



Histoire et panorama des relations agitées entre MS Excel et les statisticiens

CFIES'2010 - Bruxelles

Thierry Fahmy, ADDINSOFT
Emmanuel Jakobowicz, ADDINSOFT





Plan

- Ⓧ Petite histoire des tableurs et d'Excel
- Ⓧ Histoire des logiciels de statistique
- Ⓧ De l'ignorance à la convergence
- Ⓧ Enseigner avec MS Excel & XLSTAT : quelques exemples



Histoire des tableurs et d'Excel

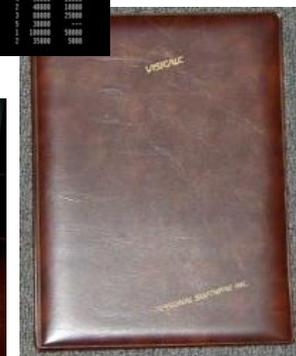
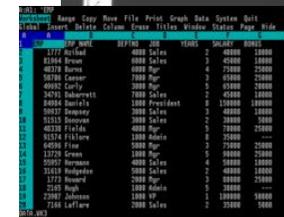
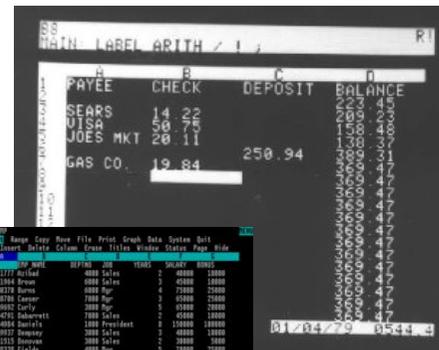


© Jean-Jacques Bernard



Histoire des tableurs (1)

- Ⓛ 1969 : LANPAR (LANguage for Programming Arrays at Random)
- Ⓛ 1978 : Daniel Bricklin, élève à la Harvard Business School a l'idée d'un calculateur interactif et visuel
- Ⓛ 1979 : La société Software Arts est créée autour du logiciel **VisiCalc** de Bricklin et Frankston qui instaure la règle A1 pour l'adressage
- Ⓛ 1982 : Microsoft sort **Multiplan**, adressage R1C1
- Ⓛ 1983 : **Lotus** Development Corporation sort la première version de 1-2-3 optimisée pour l'IBM PC
- Ⓛ 1985 : Lotus Development Corporation acquiert Software Arts et VisiCalc



Histoire des tableurs (2)

- 1985 : Microsoft remplace Multiplan par **Excel**. Excel est le premier logiciel à permettre le recalcul dynamique des valeurs liées à une cellule
- 1988 : Borland (société d'origine danoise) sort **Quattro Pro**, avec le but de devenir numéro 1 du marché à la place de Lotus 1-2-3
- 1994 : développement de StarOffice par StarDivision (Allemagne), avec le tableur **Calc**
- 1995 : IBM rachète Lotus
- 1996 : Corel acquiert Quattro Pro
- 1999 : acquisition de StarOffice par Sun Microsystems

Histoire des tableurs (3)

- 2000 : Sun Microsystems annonce la sortie de **OpenOffice**, version open source de StarOffice (version 8 depuis 2005)
- 2002 : Version 1 de OpenOffice
- 2007 : Google met à disposition Google Spreadsheets qui fera partie en 2009 de la suite Google Documents
- 2007 : Microsoft révolutionne la capacité des tableurs en permettant à une feuille de contenir plus **1 048 576 lignes** et **16384 colonnes**

Histoire d'Excel (Windows)

Version	Released	Comments
1	1985	Version 1 pour Mac
2	1987	Première version pour Windows numérotée "2" mais correspondant à la version Mac.
3	1990	Barres d'outils, possibilité de dessiner, mettre en exergue, support d'add-ins, graphiques 3D, ...
4	1992	Première version "populaire". Amélioration de la "usability".
5	1993	Mise à jour majeure. Classeurs multi-feuilles.
7	1995	Connue comme Excel 95. Première version 32-bit d'Excel. support du VBA. Sinon, proche d'Excel 5.
8	1997	Connue comme Excel 97. Nouvelle interface pour le VBA: UserForms, validation des données,
9	1999	Connue comme Excel 2000. HTML supporté comme format natif, capacité d'auto-réparation, presse papier avancé, graphiques pivot
10	2001	Connue comme Excel 2002, cette version fait partie d'Office XP. Des nouvelles fonctionnalités, mais pas visibles pour la plupart des utilisateurs. Possibilité de récupérer son fichier après un crash.
11	2003	Connue comme Microsoft Excel 2003. Les nouvelles fonctionnalités: (a) meilleur support du XML, (b) une nouvelle "list range", (c) amélioration des smart tags, et (d) correction des fonctions statistiques.
12	2007	Changement complet de l'interface. Nouveau langage de programmation. Nouvelle gestion des "barres d'outils" qui deviennent des "ribbons".
14	2010	Améliorations de l'interface et des performances

Evolution du marché (90s)



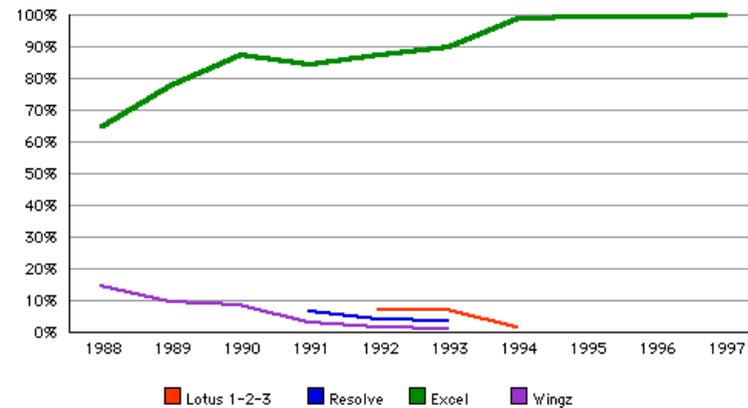
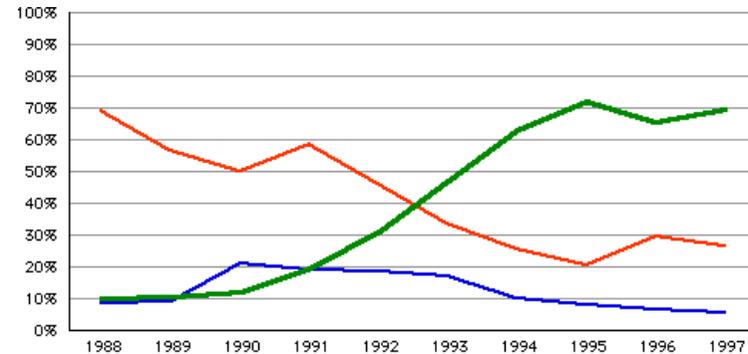
PC :

- Microsoft Excel remonte la pente face à Lotus puis le dépasse
- Quattro Pro devient négligeable



Mac :

- Microsoft Excel a toujours dominé le marché



Le marché aujourd'hui

- ① 1. Microsoft Excel (entre 90 et 95%)
 - ① 2. OpenOffice Calc
 - ① 3. Works
 - ① 4. Google Documents
 - ① 5. Lotus Symphony, Gnumeric, StarOffice, ...
-
- ① La très grande majorité des applications avancées sont encore développées sous Excel en raison
 - De la puissance du VBA
 - De la richesse du contenu disponible gratuitement
 - Des exemples et des forums accessibles
 - Des habitudes prises par des centaines de millions d'utilisateurs



Histoire des logiciels de statistique



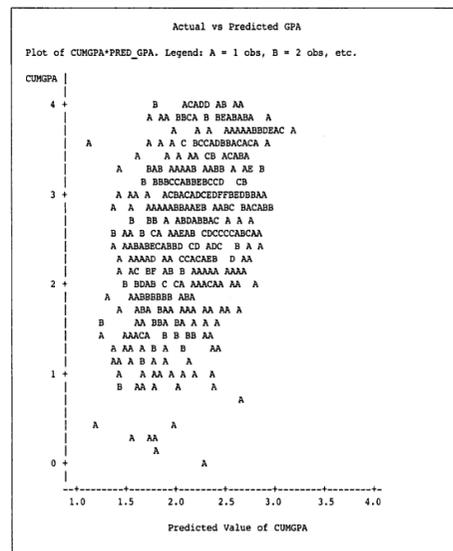
Logiciels de statistique (1)

- 1961 : BMDP est développé à UCLA comme freeware
- 1966 : SAS a été conçue par Anthony J. Barr en 1966. En tant qu'étudiant de la North Carolina State University de 1962 à 1964, Barr a créé un langage pour l'analyse de variance puis pour la régression, inspiré par la notation du statisticien Maurice Kendall, interprété pour générer du code machine. De 1966 à 1968, Barr a développé la structure fondamentale et la langue de SAS.
- En Janvier 1968, Barr and James Goodnight collaborent et améliorent les routines pour le modèle linéaire. En 1971 le monde académique l'adopte et en 1972, c'est au tour des industriels
- 1968 : SPSS (à l'origine, Statistical Package for the Social Sciences) est créé par Norman H. Nie (étudiant à Standford), C. Hadlai Hull et Dale H. Bent. C'est d'abord un langage batch pour mainframes puis IBM PC.
- 1968 : Début de GenStat
- 1972 : Minitab Statistical Software est créé par 3 enseignants de Penn State University
- 1975 : La société SPSS est créée.
- 1976 : La société SAS est créée.

Logiciels de statistique (2)

- 1. Au cours des années 1980, l'utilisation des logiciels de statistique se démocratise

```
proc glm data=forglm;
  class gender;
  model y1-y4=gender / nouni;
  repeated age 4 (8 10 12 14) / printe;
run;
```



Logiciels de statistique (3)

- ① Années 80' : Une multitude de logiciels voit le jour :
 - Statgraphics (1980), SYSTAT (1982), Stata (1985), Statistica (1985, module pour Lotus 1-2-3, StatFast pour Apple), Statview (1985), Umetrics (1987), S+ (1988)
 - En France, StatITCF, Spad, Statlab

- ① Années 90' : Passage de DOS à Windows :
 - De nombreux logiciels ne prennent pas le tournant
 - Disparition de StatITCF
 - Acquisitions (notamment par SAS et SPSS)
 - Apparition des premiers compléments pour Excel (Statbox, XLSTAT, Analyse-it)

- ① Années 2000 : une nouvelle ère
 - R prend de l'ampleur
 - Certains éditeurs se tournent vers le Data mining et le Data Warehousing, plus lucratif

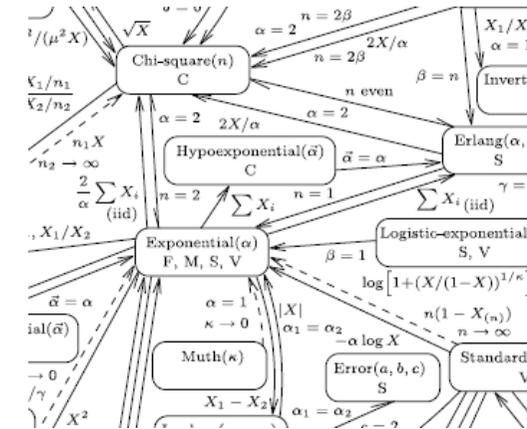


Les tableurs, non merci



Deux mondes

① D'un côté, des scientifiques travaillant sur des systèmes avec lesquels ils ont grandi



1986, Lawrence Leemis

① De l'autre des comptables, se limitant à des opérations de base (+, -, *, /)



Un passif scientifique

- ① En 1993, Excel se dote de macros complémentaires et de fonctions statistiques pour satisfaire une demande croissante
 - Ces macros sont pauvres ... et buggées
 - Les algorithmes, même les plus simples (Ecart-type) sont instables
 - Discrédit auprès de la communauté scientifique

- ① Des éditeurs profitent de cette brèche pour lancer des outils compatibles avec Excel



Finalemment ...



Convergence

- De plus en plus de logiciels de statistique adoptent une ergonomie de tableur et donnent accès au format Excel
- Les bugs des fonctions statistique d'Excel sont corrigées
- Excel est le principal format de stockage de données
- De plus en plus d'enseignants de statistique appliquée adoptent Excel pour leurs cours
- Les prix des logiciels de statistique ont augmenté, celui d'Excel baisse
- XLSTAT propose un très grand nombre de fonctions dépassant parfois celles des logiciels traditionnels



XLSTAT pour Microsoft Excel



Histoire d'XLSTAT

- 1993: Début du développement / Une alternative “user friendly” à SAS
- 1996: Premières ventes sur Internet
- 2000: Création de la société Addinsoft
- 2001: Nouvelle interface
- 2002: Premiers modules avec des partenaires: Pivot & 3DPlot
- 2004: Réécriture complète du logiciel
- 2006: Version 2006, avec le VBA comme interface et les calculs en C++
disponibilité en 7 langues. Interface pour programmation, nouvelle
version Mac
- 2007: Première conférence utilisateurs
- 2008: Lancement des modules SPC et Sim
- 2009: Nouvelles langues, nouvelle offre verticalisée

XLSTAT aujourd'hui

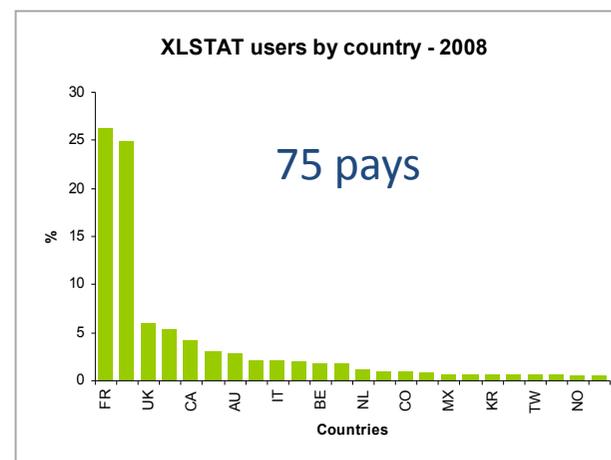
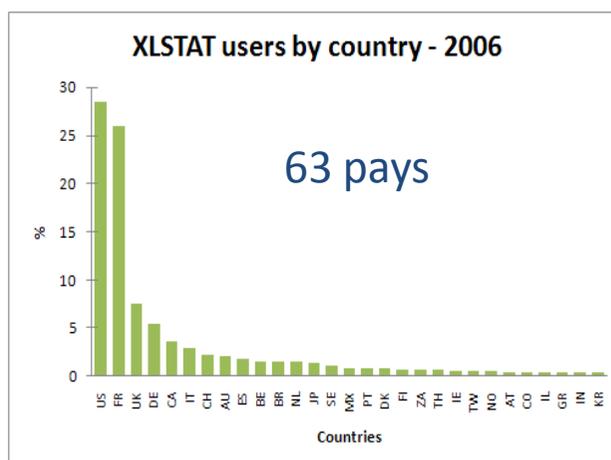
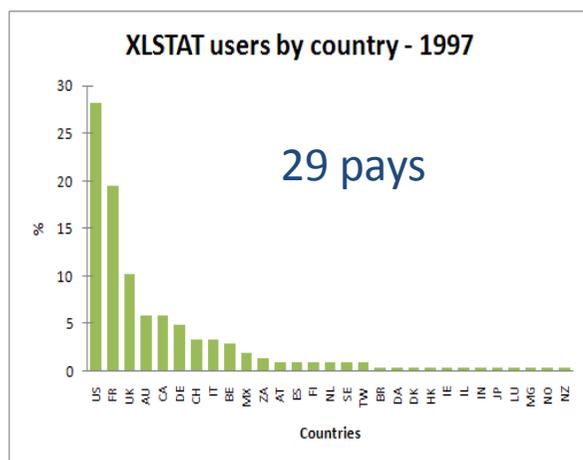
- Ⓧ Commercialisé par la société française Addinsoft
- Ⓧ Dirigé par 3 associés, français, allemand et franco-suisse
- Ⓧ 20000 utilisateurs enregistrés dans 102 pays
- Ⓧ Plus de 500 téléchargements par jour
- Ⓧ 75000 visiteurs par mois sur le site
- Ⓧ Leader mondial des solutions statistiques pour Excel
- Ⓧ 6 nouvelles versions par an
- Ⓧ Toujours plus de fonctionnalités, à l'écoute des utilisateurs

Les faits

- ❶ **MS Excel** est le tableur le plus utilisé dans le monde (90% de part de marché).
- ❷ **Excel** est le principal format de stockage des données.
- ❸ **XLSTAT** est le logiciel de statistique qui offre la meilleure intégration avec Excel.
- ❹ **XLSTAT** est le logiciel de statistique pour Excel le plus complet.
- ❺ **XLSTAT** connaît depuis 10 ans une croissance allant de 20 à 60 % par an, prenant des parts de marché à des logiciels comme SPSS ou Minitab.
- ❻ **XLSTAT** est un logiciel modulaire qui permet de réduire le coût et d'optimiser la couverture fonctionnelle.

Utilisateurs – Répartition géographique

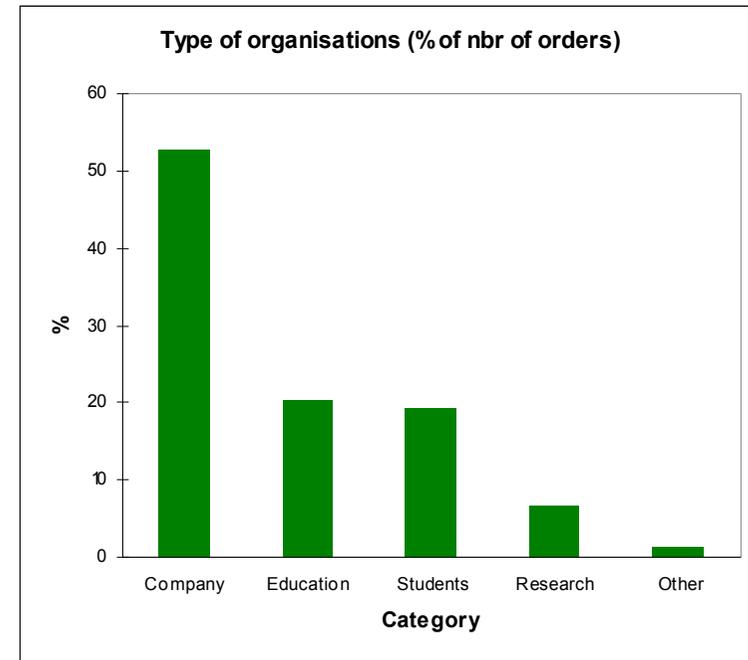
- Des clients répartis dans 102 pays (2002-2009)



- Des distributeurs dans plus de 20 pays

Utilisateurs – Répartition sectorielle

- ➊ Plus de la moitié des utilisateurs travaillent en entreprise
- ➋ Utilisateurs dans la moitié des 100 plus grandes sociétés du monde
- ➌ La plupart des utilisateurs travaillent dans des domaines liés à la recherche
- ➍ Près de 20% des utilisateurs sont des étudiants, y compris des étudiants à qui d'autres logiciels sont imposés en cours



Utilisateurs | Entreprises

France :

- Altadis, AXA, Babolat, Banque Populaire, Bel, Bonduelle, BNP, BVA, Crédit Agricole , Danone, EDF, France Telecom, GDF, Groupama, Ipsos, L'Oréal, Galeries Lafayette, Generali, Lafarge, Leroy Merlin , Michelin, Moët & Chandon, Mondiale Assistance, Pernod Ricard, Renault, RTE, Société Générale, TDF, TF1, Total, Veolia,

Reste du monde :

- Abbott, Bayer, Boeing, Cargill, Chevron, Coca Cola, Daimler, Deutsche Bank, Diageo, Exxonmobil, Firmenich, GfK, Givaudan, GlaxoSmithKline, HSBC, IFF, Illy, InBev, KFC, Kimberly Clark, Kraft, Lloyds TSB, Mars, McDonalds, Nestlé, Nexans, Novartis, Otto Group, Pepsi Cola, Philip Morris, Procter & Gamble, Sun Microsystems, Tesco, TNS Sofres, Unilever, Versar, Wells Fargo, Yahoo, Young & Rubicam, ...

Utilisateurs | Education

France :

- Ecoles d'ingénieurs : AgroParisTech, ENSAIA, ENSAR, ENITAC, ESIEA, ISA, Purpan, ...
- Ecoles de commerce : ESSEC, IESEG, INSEEC, IAE (Aix, Grenoble, Lyon), ESC Toulouse, ...
- IUT : Caen, Clermont-F., Carcassonne, Lyon, Quimper, Saint Nazaire, ...
- Universités : Aix Marseille III, Bordeaux 2/3/4, Brest, Caen, Haute Alsace, Havre, Joseph Fourier, La Rochelle, Lille 2, Lyon 1/2, Marne La Vallée, Marseille, Metz, Montpellier, Nice, Orléans, Paris 1/5/6/7/8/10/12, Reims, Rennes, Rouen, Toulouse Paul Sabatier, Louis Pasteur, Strasbourg, Tours, ...

Reste du monde :

- Cambridge, Columbia, Cornell, Harvard, McGill, Oxford, Princeton, Rice university, Stanford, Sydney, Washington, Wisconsin, Wharton, Yale, UC Davis, UCLA, UC San Diego, ...



Utilisateurs | Recherche

France :

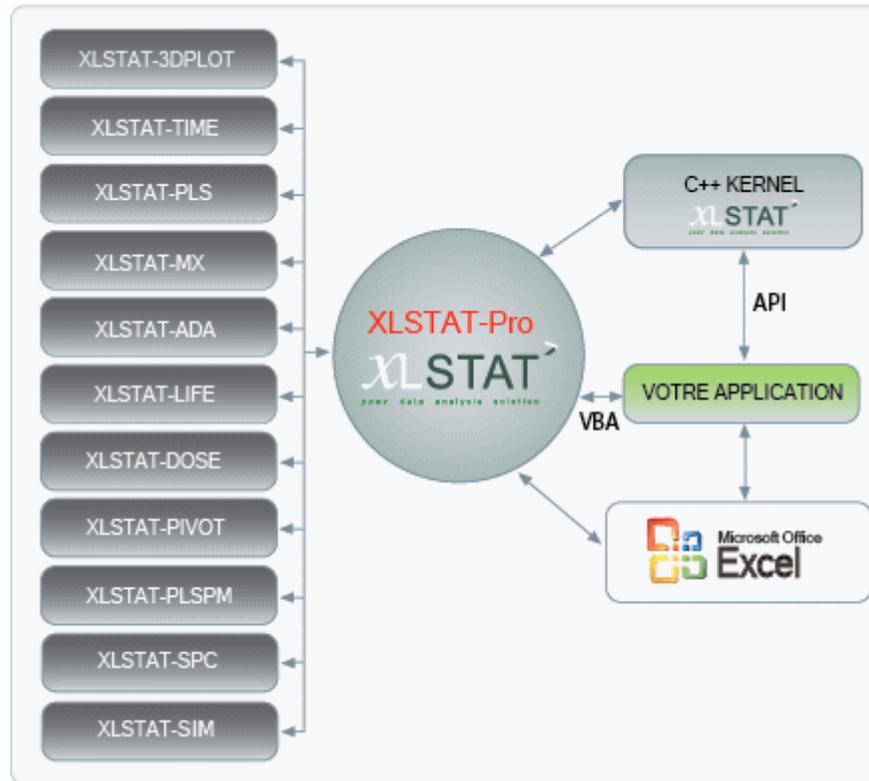
- CEA, CIRAD, CNRS (...), INERIS, INRA (Angers, Avignon, Bordeaux, Dijon, Grignon, Montpellier, Rennes, Theix, Thonon, Toulouse, Versailles), INRAP, INSERM (Bron, Clermont –F, Créteil, Nantes, Toulouse), IRD, MNHN.

Reste du monde :

- COD, CSIC (ES), CSIRO (AU), Max Planck Institut (DE), NASA, NIH, USDA, ...

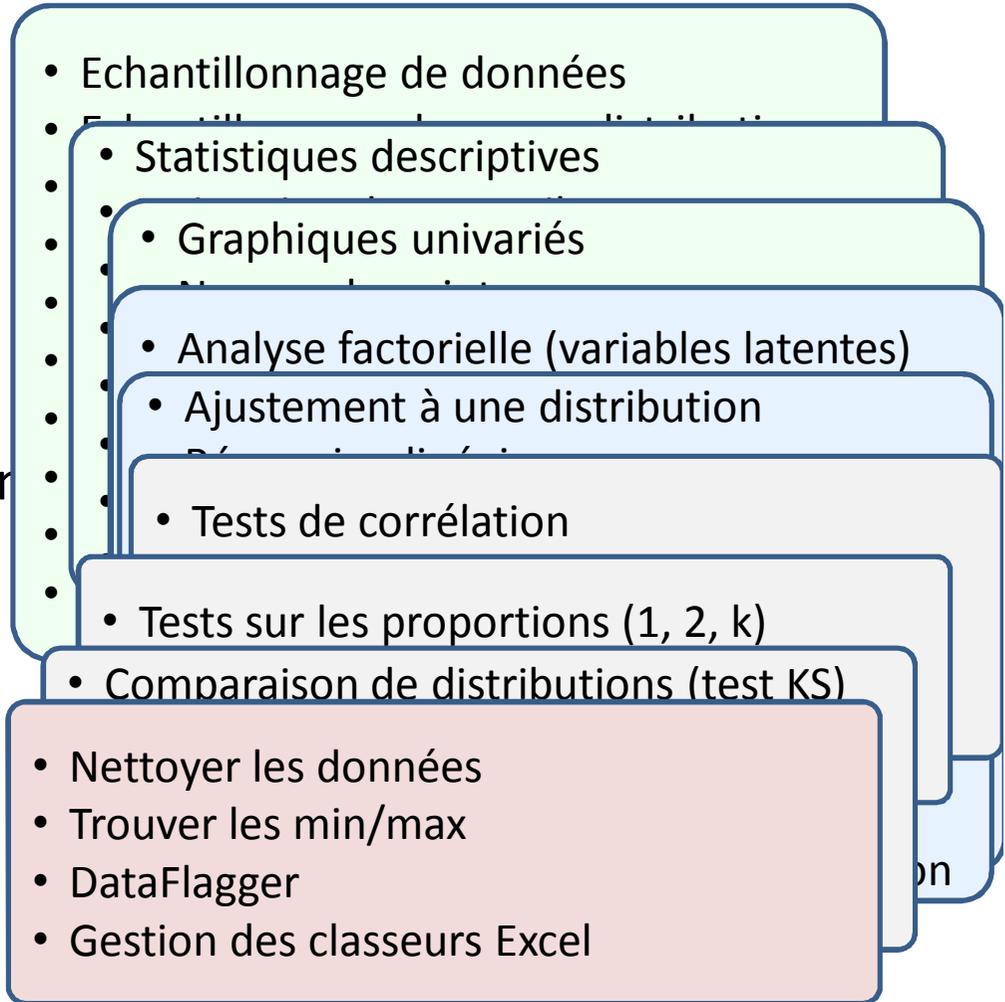


XLSTAT – Une application modulaire



XLSTAT-Pro

- Préparation des données
- Description des données
- Visualisation des données
- Analyse des données
- Modélisation des données
- Tests de corrélation/association
- Tests paramétriques
- Tests non-paramétriques
- Outils

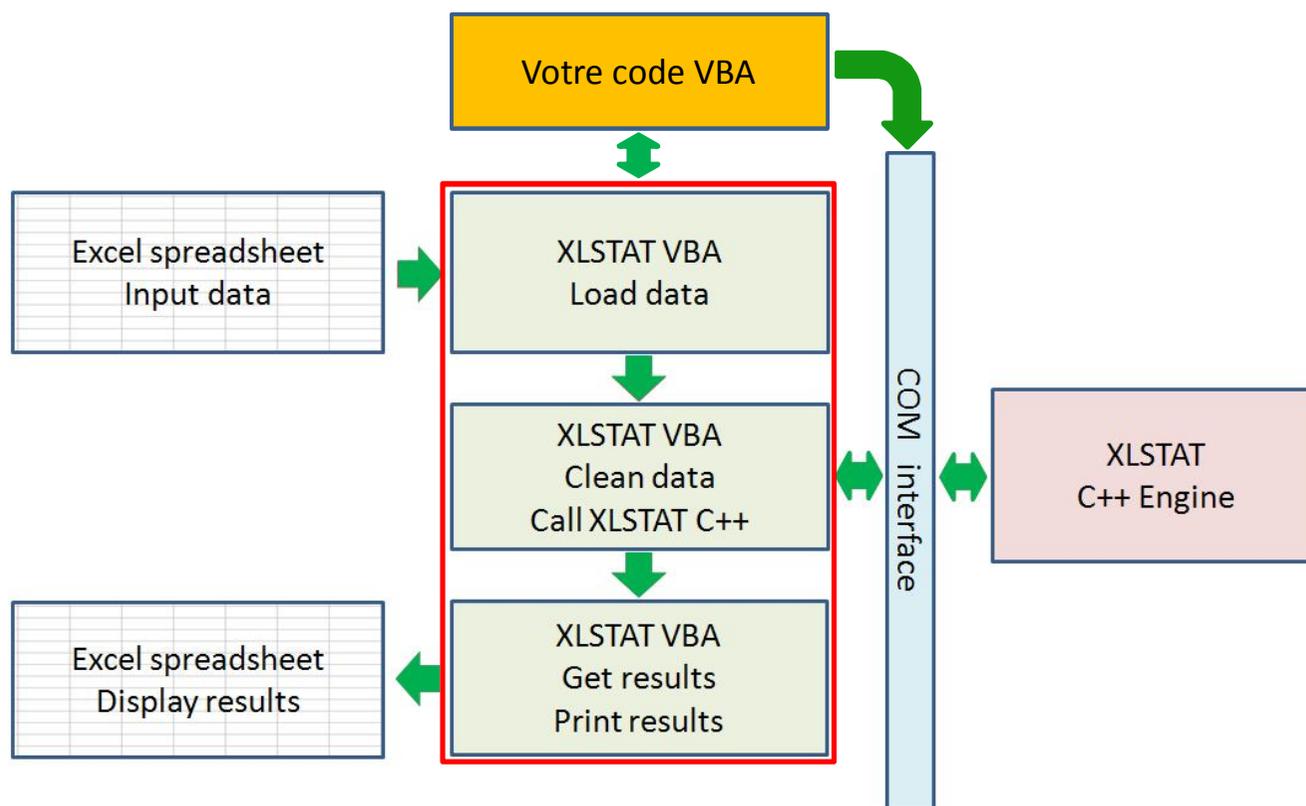


Modules complémentaires

- XLSTAT-3DPlot: Visualisation 3D
- XLSTAT-Time: Séries Chronologiques
- XLSTAT-PLS: Régression par les moindres carrés partiels
- XLSTAT-PLSPM: Approche PLS
- XLSTAT-MX: Analyse sensorielle
- XLSTAT-Dose: Analyse d'effets de dose
- XLSTAT-SPC: Statistical Process Control
- XLSTAT-Life: Analyse de survie, modèle de COX, Courbes ROC
- XLSTAT-ADA: Analyse de données avancée (PCoA, CCA, RDA ...)
- XLSTAT-Pivot: Tableaux croisés intelligents
- XLSTAT-Sim: Simulations Monte Carlo

Programmer avec XLSTAT

- Vous pouvez utiliser la couche VBA XLSTAT, ou utiliser XLSTAT comme un pont vers la couche C++



L'approche XLSTAT

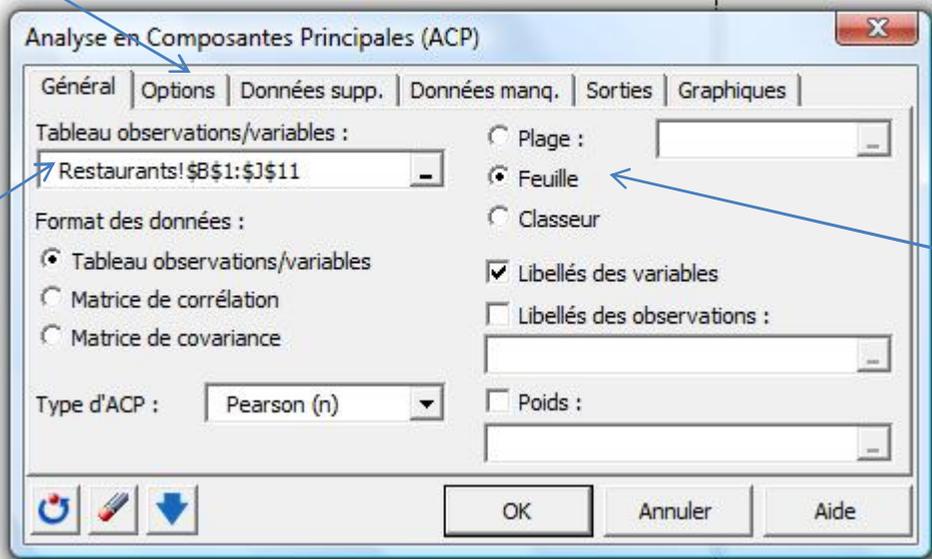
- Ⓧ Pas besoin de déclarer des variables.
- Ⓧ Les données sont dans Excel.
- Ⓧ Sélection des données avec la souris.
- Ⓧ Possibilité de mémorisation des sélections d'une session à l'autre.

L'approche XLSTAT

Les boîtes de dialogue

Des onglets donnant accès aux différentes options

Lancement à partir d'une barre d'outils ou d'un menu



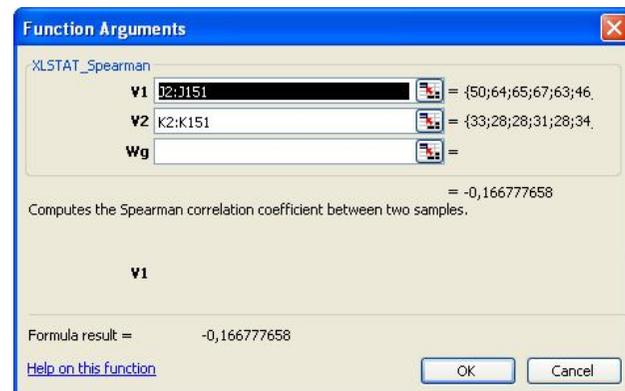
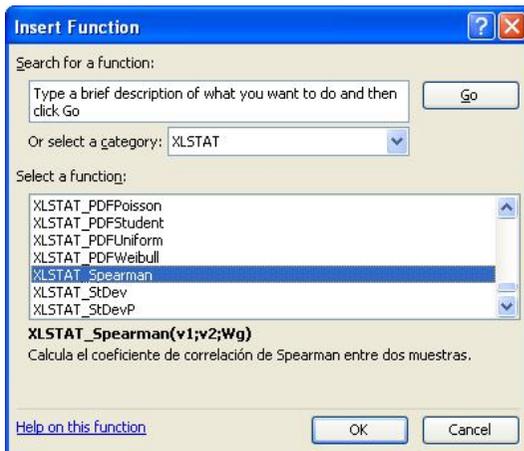
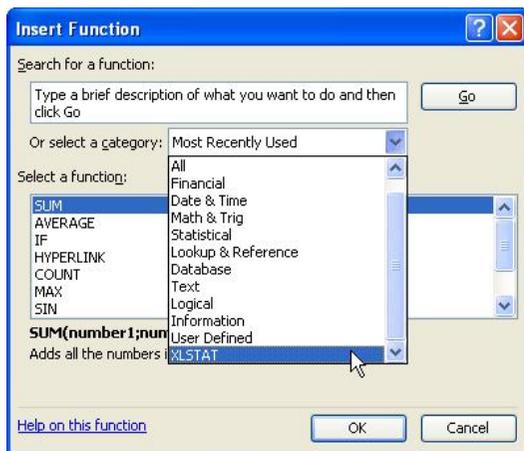
Sélection des données directement sur la feuille

Trois choix pour les sorties

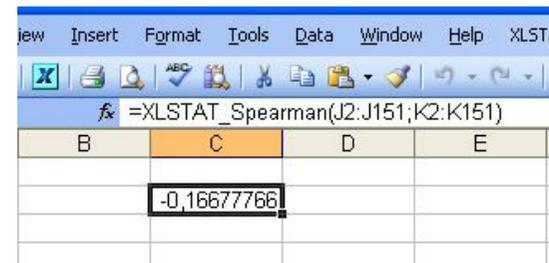


Formules XLSTAT

- ① Certaines fonctions de XLSTAT sont aussi accessibles via des formules
 - Choisissez la cellule où vous voulez ajouter la formule
 - Allez dans Excel / Insérer / Fonction
 - Choisissez la fonction
 - Entrez les arguments



- Cliquez sur OK pour afficher le résultat





Enseigner les statistiques avec Microsoft Excel et XLSTAT





Exemples

- Ⓧ Vérification de la loi des grands nombres
- Ⓧ Vérification de la loi de la moyenne
- Ⓧ Loi du U de Wilcoxon dans le cas d'ex-aequo

