796.707
6.970702614	4.50014082	0.98595862	1.404137744	0.000952632	1049.723
4.88145026	2.93725831	0.98412362	1.587637968	0.001857492	538.360
2.27408205	1.02009910	0.98227203	1.772796188	0.002967248	337.013
5.74656638	-1.0200991	0.98097913	1.902086771	0.004180936	239.181
8.36240142	-2.9372583	0.98080428	1.919571604	0.005436388	183.946
5.79326179	4.5001408	0.98219246	1.780753200	0.006741441	148.336

Pas disponible dans les anciennes versions françaises d'Excel

Que fait-on avec Excel ? (3/7)

⇒ fonctions de type tableau, comme calcul matriciel (à entrer par Shift-Ctrl-Enter !) p.ex. colinéarité

Contrairement à l'apparence ce ne sont PAS des parenthèses

`{=MMULT(XT;X)}`

	C	HAB	SIT	Y	RIX
X	1	15	0		33.0
	1	24	0		30.0
	1	19	0		37.5
	1	29	0		42.0
	1	22	1		44.5
	1	21	0		34.0
	1	20	1		40.0
	1	18	0		24.5
	1	27	1		48.0
	1	25	0		36.5

X'											X'X	X'Y		
1	1	1	1	1	1	1	1	1	1	1	10	220	3	370
15	24	19	29	22	21	20	18	27	25	220	5006	69	8288	
0	0	0	0	1	0	1	0	1	0	3	69	3	132.5	

INV(X'X)			INV(X'X) . X'Y		
3.02031802	-0.1333922	0.04770318	18.2835689		
-0.1333922	0.00618375	-0.0088339	0.725265018		
0.04770318	-0.0088339	0.48881037	9.202002356		

Que fait-on avec Excel ? (4/7)

⇒ Outil complémentaire d'analyse des données (pas installé par défaut) y compris la régression

SUMMARY OUTPUT			
<i>Regression Statistics</i>			
Multiple R	0.8309378		
R Square	0.6904576		
Adjusted R Square	0.602017		
Standard Error	4.4210168		
Observations	10		
<i>ANOVA</i>			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Regression	2	305.1822733	152.59
Residual	7	136.8177267	19.545
Total	9	442	
	<i>Coefficients</i>	<i>Standard Error</i>	<i>tStat</i>
Intercept	18.283569	7.683312583	2.3796
HAB	0.725265	0.347654593	2.0862
SIT	9.2020024	3.090952765	2.9771

R^2

R^2 ou R^2 corrigé

écart-type
résiduel

Pas disponible dans
OpenOffice.org 3.2

↑ \hat{b}_j

↑ $\hat{\sigma}(\hat{b}_j)$

↑ t

Que fait-on avec Excel ? (5/7)

⇒ Expérience de simulation (en pressant F9 aussi dans les graphiques) avec Données > Table de données + histogramme

Hundred samples were generated and the autocorrelations of lags 1, 2 and 3 were computed

SAMPLE	Sample	1	2	3
		0.013	-0.075	-0.039
	1	-0.001	0.009	-0.014
	2		0.042	-0.028
	3		-0.074	0.001
	4	0.039	-0.074	0.054
	5	0.018	0.012	-0.036

Main!F845

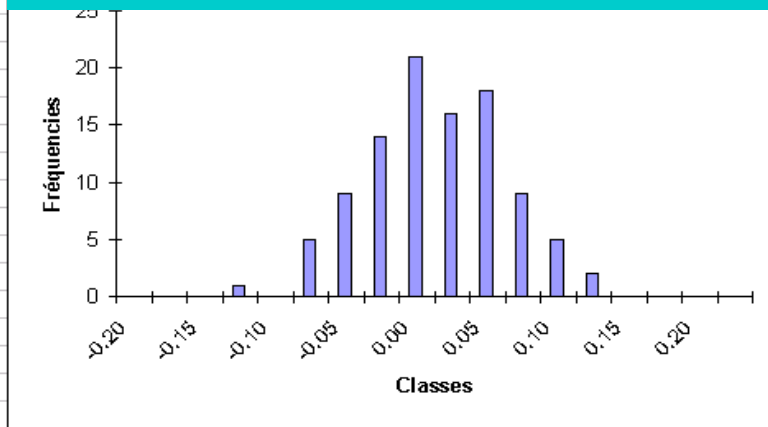
Ne fonctionne pas dans GNumeric ni OpenOffice.org 3.2

A3

Distribution of lag 1 autocorrelations over 100 samples

Classes	Frequencies	Cumulated %
-0.200	0	0
-0.175	0	0
-0.150	0	0
-0.125	1	1
-0.100	0	1
-0.075	5	6
-0.050	9	15
-0.025	14	29
0.000	21	50
0.025	16	66
0.050	18	84
0.075	9	93
0.100	5	98
0.125	2	100
0.150	0	100
0.175	0	100
0.200	0	100
	0	100
Avg	0.0025783	

=FREQUENCY(Main!F845:F944;A3:A19)



Que fait-on avec Excel ? (6/7)

⇒ Données > Table de données est aussi utile pour l'estimation d'un paramètre sur une grille (SES)

⇒ Classeurs subdivisés en

✓ feuilles de calcul

✓ feuilles graphiques

⇒ Onglets pour se déplacer entre feuilles

⇒ **Hyperliens** ←

⇒ Feuilles multilingues (F, NL, E) avec une feuille de traduction et une cellule de choix de langage

⇒ Logiciel très bien localisé

	A	B	C
1	SHORT-TERM FORECASTING METHODS,		
2	Chapter 1, tables 1.3 and 1.6, figures		
3			
4	<u>Table 1.3. Go to additive model</u>		
5	<u>Table</u>	E:\BANQUE NATIONALE\En\CHAP01\CH01EX03.XLS	
6		Change the forecast and/or the r	
7		You can change the errors but th	

Pas disponible dans
OpenOffice.org 3.2
mais bien GNumeric

Que fait-on avec Excel ? (7/7)

- ⇒ Protection des classeurs pour éviter les erreurs
- ⇒ Accès libre aux cellules pour changer des paramètres (constant de lissage) et données (extrêmes)
- ⇒ Validation des données et message
- ⇒ Utilisation des couleurs pour insister sur les différents types d'information (données, paramètres, résultats, en-têtes)
- ⇒ Scénarios employés dans le module Solver (SES + courbes de croissance)
- ⇒ Usage modéré de code Visual Basic pour Application code (restauration des données d'origine, contrôle des animations)

Pas disponible dans OpenOffice.org 2.0 mais bien 3.2 avec quelques petits problèmes à corriger

Merci pour votre attention

***Une version de cette présentation sera
bientôt disponible sur ma page web
<http://homepages.ulb.ac.be/~gmelard>***