
Guide d'utilisation

StatTools

*Compagnon statistique pour
Microsoft® Excel*

**Version 5.5
janvier, 2010**

**Palisade Corporation
798 Cascadilla St
Ithaca, NY 14850
USA
+1-607-277-8000
<http://www.palisade.com>**

Avis de copyright

Copyright © 2010, Palisade Corporation.

Marques déposées

Microsoft, Excel et Windows sont des marques déposées de Microsoft Corporation.

IBM est une marque déposée d'International Business Machines, Inc.

Palisade, TopRank, BestFit et RISKview sont des marques déposées de Palisade Corporation.

Bienvenue à StatTools pour Excel

Bienvenue

StatTools apporte à Microsoft Excel, l'étalon de l'analyse et de la modélisation, un nouvel et puissant ensemble d'outils statistiques ! StatTools est un complément Microsoft Excel de statistiques. Il permet d'analyser les données de feuilles de calcul Excel dans l'environnement familier de Microsoft Excel. Combinaison d'un puissant gestionnaire de données et d'analyses aptes à rivaliser les meilleurs logiciels de statistiques, StatTools offre le meilleur de deux mondes : la convivialité et les rapports de Microsoft Office, unis à une solide analyse statistique.

Environnement familier

Si vous savez utiliser Excel, vous saurez utiliser StatTools! StatTools opère tout comme Excel, avec ses barres d'outils, ses menus et ses fonctions de feuille de calcul personnalisées. Le tout dans le contexte même d'Excel. Contrairement aux logiciels de statistiques autonomes, StatTools n'exige aucun apprentissage intensif ni frais de formation initiaux, tout simplement parce que le travail se fait dans Excel. Les données et les variables se trouvent dans vos feuilles de calcul Excel. Les formules Excel standard sont admises pour les calculs et transformations, de même que les tris et tableaux croisés dynamiques d'Excel. Les rapports et graphiques des analyses statistiques effectuées sont au format Excel standard et répondent à toutes les fonctions de formatage Excel.

De solides statistiques sous Excel

StatTools remplace les fonctions statistiques intégrées d'Excel par ses propres calculs, robustes et rapides. La précision des calculs statistiques intégrés d'Excel laisse souvent à désirer. StatTools les évite totalement. Même les fonctions statistiques de feuille de calcul Excel telles qu'ECARTYPE() sont remplacées par de nouvelles versions StatTools robustes, telles que StatSTDEV(). Les calculs statistiques de StatTools sont conformes aux tests de précision les plus rigoureux, sous performance optimisée par .DLL C++, plutôt que par macro-calculs.

Analyses StatTools

StatTools couvre la gamme des procédures statistiques les plus courantes. Il offre une capacité sans précédent d'ajout de nouvelles analyses personnalisées. Non moins de 36 procédures statistiques de grande envergure, plus 8 utilitaires de données intégrés, gèrent les analyses statistiques les plus connues et courantes. Les fonctions statistiques proposées incluent des statistiques descriptives, des tests de normalité, des comparaisons de groupes, la corrélation, l'analyse de régression, le contrôle de qualité, les prévisions et bien davantage encore. Le tout assorti d'une bibliothèque de procédures personnalisées (composées par les utilisateurs ou par d'autres experts), pour un ensemble remarquable d'outils statistiques personnalisables, au sein même d'Excel !

Les calculs statistiques de StatTools s'effectuent en liaison interactive. Dans Excel, chaque changement de valeur s'accompagne d'un recalcul de la feuille et de la production d'une nouvelle réponse. Il en va de même pour StatTools ! Le changement d'une valeur dans l'ensemble de données donne lieu à l'actualisation automatique du rapport statistique. StatTools tire parti d'un puissant ensemble de fonctions de feuille de calcul personnalisées pour assurer l'actualisation permanente constante des statistiques affichées dans les rapports par rapport aux données.

Gestion des données StatTools

StatTools introduit un gestionnaire complet d'ensembles de données et de variables au sein même d'Excel, comparable à celui d'un progiciel statistique autonome. On peut définir un nombre indéfini d'ensembles de données avec, chacun, les variables à analyser, directement depuis les données Excel. StatTools évalue intelligemment les blocs de données ; il suggère les noms de variables, ainsi que les emplacements de données. Les ensembles de données et variables peuvent résider dans des feuilles de calcul et classeurs différents. L'utilisateur peut ainsi organiser ses données comme bon lui semble. Il suffit ensuite d'exécuter des analyses statistiques faisant référence aux variables, plutôt que de resélectionner chaque fois les données dans Excel. Mieux encore, les variables StatTools ne sont pas limitées à une seule colonne de données dans une feuille de calcul Excel : la même colonne peut être utilisée, pour une même variable, sur 255 feuilles de calcul (65 535 X 255, ou plus de 16 millions de cas dans la version StatTools Industrial ; 10 000 cas dans l'édition StatTools Professional) !

Rapports StatTools

Excel convient parfaitement à la production de rapports et de graphiques. StatTools en tire excellentement parti. StatTools utilise les formats graphiques d'Excel, dont les couleurs, les polices et le texte peuvent être aisément personnalisés. Les titres des rapports, les formats numériques et le texte peuvent être modifiés comme sur une feuille de calcul Excel ordinaire. Les tableaux et graphiques des rapports StatTools peuvent être transférés vers les documents d'autres applications par simple glissement-déplacement. Ils restent cependant liés aux données Excel : si les rapports d'analyse changent, le document s'actualise donc automatiquement.

Accès aux données et partage

Excel dispose d'excellentes fonctionnalités d'importation. L'importation de données existantes dans StatTools en est d'autant plus simple ! Les capacités Excel standard permettent la lecture de données originaires de Microsoft SQL Server, Oracle, Microsoft Access ou toute autre base de données à compatibilité ODBC. Données de fichiers texte ou d'autres applications ? Si Excel peut les lire, StatTools les lira aussi !

StatTools enregistre tous ses résultats et données dans des classeurs Excel. Comme tout autre fichier Excel, les résultats et données StatTools peuvent être envoyés à d'autres utilisateurs. Le partage ne pourrait être plus simple !

StatTools Industrial

StatTools Industrial comprend une interface de programmation par objets complète, permettant l'ajout de procédures statistiques à l'aide du langage de programmation VBA intégré d'Excel. Ces procédures personnalisées peuvent faire appel à la gestion de données intégrée de StatTools, ses outils graphiques et de rapport, tous accessibles à travers les contrôles, fonctions et méthodes StatTools personnalisés. Il est même possible de les faire figurer dans le menu StatTools pour y faciliter l'accès !

Et si vous ne comptez pas composer vos propres procédures statistiques ? StatTools Industrial permet l'exploitation des procédures d'autres utilisateurs, directement depuis le menu StatTools standard ! Les statisticiens experts ajoutent en permanence à la bibliothèque de nouvelles procédures personnalisées créées avec StatTools. Il suffit de copier un classeur assorti d'une nouvelle procédure dans le répertoire StatTools pour qu'il figure instantanément dans le menu StatTools. À l'exécution, tous les outils de gestion de données StatTools standard s'affichent, combinés à la nouvelle analyse statistique qui vous intéresse !

Table des matières

Chapitre 1 : Mise en route	1
Introduction	3
Contenu du coffret	3
Éléments du progiciel.....	3
À propos de cette version.....	3
Contexte d'exploitation.....	4
Si vous avez besoin d'aide.....	4
Configuration requise :	6
Installation	7
Généralités	7
Configuration des icônes ou raccourcis StatTools	8
DecisionTools Suite.....	8
Activation du logiciel.....	9
Chapitre 2 : Présentation de StatTools	13
Introduction	15
Menu et barre d'outils StatTools.....	15
Ensembles de données et Gestionnaire des données	16
Rapports et graphiques StatTools.....	19
Chapitre 3 : Guide de référence StatTools	23
Introduction	27
Langage macro VBA StatTools et Kit du développeur	27
Référence : Icônes StatTools	29
Barre d'outils StatTools	29
Référence : Commandes du menu StatTools	33
Introduction	33
Packs d'analyse complémentaires.....	33
Liste des commandes	35

Menu StatTools – Ensembles de données	43
Commande Gestionnaire des ensembles de données	43
Menu Utilitaires de données	49
Commande Empiler	49
Commande Désempiler	52
Commande Transformer	53
Commande Retard	55
Commande Différence	57
Commande Interaction.....	59
Commande Combinaison	61
Commande Var. nominale.....	63
Commande Échantillons aléatoires	65
Menu Statistiques de synthèse.....	67
Commande Synthèse de variable.....	67
Commande Corrélation et covariance	70
Menu Graphiques de synthèse	73
Commande Histogramme.....	73
Commande Diagramme de dispersion.....	76
Commande Boîte à moustaches.....	78
Menu Inférence statistique.....	81
Commande Intervalle de confiance - Moyenne/Écart type.....	81
Commande Intervalle de confiance - Proportion.....	84
Commande Test d'hypothèse - Moyenne / Écart type	87
Commande Test d'hypothèse - Proportion	90
Commande Sélection de taille d'échantillon	93
Commande ANOVA simple	95
Commande ANOVA double.....	98
Commande Test d'indépendance chi carré	101
Menu Tests de normalité	103
Commande Test de normalité chi carré.....	103
Commande Test de Lilliefors.....	107
Commande Graphique Q-Q normal.....	109
Menu Série temporelle et prévision.....	111
Commande Chronogramme	111
Commande Autocorrélation.....	113
Commande Test des runs.....	115
Commande Prévision	117

Menu Régression et classification.....	123
Commande Régression	124
Commande Régression logistique.....	129
Commande Analyse discriminante.....	133
Menu Contrôle de qualité.....	139
Commande Diagramme de Pareto	140
Commande Graphiques X/R	143
Commande Graphique P	147
Commande Graphique C.....	151
Commande de graphique U	154
Menu Tests non paramétriques	159
Commande Test des signes	161
Commande Test de rang de Wilcoxon.....	164
Commande Test de Mann-Whitney.....	167
Menu Utilitaires	171
Commande Paramètres d'application	171
Commande Supprimer les ensembles de données.....	177
Commande Effacer la mémoire de dialogue	177
Commande Décharger StatTools.....	177
Menu Aide.....	179
Aide StatTools	179
Manuel en ligne.....	179
Commande Activation de licence.....	179
Commande À propos.....	179
Référence : Fonctions StatTools	181
Introduction	181
Fonctions StatTools vs Fonctions Excel	181
Fonctions de distribution.....	183
Rapports « en direct »	184
Référence : Liste des fonctions statistiques	186
Fonctions disponibles	186
Description détaillée des fonctions	189
Index	207

Chapitre 1 : Mise en route

Introduction	3
Contenu du coffret.....	3
Éléments du progiciel.....	3
À propos de cette version.....	3
Contexte d'exploitation.....	4
Si vous avez besoin d'aide.....	4
Configuration requise :.....	6
Installation	7
Généralités	7
Configuration des icônes ou raccourcis StatTools	8
DecisionTools Suite.....	8
Activation du logiciel	9

Introduction

Cette introduction décrit le contenu du coffret StatTools et indique comment installer et relier StatTools à votre copie de Microsoft Excel 2000 pour Windows ou version ultérieure.

Contenu du coffret

Le coffret StatTools doit contenir les éléments suivants :

Learning Statistics with StatTools, un ouvrage de statistiques faisant appel à StatTools, rédigé par S. Christian Albright, professeur à l'Indiana University (en anglais).

le CD-ROM StatTools ou DecisionTools Suite, comportant

- *le logiciel StatTools*
- *le didacticiel StatTools*
- *le Guide de l'utilisateur StatTools (ce manuel) au format .PDF*

la licence d'exploitation de StatTools

Si votre coffret est incomplet, prenez contact avec votre revendeur StatTools ou appelez Palisade Corporation directement au +1-607-277-8000.

Éléments du progiciel

StatTools peut être acheté en autonome ou dans le cadre des versions DecisionTools Suite Professional et Industrial. Le CD-ROM StatTools contient le complément Excel StatTools, plusieurs exemples d'application de StatTools et un système d'aide StatTools en ligne indexé. Les versions DecisionTools Suite Professional et Industrial contiennent, en plus des éléments ci-dessus, une série d'autres applications.

À propos de cette version

Cette version de StatTools peut être installée en tant que programme 32 bits pour Microsoft Excel 2000 ou version ultérieure.

Contexte d'exploitation

Les descriptions contenues dans ce guide présupposent une connaissance générale du système d'exploitation Windows et du tableur Excel, notamment :

- *familiarité avec l'ordinateur et la souris*
- *compréhension des termes icônes, cliquer, double-clic, menu, fenêtre, commande, objet, etc.*
- *notions élémentaires de structure de répertoires et désignation des fichiers*

Si vous avez besoin d'aide

Un service d'assistance technique est proposé gratuitement à tous les utilisateurs enregistrés de StatTools dotés d'un plan de maintenance à jour, ou sur forfait à l'incident. Pour assurer que vous êtes bien un utilisateur enregistré de StatTools, **enregistrez-vous en ligne sur <http://www.palisade.com/support/register.asp>**.

Si vous nous contactez par téléphone, soyez prêt à nous communiquer le numéro de série de vos outils et gardez votre guide d'utilisation à portée de main. Nous pourrions vous être d'une meilleure assistance si vous vous trouvez face à votre ordinateur, prêt à exécuter les commandes du programme.

Avant d'appeler

Avant d'appeler le service d'assistance technique, passez en revue la liste de contrôle suivante :

- *Avez-vous consulté l'aide en ligne ?*
- *Avez-vous consulté ce manuel et passé en revue le didacticiel multimédia en ligne ?*
- *Avez-vous consulté le fichier LISEZMOI.WRI ? Il contient des informations sur StatTools non disponibles lors de la composition du manuel.*
- *Pouvez-vous reproduire le problème de manière cohérente ? Pouvez-vous reproduire le problème sur un autre ordinateur ou avec un autre modèle ?*
- *Avez-vous consulté notre site Web, à l'adresse <http://www.palisade.com> ? Vous y trouverez notre dernier fichier FAQ (base de données consultable de questions et réponses techniques) et les correctifs StatTools dans la section de support technique. Il est utile de consulter régulièrement notre site pour obtenir les dernières informations publiées sur StatTools et sur les autres logiciels Palisade.*

**Contacteur
Palisade**

Vos questions, commentaires ou suggestions relatifs à StatTools sont les bienvenus ! Vous pouvez prendre contact avec notre personnel d'assistance technique par l'une des méthodes suivantes :

- *Courriel : support@palisade.com*
- *Téléphone : +1-607-277-8000, du lundi au vendredi, de 9 à 17 heures, heure de l'Est des États-Unis. Suivez les instructions données pour joindre l'Assistance technique (Technical Support).*
- *Fax : +1-607-277-8001*
- *Adresse postale :*

**Technical Support
Palisade Corporation
798 Cascadilla St.
Ithaca, NY 14850
USA**

Palisade Europe :

- *Courriel : support@palisade-europe.com*
- *Téléphone : +44 1895 425050 (Royaume-Uni)*
- *Fax : +44 1895 425051 (Royaume-Uni).*
- *Adresse postale :*

**Palisade Europe
31 The Green
West Drayton
Middlesex
UB7 7PN
Royaume-Uni**

Palisade Asie-Pacifique :

- *Courriel : support@palisade.com.au*
- *Téléphone : +61 2 9929 9799 (Australie)*
- *Fax : +61 2 9954 3882 (Australie)*
- *Adresse postale :*

**Palisade Asia-Pacific Pty Limited
Suite 101, Level 1
8 Cliff Street
Milsons Point NSW 2061
Australie**

Quelle que soit la méthode choisie, veillez à indiquer le nom de votre produit, sa version et son numéro de série. La version exacte de votre produit est indiquée sous la commande Aide, À propos de... du menu StatTools proposé dans Excel.

Version étudiants

L'assistance téléphonique n'est pas disponible pour la version étudiants de StatTools. Si vous avez besoin d'aide, procédez de l'une des manières suivantes :

- ◆ *Consultez votre professeur ou assistant.*
- ◆ *Consultez le fichier FAQ sur <http://www.palisade.com>.*
- ◆ *Adressez-vous au service d'assistance technique par courriel ou par fax.*

Configuration requise :

Configuration requise pour l'installation de StatTools pour Microsoft Excel pour Windows :

- *PC Pentium ou mieux avec disque dur.*
- *Microsoft Windows 2000 SP4, Windows XP ou version ultérieure.*
- *Microsoft Excel 2000 ou version ultérieure.*

Installation

Généralités

Le programme d'installation copie les fichiers système StatTools dans un répertoire spécifié du disque dur. Sous Windows 2000 ou version ultérieure :

- 1) *Insérez le CD-ROM StatTools ou DecisionTools Suite dans le lecteur CD-ROM.*
- 2) *Cliquez sur le bouton Démarrer, puis sur Paramètres et enfin sur Panneau de configuration.*
- 3) *Cliquez deux fois sur l'icône Ajout/Suppression de programmes.*
- 4) *Cliquez sur le bouton Installer de l'onglet Installation/désinstallation.*
- 5) *Suivez les instructions d'installation affichées à l'écran.*

En cas de problème, vérifiez que vous disposez d'un espace suffisant sur le disque prévu pour l'installation. Après avoir libéré l'espace disque requis, essayez de réexécuter l'installation.

Suppression de StatTools de l'ordinateur

Pour désinstaller StatTools, utilisez l'utilitaire Ajout/Suppression de programmes du Panneau de configuration et sélectionnez l'entrée correspondant à StatTools.

Configuration des icônes ou raccourcis StatTools

Sous Windows, l'installation crée automatiquement une commande StatTools dans le menu Programmes\Palisade DecisionTools de la barre des tâches. Si toutefois vous rencontrez des problèmes en cours d'installation ou que vous souhaitez exécuter cette opération ultérieurement, procédez comme suit : Remarquez que les instructions ci-dessous concernent Windows XP Professional. Celles applicables aux autres systèmes d'exploitation varieront peut-être.

- 1) *Cliquez sur le bouton Démarrer et pointez sur Paramètres.*
- 2) *Cliquez sur Barre des tâches et Menu Démarrer, puis sur l'onglet Menu Démarrer.*
- 3) *Cliquez sur Personnaliser, Ajouter, puis sur Parcourir.*
- 4) *Repérez le fichier StatTools.EXE, cliquez dessus puis sur OK.*
- 5) *Cliquez une fois sur Suivant, puis deux fois sur le menu de votre choix.*
- 6) *Tapez le nom « StatTools » et cliquez sur Terminer.*
- 7) *Cliquez sur OK dans toutes les boîtes de dialogue ouvertes.*

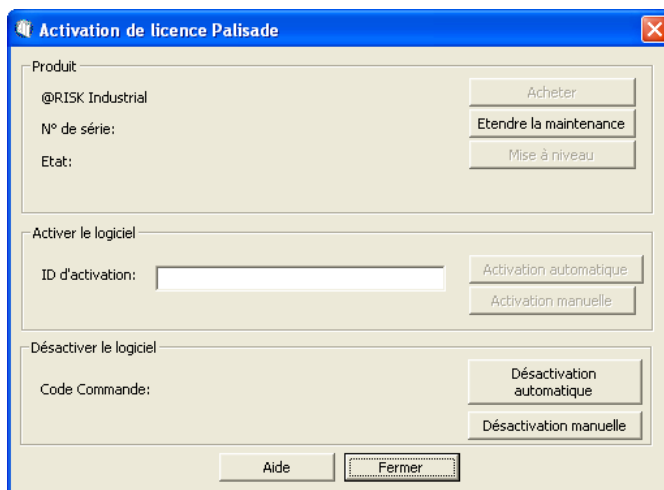
DecisionTools Suite

StatTools fait partie des outils d'analyse de risque et de décision DecisionTools Suite de Palisade Corporation. L'installation par défaut de StatTools place le programme dans un sous-répertoire du répertoire principal « Program Files\Palisade », de la même manière qu'Excel s'installe généralement dans un sous-répertoire du répertoire « Microsoft Office ».

Ce sous-répertoire de Program Files\Palisade devient le répertoire StatTools (appelé, par défaut, StatTools5). Ce répertoire contient le fichier programme de StatTools (STATTOOLS.XLA), plus les modèles types et autres fichiers nécessaires à l'exécution de StatTools. Un autre sous-répertoire de Program Files\Palisade, intitulé SYSTEM, reçoit les fichiers nécessaires à tous les programmes de la série DecisionTools Suite, y compris les fichiers d'aide et bibliothèques communs.

Activation du logiciel

L'activation est une opération de vérification de licence exigée, une seule fois, pour l'exploitation de votre logiciel sous pleine autorisation. Votre **ID d'activation** (séquence de type « 19a0-c7c1-15ef-1be0-4d7f-cd ») figure sur la facture qui vous a été envoyée par courrier ou par courriel. Si vous entrez ce code au moment de l'installation, votre logiciel s'active dès la première exécution et aucune autre intervention n'est nécessaire. Pour activer le logiciel après l'installation, choisissez la commande Activation de licence dans le menu d'aide de StatTools et entrez votre code d'activation dans la boîte de dialogue d'activation qui s'affiche.



Foire aux questions

1) Que se passera-t-il si mon logiciel n'est pas activé ?

Si vous n'entrez pas de code d'activation lors de l'installation ou que vous installez une version d'essai, votre logiciel s'exécutera en tant que tel et sera soumis aux limites de temps/nombre d'ouvertures applicables. Pour disposer d'un logiciel sous licence pleinement autorisée, vous devrez l'activer sous le code d'activation approprié.

2) Pendant combien de temps puis-je utiliser le logiciel avant de l'activer ?

Le logiciel non activé s'exécute pendant 15 jours. Toutes les fonctions sont accessibles, mais la boîte de dialogue d'activation de la licence s'ouvre à chaque démarrage du programme pour vous indiquer le temps d'exploitation restant sans activation. Au bout de la période d'essai de 15 jours, le logiciel ne s'exécutera plus que s'il est activé.

3) Comment vérifier l'état d'activation de mon logiciel ?

La commande Activation de licence du menu d'aide de StatTools donne accès à la boîte de dialogue d'activation. Le logiciel activé y figure sous l'état **Activé** et la version d'essai, sous l'état **Non activé**. Si le logiciel n'est pas activé, la durée restante de la période d'essai est indiquée.

4) Comment activer mon logiciel ?

En l'absence de code d'activation, cliquez sur le bouton Acheter de la boîte de dialogue Activation de licence. En cas d'achat en ligne, vous recevrez immédiatement un code d'activation et un lien (facultatif) de téléchargement du programme d'installation, au cas où la réinstallation du logiciel serait nécessaire. Pour acheter StatTools par téléphone, prenez contact avec votre représentation Palisade locale, au numéro indiqué dans ce chapitre sous **Contacteur Palisade**.

L'activation peut se faire sur Internet ou par courriel :

- **Si vous avez accès à Internet**

Dans la boîte de dialogue Activation de licence, tapez ou collez votre code d'activation et cliquez sur « Automatique via Internet ». Un message de confirmation devrait s'afficher après quelques secondes et la boîte de dialogue Activation de licence doit refléter l'état activé du logiciel.

- **Si vous n'avez pas accès à Internet**

Pour activer votre logiciel par courriel, procédez comme suit :

1. Cliquez sur « **Activation manuelle** » pour ouvrir le fichier de demande request.xml, à enregistrer sur disque ou copier dans le Presse-Papiers Windows. (Ne manquez pas de noter le lieu d'enregistrement de ce fichier sur votre ordinateur.)
2. **Copiez ou joignez le fichier XML** à un courriel adressé à *activation@palisade.com*. Vous devriez recevoir rapidement une confirmation automatique par retour de courriel.
3. **Enregistrez le fichier response.xml** joint au courriel de réponse sur votre disque dur.
4. **Cliquez sur le bouton Traiter** qui apparaît maintenant dans la boîte de dialogue d'activation de licence Palisade et naviguez jusqu'au fichier response.xml. Sélectionnez le fichier et cliquez sur OK.

Un message de confirmation devrait apparaître et la boîte de dialogue Activation de licence doit refléter l'état activé du logiciel.

5) Comment transférer ma licence logicielle sur un autre ordinateur ?

Le transfert d'une licence, ou **réhébergement**, peut s'effectuer en deux étapes à travers la boîte de dialogue Activation de licence de Palisade : par *désactivation* sur le premier ordinateur, puis *activation* sur le second. Un exemple type de réhébergement consiste à transférer StatTools d'un PC de bureau sur portable. Pour transférer la licence de *l'ordinateur1* à *l'ordinateur2*, veillez à ce que le logiciel soit installé sur les deux ordinateurs et à ce que les deux soient connectés à Internet pendant l'opération de désactivation/activation.

1. Sur *l'ordinateur1*, cliquez sur **Désactivation automatique** dans la boîte de dialogue Activation de licence. Attendez que s'affiche le message de confirmation.
2. Sur *l'ordinateur2*, cliquez sur **Activation automatique**. Attendez que s'affiche le message de confirmation.

Si les ordinateurs n'ont pas accès à Internet, suivez la procédure décrite plus haut pour l'activation par courriel.

6) J'ai accès à Internet mais je ne réussis pas à activer/désactiver automatiquement.

Votre pare-feu doit être configuré de manière à autoriser l'accès TCP au serveur de licences. Pour les installations mono-utilisateur (hors réseau), il s'agit de <http://service.palisade.com:8888> (port TCP 8888 sur <http://service.palisade.com>).

Chapitre 2 : Présentation de StatTools

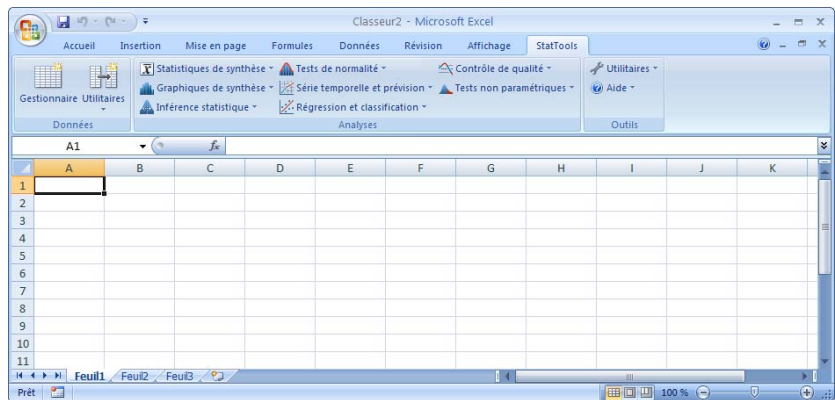
Introduction	15
Menu et barre d'outils StatTools.....	15
Ensembles de données et Gestionnaire des données.....	16
Données multi-plages.....	17
Données empilées et désempilées.....	17
Traitement des valeurs manquantes.....	18
Rapports et graphiques StatTools.....	19
Formules vs valeurs.....	20
Commentaires dans les cellules.....	21

Introduction

StatTools introduit une grande puissance statistique dans l'environnement familier de Microsoft Excel. Les procédures StatTools -- telles que la création de diagrammes de dispersion, le test d'une variable en normalité et l'exécution d'une analyse de régression -- s'exécutent sur les données Excel existantes. Les rapports et graphiques des analyses effectuées se créent dans Excel aussi.

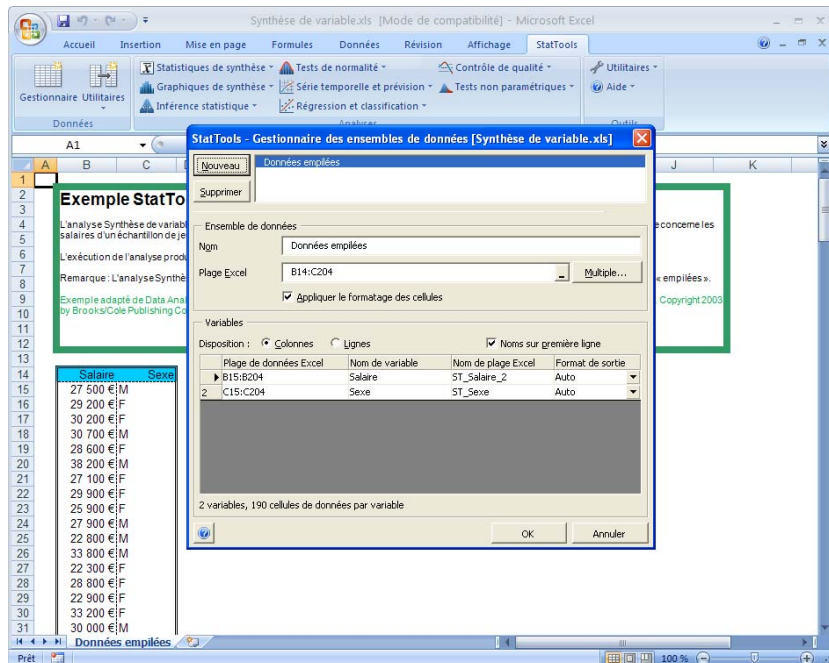
Menu et barre d'outils StatTools

Une fois le programme installé, le menu et les commandes de StatTools s'ajoutent à la barre de menus des versions Excel 2003 et antérieures. Une barre d'outils StatTools s'affiche aussi. Dans les versions 2007 et ultérieures, un ruban StatTools s'affiche.



Ensembles de données et Gestionnaire des données

StatTools est similaire à la plupart des progiciels statistiques autonomes en ce qu'il se structure autour de variables. Pour la plupart des analyses, on travaille sur un ensemble de données, ou un ensemble de variables statistiques, souvent disposé dans des colonnes contiguës, avec les noms de variable figurant sur la première ligne de l'ensemble. Le **Gestionnaire des ensembles de données** de StatTools sert à définir les ensembles de données et les variables. Les variables prédéfinies peuvent ensuite servir aux analyses statistiques, sans exiger la resélection continue des données à analyser.



Dans un ensemble de données, chaque variable est désignée par un nom et est associée à une plage de cellules Excel. La disposition typique des variables est **une variable par colonne**, mais elles peuvent aussi être disposées par ligne. Un ensemble de données peut comporter plusieurs blocs de cellules et permettre la disposition des données sur différentes feuilles d'un même classeur.

Lors de la définition d'un ensemble de données, StatTools tente d'identifier les variables dans un bloc de cellules voisin de la sélection Excel courante. Il facilite et accélère ainsi la configuration d'un ensemble où les noms de variable s'inscrivent sur la première ligne et les variables se disposent en colonnes.

Toutes les colonnes de l'ensemble de données ne doivent pas nécessairement être de longueur égale. Deux variables, *Poids_Hommes* et *Poids_Femmes* par exemple, peuvent présenter différents nombres d'observations. Pour de nombreuses analyses toutefois, StatTools traite les cellules blanches des colonnes plus courtes comme données manquantes.

Données multi-plages

Une colonne de feuille de calcul Excel 2003 ou version antérieure admet un maximum de 65 536 points de données pour une variable. En présence de valeurs plus nombreuses, si le passage à Excel 2007 n'est pas possible, StatTools admet l'affectation de plusieurs plages de cellules à un même ensemble de données. Ainsi, il est possible de « répéter » un ensemble de données sur plusieurs feuilles de calcul, en affectant les mêmes colonnes des différentes feuilles aux valeurs de l'ensemble. Cette capacité permet également d'affecter différents blocs de cellules d'une même feuille de calcul à un même ensemble de données. L'approche est utile quand les données sont réparties en différents endroits d'une même feuille.

Données empilées et désempilées

StatTools gère les données empilées et désempilées. Pour certaines procédures statistiques, il est plus facile de travailler avec des données empilées et pour d'autres, avec des données désempilées. Par exemple, si l'on compare les revenus moyens des ménages de différents quartiers, l'approche désempilée présenterait une variable (ou colonne) *Revenu* distincte pour chaque quartier. Sous le format empilé, il y aurait une variable valeur *Revenu* et une variable catégorie *Quartier* indiquant le quartier de chaque ménage.

L'utilitaire d'empilage de StatTools permet d'empiler les variables en deux colonnes : une colonne de valeur, *Revenu*, et une colonne de catégorie, *Quartier*. Suivant le type d'analyse, l'ensemble de données empilé peut être plus facile à manier que sa version désempilée.

**Traitement des
valeurs
manquantes**

Si l'ensemble de données présente des valeurs manquantes (comme c'est fréquemment le cas dans l'analyse statistique), StatTools traite ces données de la manière appropriée, suivant la tâche. Ainsi, les mesures de synthèse telles que la moyenne et l'écart type ignorent les valeurs manquantes. En revanche, une analyse de régression impliquant trois variables n'utilise que les lignes de l'ensemble de données qui ne présentent aucune donnée manquante pour aucune des trois variables. Autre exemple encore, dans le cas d'un diagramme de dispersion à deux variables, seuls les points pour lesquels les deux variables ne présentent aucune valeur manquante sont tracés.

Remarque : Certaines procédures StatTools n'admettent pas les valeurs manquantes. Voir la manière dont chaque procédure traite les valeurs manquantes dans la section Référence de ce manuel.

Rapports et graphiques StatTools

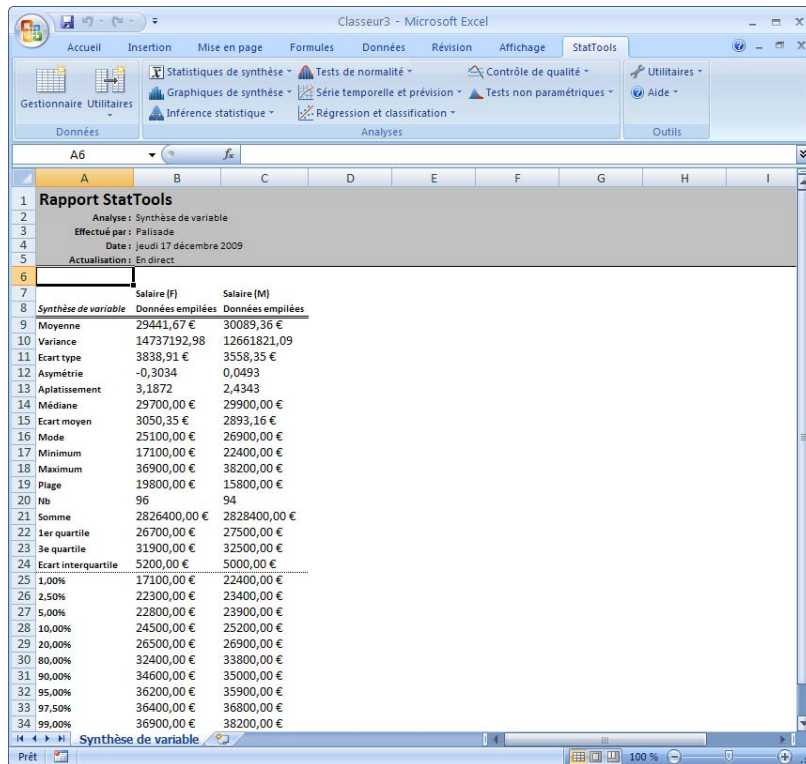
Pour chaque sortie numérique (rapport d'analyse de régression ou tableau de statistiques de synthèse), StatTools propose différentes options de destination du rapport :

- **Nouveau classeur** : un nouveau classeur se crée (si nécessaire) et chaque rapport y aboutit sur une nouvelle feuille de calcul.
- **Nouvelle feuille du classeur actif** : chaque rapport aboutit sur une nouvelle feuille du classeur actif.
- **Après la dernière colonne utilisée** : chaque rapport aboutit sur la feuille active, à droite de la dernière colonne utilisée.
- **À partir de la cellule** : l'utilisateur sélectionne la cellule destinée à marquer le coin supérieur gauche du rapport ou du graphique.

Les graphiques créés par StatTools accompagnent les rapports. Les graphiques se créent au format Excel et peuvent être personnalisés à l'aide des commandes graphiques Excel standard.

Formules vs valeurs

Par défaut, StatTools essaie de rendre les résultats aussi « actuels » que possible. Dans la mesure du possible, les rapports sont dotés de formules qui les relient aux données originales. Supposons par exemple une variable *Poids*, pour laquelle on désire des mesures de synthèse, telles que la moyenne et l'écart type. La procédure Statistiques de synthèse donne à la plage des poids le nom *Poids* et entre les formules dans les cellules de sortie : **=StatMean(Poids)** et **=StatStdDev(Poids)**. StatMean et StatStdDev sont des fonctions StatTools intégrées de calcul de la moyenne et de l'écart type. Elles remplacent les fonctions Excel standard intégrées correspondantes.



The screenshot shows the Microsoft Excel interface with the StatTools ribbon active. The ribbon includes options for 'Statistiques de synthèse', 'Graphiques de synthèse', 'Inférence statistique', 'Tests de normalité', 'Série temporelle et prévision', 'Régression et classification', 'Contrôle de qualité', 'Tests non paramétriques', 'Utilitaires', and 'Aide'. The main window displays a report titled 'Rapport StatTools' for 'Analyse : Synthèse de variable'. The report includes the following data:

Synthèse de variable	Données empliées	Données empliées
Moyenne	29441,67 €	30089,36 €
Variance	14737192,98	12661821,09
Ecart type	3838,91 €	3558,35 €
Asymétrie	-0,3034	0,0493
Aplatissement	3,1872	2,4343
Médiane	29700,00 €	29900,00 €
Ecart moyen	3050,35 €	2893,16 €
Mode	25100,00 €	26900,00 €
Minimum	17100,00 €	22400,00 €
Maximum	36900,00 €	38200,00 €
Plage	19800,00 €	15800,00 €
Nb	96	94
Somme	2826400,00 €	2828400,00 €
1er quartile	26700,00 €	27500,00 €
3e quartile	31900,00 €	32500,00 €
Ecart interquartile	5200,00 €	5000,00 €
1,00%	17100,00 €	22400,00 €
2,50%	22300,00 €	23400,00 €
5,00%	22800,00 €	23900,00 €
10,00%	24500,00 €	25200,00 €
20,00%	26500,00 €	26900,00 €
30,00%	32400,00 €	33800,00 €
40,00%	34600,00 €	35000,00 €
50,00%	36200,00 €	35900,00 €
60,00%	36400,00 €	36800,00 €
70,00%	36900,00 €	38200,00 €

Les formules remplissent deux rôles dans les rapports : Elles aident à comprendre les procédures statistiques et les fonctions StatTools dans Excel. On ne voit pas simplement un résultat numérique : on voit comment il se forme. Elles présentent ensuite l'avantage qu'en cas de changement des données, les résultats changent automatiquement, sans qu'il soit nécessaire de réexécuter la procédure.

L'approche n'est cependant pas toujours pratique. Dans le cas des régressions, notamment. StatTools ne fournit pas les formules utilisées pour créer une sortie de régression ; il n'en produit que les résultats numériques. Si les données changent, il faut donc réexécuter les procédures.

StatTools offre aussi l'option de désactiver l'actualisation en direct. Cette désactivation est utile si le temps de recalcul Excel est excessif lors du changement des données.

**Commentaires
dans les cellules**

Il est possible, dans Excel, d'ajouter un commentaire « contextuel » à une cellule. Un petit triangle rouge dans le coin supérieur droit de la cellule indique la présence d'un commentaire. Pour le lire, il suffit de placer le curseur sur la cellule. StatTools tire parti de cette fonction pour l'ajout d'une aide contextuelle : le nec plus ultra de l'aide en ligne !

Si vous vous trouvez jamais face à une feuille de calcul dont les commentaires restent toujours visibles et masquent les données, choisissez Outils/Options puis, sous l'onglet Afficher, cliquez sur le bouton Indicateur de commentaire seulement.

Chapitre 3 : Guide de référence StatTools

Introduction	27
Langage macro VBA StatTools et Kit du développeur	27
Référence : Icônes StatTools	29
Barre d'outils StatTools	29
Référence : Commandes du menu StatTools	33
Introduction	33
Packs d'analyse complémentaires.....	33
Liste des commandes	35
Menu StatTools – Ensembles de données	43
Commande Gestionnaire des ensembles de données	43
Menu Utilitaires de données	49
Commande Empiler.....	49
Commande Désempiler	52
Commande Transformer.....	53
Commande Retard	55
Commande Différence	57
Commande Interaction.....	59
Commande Combinaison.....	61
Commande Var. nominale.....	63
Commande Échantillons aléatoires	65
Menu Statistiques de synthèse	67
Commande Synthèse de variable.....	67
Commande Corrélation et covariance	70
Menu Graphiques de synthèse	73
Commande Histogramme.....	73
Commande Diagramme de dispersion	76
Commande Boîte à moustaches.....	78
Menu Inférence statistique	81

Commande Intervalle de confiance - Moyenne/Écart type	81
Commande Intervalle de confiance - Proportion.....	84
Commande Test d'hypothèse - Moyenne / Écart type	87
Commande Test d'hypothèse - Proportion	90
Commande Sélection de taille d'échantillon	93
Commande ANOVA simple	95
Commande ANOVA double.....	98
Commande Test d'indépendance chi carré	101
Menu Tests de normalité	103
Commande Test de normalité chi carré.....	103
Commande Test de Lilliefors.....	107
Commande Graphique Q-Q normal.....	109
Menu Série temporelle et prévision.....	111
Commande Chronogramme	111
Commande Autocorrélation.....	113
Commande Test des runs.....	115
Commande Prévision	117
Menu Régression et classification	123
Commande Régression.....	124
Commande Régression logistique	129
Commande Analyse discriminante.....	133
Menu Contrôle de qualité	139
Commande Diagramme de Pareto	140
Commande Graphiques X/R	143
Commande Graphique P	147
Commande Graphique C.....	151
Commande de graphique U	154
Menu Tests non paramétriques	159
Commande Test des signes	161
Commande Test de rang de Wilcoxon.....	164
Commande Test de Mann-Whitney.....	167
Menu Utilitaires.....	171
Commande Paramètres d'application.....	171
Commande Supprimer les ensembles de données	177
Commande Effacer la mémoire de dialogue	177
Commande Télécharger StatTools.....	177
Menu Aide	179
Aide StatTools.....	179
Manuel en ligne.....	179
Commande Activation de licence.....	179

Commande À propos.....	179
Référence : Fonctions StatTools	181
Introduction	181
Fonctions StatTools vs Fonctions Excel	181
Fonctions de distribution.....	183
Rapports « en direct »	184
Référence : Liste des fonctions statistiques	186
Fonctions disponibles	186
Description détaillée des fonctions	189

Introduction

Ce chapitre décrit les icônes, les commandes et les fonctions statistiques de StatTools. Il se divise en quatre sections :

- 1) *Référence : Icônes StatTools*
- 2) *Référence : Synthèse des procédures StatTools*
- 3) *Référence : Commandes StatTools*
- 4) *Référence : Fonctions StatTools*

Langage macro VBA StatTools et Kit du développeur

StatTools s'accompagne aussi d'un puissant langage macro VBA utile

- 1) **à l'automatisation des analyses StatTools**
- 2) **au développement de nouvelles analyses statistiques tirant parti du gestionnaire d'ensembles de données, des rapports et des graphiques StatTools.** Ces calculs personnalisés peuvent assurer des analyses absentes des procédures intégrées de StatTools. Au besoin, ces analyses peuvent être ajoutées au menu et aux barres d'outils de StatTools.

Pour plus de détails sur le langage macro VBA StatTools et le kit du développeur, voir la documentation en ligne fournie avec le logiciel.

Référence : Icônes StatTools

Barre d'outils StatTools

Les icônes StatTools servent à définir les ensembles de données et les variables, puis à exécuter des procédures statistiques sur ces variables. Ces icônes figurent sur la « barre d'outils » du tableur, sous forme de barre d'outils personnalisée dans Excel 2003 et versions antérieures ou de ruban dans Excel 2007. Cette section décrit brièvement chaque icône : elle explique les fonctions exécutées par chacune et la commande de menu équivalente.

Les icônes décrites ci-dessous figurent sur la barre d'outils StatTools installée dans Excel 2003 et versions antérieures.

Icône Fonction et commande équivalente



Définir - ou modifier ou supprimer - un ensemble de données et des variables

Commande équivalente : Gestionnaire des ensembles de données



Exécuter un utilitaire de données

Commande équivalente : Utilitaires



Exécuter une procédure de statistiques de synthèse

Commande équivalente : Statistiques de synthèse



Créer des graphiques de synthèse des variables

Commande équivalente : Graphiques de synthèse



Exécuter une procédure d'inférence statistique

Commande équivalente : Inférence statistique



Exécuter un test de normalité sur les variables

Commande équivalente : Tests de normalité



Exécuter une procédure de série temporelle ou prévision

Commande équivalente : Série temporelle et prévision



Exécuter une procédure de régression ou classification

Commande équivalente : Régression et classification



Exécuter une procédure de contrôle de qualité

Commande équivalente : Contrôle de qualité



Exécuter un test non paramétrique

Commande équivalente : Tests non paramétriques



Afficher les utilitaires StatTools

Commande équivalente : Utilitaires



Afficher le fichier d'aide de StatTools

Commande équivalente : Aide

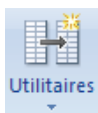
Les icônes décrites ci-dessous figurent sur le ruban StatTools installé dans Excel 2007.

Icône Fonction et commande équivalente



Définir - ou modifier ou supprimer - un ensemble de données et des variables

Commande équivalente : Gestionnaire des ensembles de données



Exécuter un utilitaire de données

Commande équivalente : Utilitaires



Exécuter une procédure de statistiques de synthèse

Commande équivalente : Statistiques de synthèse



Créer des graphiques de synthèse des variables

Commande équivalente : Graphiques de synthèse

**Exécuter une procédure d'inférence statistique**

Commande équivalente : Inférence statistique

**Exécuter un test de normalité sur les variables**

Commande équivalente : Tests de normalité

**Exécuter une procédure de série temporelle ou prévision**

Commande équivalente : Série temporelle et prévision

**Exécuter une procédure de régression ou classification**

Commande équivalente : Régression et classification

**Exécuter une procédure de contrôle de qualité**

Commande équivalente : Contrôle de qualité

**Exécuter un test non paramétrique**

Commande équivalente : Tests non paramétriques

**Afficher les utilitaires StatTools**

Commande équivalente : Utilitaires

**Afficher le fichier d'aide de StatTools**

Commande équivalente : Aide

Référence : Commandes du menu StatTools

Introduction

Cette section du Guide de référence décrit les commandes StatTools telles qu'elles figurent dans le menu StatTools (Excel 2003 et versions antérieures) ou sur le ruban StatTools d'Excel (versions 2007 et ultérieures). Les commandes sont expliquées dans leur ordre d'apparition dans le menu, en commençant par Gestionnaire des ensembles de données. Les icônes StatTools donnent accès à beaucoup des commandes. La section **Référence : Icônes StatTools** présentée dans ce chapitre identifie la commande équivalente de chaque icône **StatTools**.

Plusieurs commandes **StatTools** sont également disponibles dans un menu contextuel flottant invoqué d'un clic du bouton droit de la souris dans Excel.

Packs d'analyse complémentaires

La version StatTools Industrial peut être complétée de « packs d'analyse » complémentaires, contenant de nouvelles analyses absentes du logiciel de base. Les commandes trouvées dans ces ensembles s'ajoutent au menu StatTools. Si vous utilisez les packs d'analyse complémentaires, il se peut donc que votre menu ne se présente pas exactement comme décrit ici. Pour plus de détails sur le développement de procédures complémentaires pour StatTools, voir la documentation en ligne relative au langage macro VBA StatTools et au kit du développeur.

Liste des commandes

Les procédures proposées dans StatTools s'organisent en groupes naturels. Chaque groupe représente une entrée de menu StatTools. Si un groupe comporte plus d'un élément, ces éléments se listent dans un sous-menu. Cette section présente une brève description de chaque procédure comprise dans chaque groupe. Pour plus de détails, voir la section **Référence : Commandes StatTools**.

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multi-plages ?	Données incorrectes	Nbre variables
Statistiques de synthèse							
Commande Synthèse de variable	Génère des statistiques de synthèse incluant les mesures habituelles de moyenne, médiane et écart type, ainsi que différentes options de quartile et centile.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-100
Commande Corrélation et covariance	Crée une table des corrélations et/ou des covariances pour l'ensemble de variables sélectionnées.	Admises au début, au milieu et à la fin des données.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	Omises.	1-250
Graphiques de synthèse							
Commande Histogramme	Crée un histogramme pour chaque variable sélectionnée. Offre l'option de définir les catégories ou « intervalles » de l'histogramme.	Admises au début, au milieu et à la fin des données.	Partiellement en direct – les changements de données actualisent le graphique quand les données se trouve dans la plage de l'axe X du graphique.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-100
Commande Diagramme de dispersion	Crée un diagramme de dispersion pour chaque paire de variables sélectionnée.	Admises au début, au milieu et à la fin des données.	En direct.	Données désempilées seulement. Max 32 000 cas.	Non.	Non admises.	1-10
Commande Boîte à moustaches	Crée une simple boîte (pour une variable) ou plusieurs boîtes l'une à côté de l'autre (pour plusieurs variables).	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Non.	Omises.	1-10

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multiples ?	Données incorrectes	Nbre variables
Inférence statistique							
Commande Intervalle de confiance – Moyenne/Écart type	Calcule un intervalle de confiance pour la moyenne et l'écart type de variables simples, ou la différence entre les moyennes pour les paires de variables. Les intervalles de confiance se calculent par analyse à un échantillon, à deux échantillons ou à échantillons appariés.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).
Commande Intervalle de confiance – Proportion	Analyse la proportion d'éléments d'un échantillon qui appartient à une catégorie donnée (analyse à un échantillon) ou compare deux échantillons concernant la proportion d'éléments compris dans une catégorie donnée (analyse à deux échantillons).	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).
Commande Test d'hypothèse – Moyenne / Écart type	Effectue des tests d'hypothèse pour la moyenne et l'écart type de variables simples, ou la différence entre les moyennes pour les paires de variables. Les tests s'effectuent par analyse à un échantillon, à deux échantillons ou à échantillons appariés.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).
Commande Test d'hypothèse – Proportion	Analyse la proportion d'éléments d'un échantillon qui appartient à une catégorie donnée (analyse à un échantillon) ou compare deux échantillons concernant la proportion d'éléments compris dans une catégorie donnée (analyse à deux échantillons).	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multi-plages ?	Données incorrectes	Nbre variables
Commande Sélection de taille d'échantillon	Détermine la ou les tailles d'échantillon nécessaires à l'obtention d'un intervalle de confiance à demi-longueur prescrite. La procédure s'effectue pour les intervalles de confiance d'une moyenne, d'une proportion, de la différence entre deux moyennes et de la différence entre deux proportions.	n/a	n/a	n/a	n/a	n/a	n/a
Commande ANOVA simple	Extension de l'analyse à deux échantillons pour la comparaison de deux moyennes de population. Teste l'égalité de deux moyennes ou davantage.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	2-50
Commande ANOVA double	Effectue une analyse ANOVA double. Cette analyse s'effectue généralement dans le contexte d'un concept expérimental à deux « facteurs » réglés, chacun, à différents « niveaux de traitement ».	Non admises.	En direct.	Données empilées. Max 16M cas. L'expérience doit être équilibrée.	Oui.	Non admises.	2 variables catégorielles, 1 variable valeur.
Commande Test d'indépendance chi carré	Détermine par test chi carré si les attributs de ligne et de colonne d'un tableau de contingence sont statistiquement indépendants.	Non.	En direct (pour autant que la taille du tableau ne change pas).	n/a	n/a	n/a	n/a
Tests de normalité							
Commande Test de normalité chi carré	Effectue un test de normalité chi carré pour une variable sélectionnée.	Admises au début, au milieu et à la fin des données.	Partiellement en direct (le positionnement des intervalles ne change pas mais l'occupation et les graphiques, si).	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1
Commande Test de Lilliefors	Offre un test de normalité plus puissant que le test de qualité d'ajustement chi carré. (La puissance du test tient au fait qu'il est plus susceptible de détecter la non-normalité.)	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-10

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multiples ?	Données incorrectes	Nbre variables
Commande Graphique Q-Q normal	Crée un tracé double quantile (Q-Q) pour une variable sélectionnée. Test informel de normalité.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Non.	Omises.	1
Série temporelle et prévision							
Commande Chronogramme	Crée un chronogramme d'une ou plusieurs variables de série temporelle (toutes sur un même graphique).	Admises au début, au milieu et à la fin des données.	En direct.	Données désempilées. Max 32 000 cas.	Non.	Non admises.	1-100
Commande Autocorrélation	Calcule un nombre quelconque d'autocorrélations pour une variable de série temporelle, indique (le cas échéant) celles significativement non nulles et trace (facultativement) un graphique à barres (appelé corrélogramme) des autocorrélations.	Admises au début ou à la fin des données.	En direct.	Données désempilées. Max 32 000 cas.	Non.	Non admises.	1-10
Commande Test des runs	Effectue un test des runs pour vérifier si une variable (généralement de série temporelle) est aléatoire.	Admises au début ou à la fin des données.	En direct.	Données désempilées. Max 16M cas.	Oui.	Non admises.	1 ou plus.
Commande Prévision	Prévoit des données de série temporelle selon la méthode des moyennes mobiles, lissage exponentiel simple, lissage exponentiel de Holt pour la capture de tendance et lissage exponentiel de Winters pour la capture de saisonnalité.	Admises au début des données seulement.	En direct.	Données désempilées. Max 32 000 cas.	Oui.	Non admises.	1 ou plus.
Régression et classification							
Commande Régression	Exécute différents types d'analyse de régression : multiple, en escalier, avant, arrière et par bloc.	Admises au début, au milieu et à la fin des données.	Statiques.	Données désempilées. Max 16M cas.	Oui.	Non admises.	1 dépendante ; 1-250 indépendant es.
Commande Régression logistique	Effectue une analyse de régression logistique sur un ensemble de données. Il s'agit essentiellement d'un type non linéaire d'analyse de régression où la variable réponse est binaire : 0 ou 1.	Admises au début, au milieu et à la fin des données.	Statiques.	Données empilées. Max 16M cas.	Oui.	Non admises.	1 dépendante ; 1-250 indépendant es.

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multi-plages ?	Données incorrectes	Nbre variables
Commande Analyse discriminante	Effectue une analyse discriminante sur un ensemble de données. Il doit y avoir une variable « catégorielle » qui spécifie, parmi deux groupes ou davantage, celui auquel chaque observation appartient, ainsi qu'une ou plusieurs variables explicatives pouvant servir à prédire cette appartenance.	Admises au début, au milieu et à la fin des données.	Statiques.	Données emplies. Max 16M cas.	Oui.	Non admises sauf dans les variables dépendantes.	1 dépendante ; 1-250 indépendantes.
Contrôle de qualité							
Commande Diagramme de Pareto	Produit un diagramme de Pareto permettant de voir l'importance relative des données catégorisées.	Admises au début, au milieu et à la fin des données.	Statiques.	Données désemplées.	Oui.	Omises.	1 catégorie, ou 1 valeur et 1 catégorie.
Commande Graphiques X/R	Produit des graphiques de contrôle X/R permettant de voir si un processus est conforme à ses limites de contrôle statistique.	Non admises.	Statiques.	Données désemplées. Max 32 000 cas.	Non.	Non admises.	2-25
Commande Graphique P	Produit des graphiques de contrôle P permettant de voir si un processus est conforme à ses limites de contrôle statistique.	Non admises.	Statiques.	Données désemplées. Max 32 000 cas.	Non.	Non admises.	1 variable 1 variable taille.
Commande Graphique C	Produit des graphiques de contrôle C permettant de voir si un processus est conforme à ses limites de contrôle statistique.	Non admises.	Statiques.	Données désemplées. Max 32 000 cas.	Non.	Non admises.	1
Commande de graphique U	Produit des graphiques de contrôle U permettant de voir si un processus est conforme à ses limites de contrôle statistique.	Non admises.	Statiques.	Données désemplées. Max 32 000 cas.	Non.	Non admises.	1
Tests non paramétriques							
Commande Test des signes	Effectue des tests d'hypothèse pour la médiane d'une simple variable ou la médiane des différences pour une paire de variables.	Admises au début, au milieu et à la fin des données.	En direct.	Données emplies et désemplées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multi- plages ?	Données incorrectes	Nbre variables
Commande Test de rang de Wilcoxon	Effectue des tests d'hypothèse semblables à ceux du Test des signes, mais sous hypothèse de distribution de probabilités symétrique.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).
Commande Test de Mann-Whitney	Effectue un test d'hypothèse sur deux échantillons.	Admises au début, au milieu et à la fin des données.	En direct.	Données empilées et désempilées. Max 16M cas.	Oui.	Omises.	1-250 (analyse à 1 échantillon). Exactement 2 (analyse à 2 échantillons ou à échantillons appariés).
Utilitaires de données							
Commande Empiler	Prend un ensemble de données à variables distinctes pour chaque groupe dans des colonnes distinctes et permet leur « empilage » en deux colonnes : une colonne de « catégorie » et une colonne de « valeur ». Suivant le type d'analyse, l'ensemble de données empilé peut être plus facile à traiter que sa version désempilée.	Oui – partout dans la variable.	Statiques.	Données désempilées seulement. Max 65 535 cas.	Non.	n/a	1-100
Commande Désempiler	Procédure inverse à celle d'empilage.	Oui – partout dans la variable.	n/a	Données empilées seulement. Max 16M cas.	Oui.	n/a	1-32
Commande Variable nominale	Crée des variables nominales (dummy) (0-1) basées sur celles existantes.	Oui – partout dans la variable.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	n/a	1
Commande Combinaison	Crée une nouvelle variable au départ de deux variables numériques, d'une variable catégorielle et une numérique, ou de deux variables catégorielles.	Oui – partout dans la variable.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	Non admises.	2-32 d'un même ensemble de données.

Procédure	Description	Données manquantes ?	Rapports en direct vs statiques	Données	Données multi-plages ?	Données incorrectes	Nbre variables
Commande Interaction	Crée une nouvelle variable issue du produit, de la somme, de la moyenne, du minimum, du maximum ou de l'étendue min-max d'une ou plusieurs variables.	Oui – partout dans la variable.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	n/a	2 var. valeur, ou 1 valeur et 1 catégorie, ou 2 var. catégorie.
Commande Retard	Crée une nouvelle variable déphasée ou retardée sur la base d'une variable existante. Une variable retardée est tout simplement une version de la variable originale « repoussée » d'un nombre de lignes égal au retard.	Oui – partout dans la variable.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	Omises.	1
Commande Transformer	Applique l'une de quatre transformations non linéaires aux variables sélectionnées – logarithme népérien, carré, racine carrée ou réciproque – pour créer une nouvelle variable.	Oui – partout dans la variable.	En direct ou statiques.	Données désempilées seulement. Max 16M cas.	Oui.	Omises.	1-100
Commande Différence	Crée des variables de différence au départ d'une variable originale.	Oui – partout dans la variable.	En direct.	Données désempilées seulement. Max 16M cas.	Oui.	n/a	1
Commande Échantillons aléatoires	Permet la génération d'échantillons aléatoires au départ d'un ensemble de données, sous échantillonnage avec ou sans remplacement.	Oui – partout dans la variable.	Statiques.	Données empilées seulement. Max 16M cas.	Oui.	Omises.	1-32

Menu StatTools – Ensembles de données

Commande Gestionnaire des ensembles de données

Définit les ensembles de données et variables StatTools, ou modifie ou supprime un ensemble de données et des variables existants.

La commande **Gestionnaire des ensembles de données** sert à définir les ensembles de données et les variables. Les ensembles de données et variables définis ici peuvent ensuite être soumis aux analyses des procédures StatTools. Dans la boîte de dialogue du Gestionnaire, on peut ajouter ou supprimer des ensembles de données, nommer un ensemble, spécifier la disposition des variables dans un ensemble et nommer ces variables.

Définitions

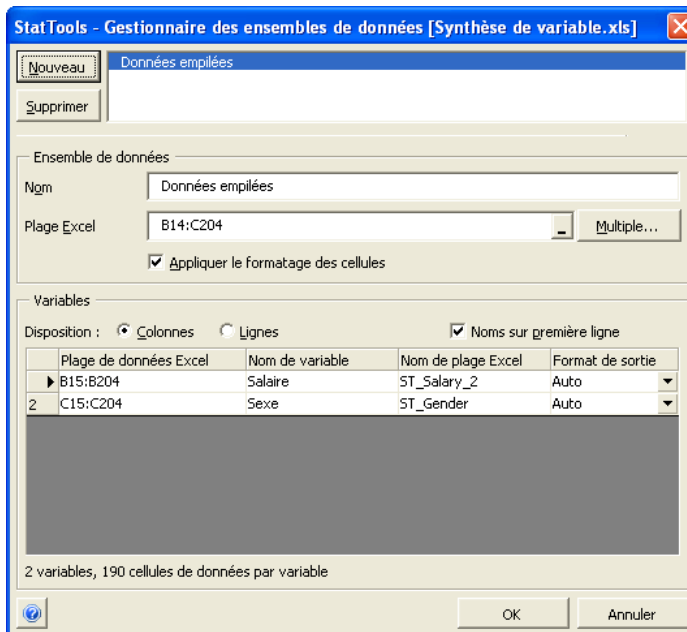
StatTools est similaire à la plupart des progiciels statistiques autonomes en ce qu'il se structure autour de variables. Pour la plupart des analyses, on travaille sur un ensemble de données, ou un ensemble de variables statistiques, souvent disposé dans des colonnes contiguës, avec les noms de variable figurant sur la première ligne de l'ensemble. Les variables prédéfinies peuvent ensuite servir aux analyses statistiques, sans exiger la resélection continue des données à analyser.

Dans un ensemble de données, chaque variable est désignée par un nom et est associée à une plage de cellules Excel. La **disposition** sélectionnée spécifie la manière dont les variables s'organisent dans l'ensemble de données. La disposition typique des variables est **Colonnes** (une variable par colonne), mais elles peuvent aussi être disposées par **Lignes**. Un ensemble de données peut comporter plusieurs blocs de cellules et permettre la disposition des données sur différentes feuilles d'un même classeur.

Lors de la définition d'un ensemble de données, StatTools tente d'identifier les variables dans un bloc de cellules voisin de la sélection Excel courante. Il facilite et accélère ainsi la configuration d'un ensemble où les noms de variable s'inscrivent sur la première ligne et les variables se disposent en colonnes.

Toutes les colonnes de l'ensemble de données ne doivent pas nécessairement être de longueur égale. Deux variables, *Poids_Hommes* et *Poids_Femmes* par exemple, peuvent présenter différents nombres d'observations. Pour de nombreuses analyses toutefois, StatTools traite les cellules blanches des colonnes plus courtes comme données manquantes.

Boîte de dialogue du Gestionnaire



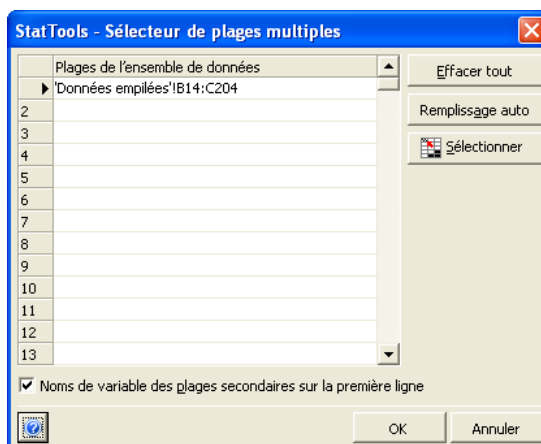
Sous le titre **Ensemble de données** de la boîte de dialogue du Gestionnaire des ensembles de données :

- **Nouveau / Supprimer** ajoute un nouvel ensemble de données, ou en supprime un existant.
- **Nom** spécifie le nom de l'ensemble de données.
- **Plage Excel** spécifie la plage Excel associée à un ensemble de données. Si plusieurs plages de cellules ont été affectées à un ensemble de données, la valeur de ce champ est précédée de l'étiquette **Multiple**.
- **Appliquer le formatage des cellules** ajoute une grille et des couleurs d'identification des ensembles de données.
- Un clic sur le bouton **Multiple** ouvre la boîte de dialogue **Sélecteur de plages multiples**. Cette boîte de dialogue permet l'entrée des plages de cellules individuelles qui constituent l'ensemble de données à plages de cellules multiples.

Ensembles de données à plages multiples

StatTools admet l'affectation de plusieurs plages de cellules à un même ensemble de données. Un ensemble de données à plages multiples peut être utile dans les cas suivants :

- 1) quand chaque variable d'un ensemble de données comporte plus de 65 536 points de données (sous Excel 2003 ou version antérieure), exigeant l'expansion des données sur plusieurs feuilles de calcul d'un même classeur ;
- 2) quand les données d'une variable se trouvent dans plusieurs blocs dispersés à travers les différentes feuilles de calcul d'un classeur.

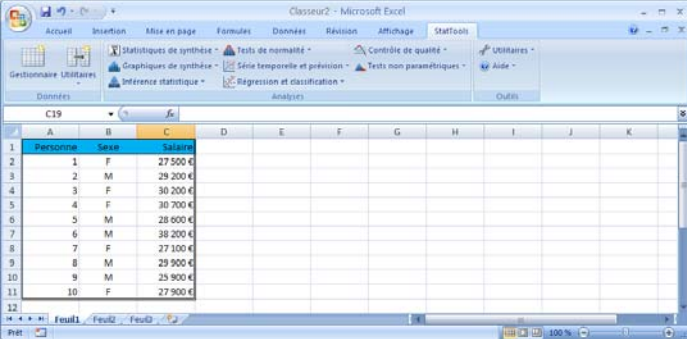


Dans la boîte de dialogue Sélecteur de plages multiples :

- **Effacer tout** efface toutes les plages entrées.
- **Remplissage auto** applique la première plage entrée (sur la première ligne) à toutes les feuilles de calcul visibles du classeur actif et entre les références *NomFeuille!PlageCellules* dans la grille.
- **Sélectionner** affiche un sélecteur pour la sélection d'un bloc de cellules à utiliser comme plage d'ensemble de données.
- **Noms de variable des plages secondaires dans la première colonne (sur la première ligne)** – Pour les ensembles de données à plages multiples, les noms de variable peuvent être étiquetés dans les colonnes (ou, suivant la disposition sélectionnée, sur les lignes) de chaque plage listée, ou dans/sur celles de la première plage sélectionnée seulement. La première plage sélectionnée est celle entrée sur la première ligne de la boîte de dialogue Sélecteur de plages multiples.

Sous le titre **Variables** de la boîte de dialogue du Gestionnaire des ensembles de données :

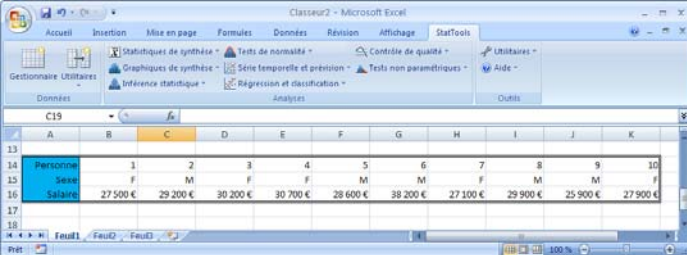
- **Disposition** spécifie la manière dont les variables sont structurées dans la plage Excel occupée par l'ensemble de données :
 - **Colonnes** représente la disposition type, dans laquelle chaque colonne de la plage Excel contient les données d'une variable. Le nom des variables est généralement indiqué en haut de chaque colonne.



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K
1	Personne	Sexe	Salaire								
2		1	F	27 500 €							
3		2	M	29 200 €							
4		3	F	30 200 €							
5		4	F	30 700 €							
6		5	M	28 600 €							
7		6	M	38 200 €							
8		7	F	27 100 €							
9		8	M	29 900 €							
10		9	M	25 900 €							
11		10	F	27 900 €							

- Sous la disposition **Lignes**, chaque ligne de l'ensemble de données contient les données d'une variable. Cette disposition est souvent choisie pour la présentation des données de série temporelle dans Excel.



The screenshot shows an Excel spreadsheet with the following data:

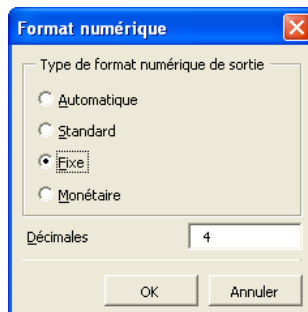
	A	B	C	D	E	F	G	H	I	J	K
13											
14	Personne	3	2	3	4	5	6	7	8	9	10
15	Sexe	F	M	F	F	M	M	F	M	M	F
16	Salaire	27 500 €	29 200 €	30 200 €	30 700 €	28 600 €	38 200 €	27 100 €	29 900 €	25 900 €	27 900 €
17											
18											

- L'option **Noms dans première colonne (ou sur première ligne)** s'applique quand les noms des variables d'un ensemble de données sont inclus en haut des colonnes (ou dans les cellules les plus à gauche si la disposition choisie est **Lignes**).

Paramètres des variables

Chaque ligne de la grille affichée dans la boîte de dialogue Gestionnaire des ensembles de données liste les variables d'un ensemble, y compris le nom de chaque variable, la plage Excel qui contient les points de données et le nom de plage Excel utilisé pour identifier les données de la variable dans les formules Excel.

- Le **Nom de plage Excel** indiqué est celui utilisé dans les formules Excel créées dans les rapports et graphiques de StatTools. Ces formules permettent l'actualisation automatique « en direct » des rapports quand les données d'une variable changent. La définition de noms de plage textuels descriptifs peut rendre les formules plus intelligibles au lecteur.
- **Format de sortie** spécifie le format des valeurs affichées pour une variable dans les rapports des analyses StatTools. Sous l'option **Auto**, StatTools sélectionne le format « optimal » en fonction du formatage numérique appliqué aux cellules contenant les valeurs de la variable dans Excel. La flèche de zone déroulante, dans la case de format de sortie, permet de sélectionner le format spécifique à utiliser :



Standard est l'équivalent du format numérique Standard d'Excel. **Fixe** affiche la précision au nombre de **Décimales** entré. **Monétaire** est l'équivalent du format numérique Monétaire d'Excel, sous précision au nombre de **Décimales** entré.

Remarque : Le format de sortie désiré peut aussi être introduit directement dans la boîte de dialogue du Gestionnaire, selon la notation *Format_sélectionné(Décimales)*. Par exemple : *Monétaire(4)*.

**Capacité
d'ensembles de
données et
variables**

En l'espace d'une session, StatTools admet :

- un maximum de 256 ensembles de données, dans un même classeur ;
- un maximum de 256 variables par ensemble de données. Toutes les données d'un même ensemble de données doivent se trouver dans le même classeur ;
- un maximum de 16 777 216 points de données par variable.

Les capacités de données effectives peuvent être inférieures aux valeurs indiquées ci-dessus suivant la configuration du système et la version d'Excel. Les analyses StatTools en soi peuvent être soumises à d'autres limites. Les limites de mémoire d'Excel même peuvent aussi affecter ces capacités.

Remarque : La boîte de dialogue Gestionnaire des ensembles de données liste tous les ensembles de données et toutes les variables du classeur actif (dont le nom est indiqué dans le titre de la boîte de dialogue). Pour lister les ensembles de données d'un autre classeur, activez-le dans Excel et rouvrez la boîte de dialogue du Gestionnaire.

Menu Utilitaires de données

Commande Empiler

Convertit un ensemble de variables du format désempilé au format empilé.

La commande Empiler permet de convertir des données de format « désempilé », où l'ensemble de données comporte au moins deux variables de type **valeur**, au format « empilé », où il comporte une variable de type **catégorie** et une variable valeur. Par exemple, si l'on compare les revenus moyens des ménages de différents quartiers, l'approche désempilée présente une variable (ou colonne) *Revenu* distincte pour chaque *Quartier*. Ces colonnes ne doivent pas être de longueur égale : chaque quartier peut présenter une taille d'échantillon distincte. Sous le format empilé, il y aurait une variable valeur *Revenu* et une variable catégorie *Quartier* indiquant le quartier de chaque ménage.

Cette procédure sert ainsi essentiellement à « empiler » les variables en deux colonnes : une colonne de valeur, *Revenu*, et une colonne de catégorie, *Quartier*. Suivant le type d'analyse, l'ensemble de données empilé peut être plus facile à traiter que sa version désempilée.

Variables empilées et désempilées

Classeur2 - Microsoft Excel

Statistiques de synthèse - Tests de normalité - Contrôle de qualité - Utilitaires -
 Graphiques de synthèse - Série temporelle et prévision - Tests non paramétriques - Aide -
 Inférence statistique - Régression et classification - Analyses - Outils

A36

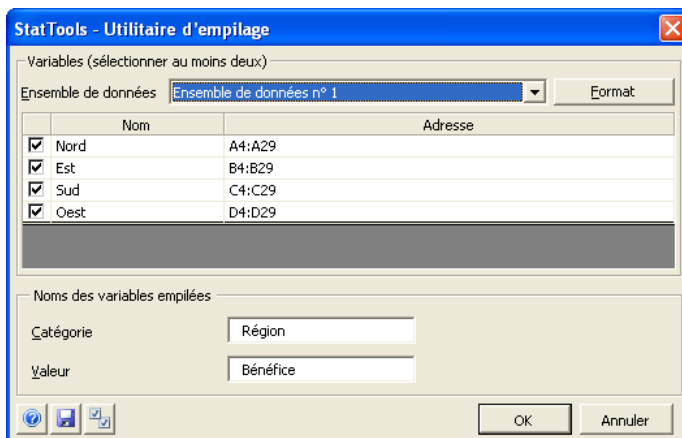
Bénéfice (données empilées):				Bénéfice (données désempilées):	
Nord	Est	Sud	Oest	Région	Bénéfice
46 809 €	43 108 €	73 115 €	28 650 €	Est	43 108 €
45 812 €	4 524 €	35 747 €	5 046 €	Est	4 524 €
49 132 €	55 628 €	129 974 €	2 702 €	Est	55 628 €
15 459 €	46 488 €	34 920 €	175 693 €	Est	46 488 €
39 503 €	80 831 €	113 469 €	126 656 €	Est	80 831 €
9 547 €	58 703 €	238 111 €	129 321 €	Est	58 703 €
49 880 €	45 000 €	55 385 €	107 794 €	Est	45 000 €
41 866 €	65 527 €	104 381 €	89 853 €	Est	65 527 €
3 129 €	58 842 €	88 441 €	84 903 €	Est	58 842 €
23 115 €	12 067 €	139 580 €	72 781 €	Est	12 067 €
47 886 €	26 313 €	176 853 €	25 464 €	Est	26 313 €
22 550 €	38 457 €	234 325 €	125 144 €	Est	38 457 €
15 975 €	8 241 €		82 216 €	Est	8 241 €
716 €	52 476 €		81 951 €	Est	52 476 €
27 065 €	88 298 €		82 222 €	Est	88 298 €
17 569 €	92 347 €		147 837 €	Est	92 347 €
2 020 €	59 968 €		147 412 €	Est	59 968 €
53 883 €	69 574 €		175 167 €	Est	69 574 €
56 984 €	3 055 €		159 937 €	Est	3 055 €
29 041 €	28 375 €		187 547 €	Est	28 375 €
	94 943 €		137 448 €	Est	94 943 €
	54 889 €		115 696 €	Est	54 889 €
	86 583 €		97 260 €	Est	86 583 €
	56 665 €			Est	56 665 €
	28 078 €			Est	28 078 €
	71 380 €			Est	71 380 €
				Oest	28 650 €
				Oest	5 046 €
				Oest	2 702 €
				Oest	175 693 €
				Oest	126 656 €
				Oest	129 321 €
				Oest	107 794 €

Sélectionnez une destination et appuyez sur ENTRÉE ou cliquez sur Coller

100 %

Boîte de dialogue Utilitaire d'empilage

L'empilage s'effectue depuis la boîte de dialogue **Utilitaire d'empilage** :



Au moins deux variables doivent être sélectionnées pour l'empilage. Les données sélectionnées sont toujours initialement traitées comme désempilées. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue Utilitaire d'empilage :

- Le volet **Noms des variables empilées** spécifie les variables de type **catégorie** et **valeur** appelées à former l'ensemble de données empilé à deux variables. Les noms définis ici figureront en haut des colonnes des variables de catégorie et valeur.

Le bouton OK empile les variables et un nouvel ensemble de données se crée pour les données empilées.

Variables « catégorie » et « valeur »

Des variables **Catégorie** et **Valeur** sont nécessaires pour la formation d'un ensemble de données empilé. La variable Catégorie (parfois appelée variable « code ») est tout simplement un identificateur descriptif d'un ensemble correspondant de variable(s) de type valeur. La variable Catégorie est souvent une étiquette textuelle. Les variables de valeur (parfois appelées variables de « mesure ») sont en revanche des variables numériques standard analysables dans les procédures statistiques.

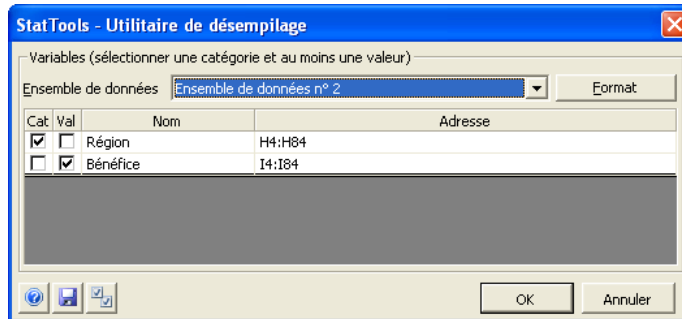
Commande Désempiler

Convertit un ensemble de variables du format empilé au format désempilé.

La commande Désempiler effectue exactement l'inverse de la commande Empiler. Par exemple, si l'on part d'une variable catégorie *Sexe* et d'une variable valeur *Poids*, cette commande les désempile en colonnes *Poids_Hommes* et *Poids_Femmes* séparées.

**Boîte de dialogue
Utilitaire de
désempilage**

Le désempilage s'effectue depuis la boîte de dialogue **Utilitaire de désempilage** :



Les données sélectionnées sont toujours initialement traitées comme empilées. Au moins deux variables doivent être sélectionnées pour le désempilage. L'une de ces variables est identifiée comme la variable **Catégorie** (en cochant la case **Cat**) et au moins une variable est identifiée comme variable **Valeur** (en cochant la case **Val**). Les variables peuvent provenir d'ensembles de données différents.

Le bouton OK désempile les variables et un ou plusieurs nouveaux ensembles de données se créent pour les données désempilées.

Commande Transformer

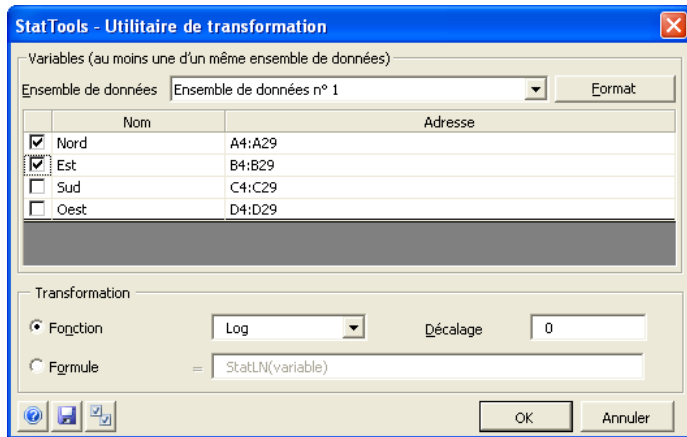
Transforme une ou plusieurs variables en nouvelles variables et valeurs sur la base d'une fonction de transformation définie.

La commande Transformer permet de soumettre une variable à l'une des quatre transformations suivantes : logarithme népérien, carré, racine carrée ou réciproque. Une formule peut aussi être entrée pour calculer la valeur d'une variable transformée.

Si certaines valeurs sont manquantes pour la variable de base, les valeurs correspondantes de la variable transformée le sont aussi.

La transformation des variables s'effectue dans la boîte de dialogue **Utilitaire de transformation** :

**Boîte de dialogue
Utilitaire de
transformation**



L'ensemble de données sélectionné est toujours au format désempilé. Les variables d'un seul ensemble de données à la fois peuvent être transformées.

Dans la boîte de dialogue Utilitaire de transformation :

- **Fonction** indique l'opération mathématique à effectuer sur chaque valeur des variables sélectionnées pour générer la nouvelle valeur transformée. Quatre fonctions de transformation sont proposées : logarithme népérien, carré, racine carrée ou réciproque. L'option **Formule** permet l'entrée d'une formule personnalisée pour le calcul d'une nouvelle valeur de variable basé sur une expression mathématique telle que

$$(Variable*1,5)^2$$

On remarquera que dans l'équation, le mot-clé « Variable » est un paramètre fictif représentant la valeur réelle de la variable à transformer.

Ensemble de données avec nouvelles variables transformées

Bénéfice (données emplies):					
Nord	Est	Sud	Oest	Log(Nord)	Log(Est)
46 809 €	43 108 €	73 115 €	28 650 €	10,75383077	10,67146387
45 812 €	4 524 €	35 747 €	5 046 €	10,73230134	8,417151837
49 132 €	55 628 €	129 974 €	2 702 €	10,80226583	10,92644195
15 459 €	46 488 €	34 920 €	175 693 €	9,645946637	10,74694949
39 503 €	80 831 €	113 469 €	126 656 €	10,5841319	11,30011583
9 547 €	58 703 €	238 111 €	129 321 €	9,163982248	10,98024611
49 880 €	45 000 €	55 385 €	107 794 €	10,8173754	10,71441777
41 866 €	65 527 €	104 381 €	89 853 €	10,64222932	11,09021755
3 129 €	58 842 €	88 441 €	84 903 €	8,048468744	10,98261116
23 115 €	12 067 €	139 580 €	72 781 €	10,04823704	9,398229733
47 886 €	26 313 €	176 853 €	25 464 €	10,77657847	10,17781839
22 550 €	38 457 €	234 325 €	125 144 €	10,02349034	10,55729601
15 975 €	8 241 €		82 216 €	9,678780279	9,016876975
716 €	52 476 €		81 951 €	6,573680167	10,8681112
27 065 €	88 298 €		82 222 €	10,20599666	11,38847274
17 569 €	92 347 €		147 837 €	9,773891264	11,4333085
2 020 €	59 968 €		147 412 €	7,61085279	11,00156637
53 883 €	69 574 €		175 167 €	10,89457031	11,15014621
56 984 €	3 055 €		159 937 €	10,95052581	8,024534872
29 041 €	28 375 €		187 547 €	10,2764639	10,25326375
	94 943 €		137 448 €		11,46103199
	54 889 €		115 696 €		10,91306824
	86 583 €		97 260 €		11,36885877
	56 665 €				10,94491202
	28 078 €				10,24274163
	71 380 €				11,175773

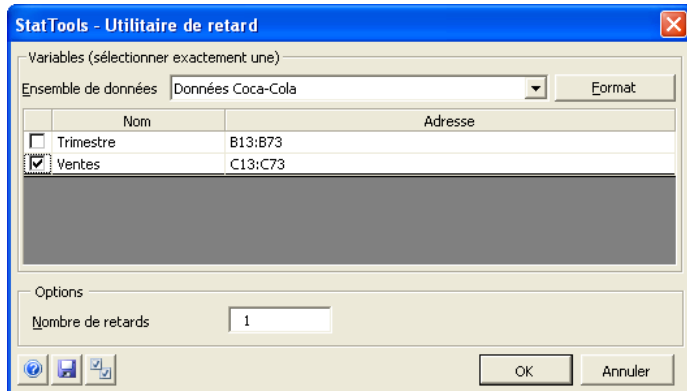
Commande Retard

Crée une nouvelle variable déphasée/retardée sur la base d'une variable existante.

La commande Retard permet de créer une nouvelle variable déphasée ou « retardée » sur la base d'une variable existante. Une variable retardée est tout simplement une version de la variable originale « repoussée » d'un nombre de lignes égal au retard. Par exemple, la version retard 3 des ventes de novembre 1998 représente les ventes de trois mois plus tôt, soit août 1998.

Boîte de dialogue Utilitaire de retard

Les retards de variables se définissent dans la boîte de dialogue **Utilitaire de retard** :



L'ensemble de données sélectionné est toujours au format désempilé. Une seule variable peut être retardée à la fois.

Dans la boîte de dialogue Utilitaire de retard :

- **Nombre de retards** représente le nombre de périodes de retard à appliquer aux valeurs lors de la création des nouvelles variables. Une nouvelle variable se crée pour chacun des retards spécifiés.

**Variable retardée
à droite de
l'ensemble de
données source**

The screenshot shows the Microsoft Excel interface with the StatTools ribbon selected. The ribbon includes options for 'Statistiques de synthèse', 'Tests de normalité', 'Contrôle de qualité', 'Graphiques de synthèse', 'Série temporelle et prévision', 'Tests non paramétriques', 'Inférence statistique', and 'Régression et classification'. The active window is 'Prévision.xls' in 'Mode de compatibilité'. The spreadsheet content is as follows:

Exemple StatTools : Prévision

L'analyse Prévision de StatTools permet de projeter les valeurs futures d'une variable de série temporelle en fonction de ses valeurs passées.

Dans cet exemple, les chiffres de ventes trimestriels de Coca-Cola sont présentés (en millions). L'analyse de prévision permet de projeter les chiffres de ventes à venir.

Exemple adapté de Data Analysis and Decision Making with Microsoft Excel, par S. Christian Albright, Wayne L. Winston et Christopher Zappe. Copyright 2003 by Brooks/Cole Publishing Company.

Trimestre	Ventes	Retard1(Ventes)
T1-86	1734,83	
T2-86	2244,96	1734,83
T3-86	2533,8	2244,96
T4-86	2154,96	2533,8
T1-87	1547,82	2154,96
T2-87	2104,41	1547,82
T3-87	2014,36	2104,41
T4-87	1991,75	2014,36
T1-88	1869,05	1991,75
T2-88	2313,63	1869,05
T3-88	2128,32	2313,63
T4-88	2026,83	2128,32
T1-89	1910,6	2026,83
T2-89	2331,16	1910,6
T3-89	2206,55	2331,16
T4-89	2173,97	2206,55
T1-90	2148,28	2173,97
T2-90	2739,31	2148,28
T3-90	2792,75	2739,31
T4-90	2556,01	2792,75
T1-91	2480,97	2556,01
T2-91	3039,52	2480,97
T3-91	3172,12	3039,52
T4-91	2879	3172,12
T1-92	2772	2879
T2-92	3550	2772
T3-92	3508	3550
T4-92	3243,86	3508
T1-93	3056	3243,86
T2-93	3899	3056

Commande Différence

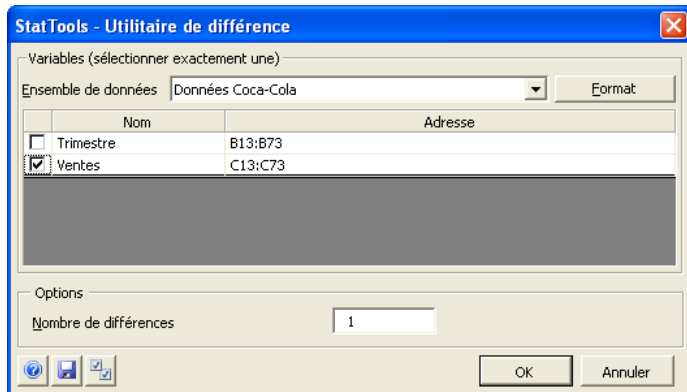
Crée des variables de différence au départ d'une variable originale.

La commande Différence permet de créer un nombre quelconque de variables de différence au départ d'une variable originale. Elle est principalement utile dans le contexte des variables de série temporelle. On sélectionne une variable à différencier et le nombre de différences désiré (généralement 1 ou 2). La procédure crée le nombre de nouvelles variables de différence indiqué. Chaque variable de différence contient les différences de la variable sélectionnée. Dans le cas de données mensuelles, par exemple, la valeur de différence de mars 1997 est la valeur mars 1997 originale moins la valeur février 1997 originale. De même, la seconde variable de différence (si elle est demandée) contient les différences des premières différences.

La procédure est souvent utile à l'analyse de série temporelle, quand la variable originale n'est pas « stationnaire » dans le temps. Par exemple, une série temporelle à tendance à la hausse n'est pas stationnaire. La différence permet souvent d'atteindre l'état stationnaire. La seconde différence est parfois utile, mais elle est moins courante. Au troisième degré et au-delà, la différence n'est presque jamais nécessaire.

Boîte de dialogue Utilitaire de différence

Les variables de différence se définissent dans la boîte de dialogue **Utilitaire de différence** :



L'ensemble de données sélectionné est toujours au format désempilé. Une seule variable peut être utilisée à la fois pour créer des variables de différence.

Dans la boîte de dialogue Utilitaire de différence :

- **Nombre de différences** désigne le nombre de différences à créer.

Ensemble de données avec variable de différence

Exemple StatTools : Prédiction

L'analyse Prédiction de StatTools permet de projeter les valeurs futures d'une variable de série temporelle en fonction de ses valeurs passées. Dans cet exemple, les chiffres de ventes trimestriels de Coca-Cola sont présentés (en millions). L'analyse de prédiction permet de projeter les chiffres de ventes à venir.

Exemple adapté de Data Analysis and Decision Making with Microsoft Excel, par S. Christian Albright, Wayne L. Winston et Christopher Zappe. Copyright 2003 by Brooks/Cole Publishing Company.

Trimestre	Ventes	Différence1(Ventes)
T1-86	1734,83	
T2-86	2244,96	510,13
T3-86	2533,8	288,84
T4-86	2154,96	-378,84
T1-87	1547,82	-607,14
T2-87	2104,41	556,59
T3-87	2014,36	-90,05
T4-87	1991,75	-22,61
T1-88	1869,05	-122,7
T2-88	2313,63	444,58
T3-88	2128,32	-185,31
T4-88	2026,83	-101,49
T1-89	1910,6	-116,23
T2-89	2331,16	420,56
T3-89	2206,55	-124,61
T4-89	2173,97	-32,58
T1-90	2148,28	-25,69
T2-90	2739,31	591,03
T3-90	2792,75	53,44
T4-90	2556,01	-236,74
T1-91	2480,97	-75,04
T2-91	3039,52	558,55
T3-91	3172,12	132,6
T4-91	2879	-293,12
T1-92	2772	-107
T2-92	3550	778
T3-92	3508	-42
T4-92	3243,86	-264,14
T1-93	3056	-187,86
T2-93	3899	843

Commande Interaction

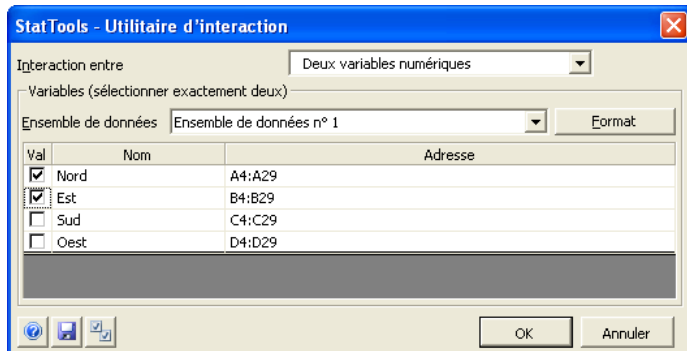
Crée une variable d'interaction au départ d'une ou de plusieurs variables originales.

La commande Interaction permet de créer une variable d'interaction au départ d'une ou de plusieurs variables originales. Les variables d'interaction se créent au départ de deux variables numériques, d'une variable numérique et d'une catégorielle, ou de deux variables catégorielles.

Si les deux variables sont numériques (non catégorielles), l'utilitaire en crée le produit. Si une variable est numérique et l'autre catégorielle, il crée les produits de la variable numérique avec chaque valeur nominale (dummy) correspondant aux catégories de la variable catégorielle. Et si les deux variables sont catégorielles, il crée les produits de toutes les paires de valeurs nominales (dummy) au départ des deux variables.

Boîte de dialogue Utilitaire d'interaction

Les variables d'interaction se définissent dans la boîte de dialogue **Utilitaire d'interaction** :



L'ensemble de données sélectionné est toujours au format désempilé. Une ou plusieurs variables à la fois peuvent servir à la création d'une variable d'interaction.

Dans la boîte de dialogue Utilitaire d'interaction :

- **Interaction entre** permet la sélection du type de chaque variable à sélectionner : **Deux variables numériques**, **Une variable numérique et une variable catégorielle** ou **Deux variables catégorielles**.

**Comment se crée
la variable
d'interaction ?**

Une variable d'interaction se forme au départ des deux variables sélectionnées dans la boîte de dialogue. Trois options sont proposées pour ces deux variables. Il s'agit, pour la première, de deux variables de « mesure » numérique. La variable d'interaction en est alors le produit. Pour la deuxième, une variable peut être une « mesure » numérique et l'autre, une catégorie. StatTools crée dans ce cas des variables nominales (dummy) internes pour chaque catégorie de la variable catégorielle et multiplie *chaque* variable nominale par la variable numérique. Pour la troisième option, les deux variables peuvent être catégorielles. StatTools crée alors des variables nominales (dummy) internes pour chaque catégorie de chaque variable catégorielle et multiplie chaque variable nominale de la première par la chaque variable nominale de la seconde. Ainsi, si les deux variables catégorielles comptent, respectivement, 2 et 5 catégories, StatTools crée $2 \times 5 = 10$ variables d'interaction.

**Ensemble de
données avec
variable
d'interaction**

	A	B	C	D	E	F	G	H	I	
1	Bénéfice (données empilées):									
2										
3	Nord	Est	Sud	Oest	Interaction(Nord,Est)					
4	46 809 €	43 108 €	73 115 €	28 650 €	2017842372					
5	45 812 €	4 524 €	35 747 €	5 046 €	207253488					
6	49 132 €	55 628 €	129 974 €	2 702 €	2733114896					
7	15 459 €	46 488 €	34 920 €	175 693 €	718657992					
8	39 503 €	80 831 €	113 469 €	126 656 €	3193066993					
9	9 547 €	58 703 €	238 111 €	129 321 €	560437541					
10	49 880 €	45 000 €	55 385 €	107 794 €	2244600000					
11	41 866 €	65 527 €	104 381 €	89 853 €	2743353382					
12	3 129 €	58 842 €	88 441 €	84 903 €	184116618					
13	23 115 €	12 067 €	139 580 €	72 781 €	278928705					
14	47 886 €	26 313 €	176 853 €	25 464 €	1260024318					
15	22 550 €	38 457 €	234 325 €	125 144 €	867205350					
16	15 975 €	8 241 €		82 216 €	131649975					
17	716 €	52 476 €		81 951 €	37572816					
18	27 065 €	88 298 €		82 222 €	2389785370					
19	17 569 €	92 347 €		147 837 €	1622444443					
20	2 020 €	59 968 €		147 412 €	121135360					
21	53 883 €	69 574 €		175 167 €	3748855842					
22	56 984 €	3 055 €		159 937 €	174086120					
23	29 041 €			28 375 €	187 547 €					
24				94 943 €	137 448 €					
25				54 889 €	115 696 €					
26				86 583 €	97 260 €					
27				56 665 €						
28				28 078 €						
29				71 380 €						
30										

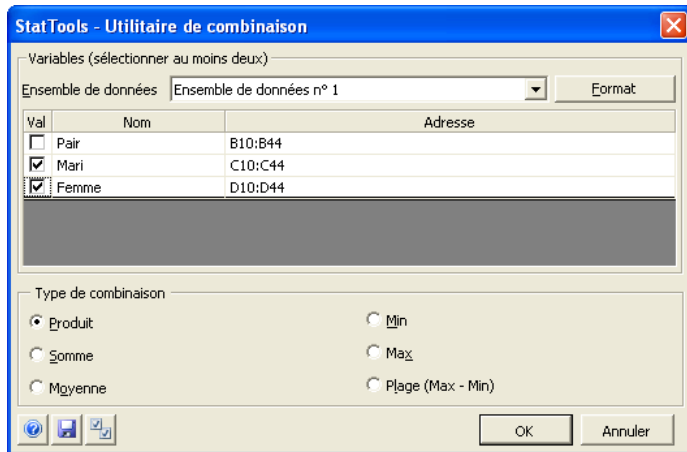
Commande Combinaison

Crée une variable de combinaison au départ d'une ou de plusieurs variables originales.

La commande Combinaison permet de créer une variable de combinaison au départ d'une ou de plusieurs variables originales. La variable de combinaison est issue du produit, de la somme, de la moyenne, du minimum, du maximum ou de l'étendue min-max d'une ou de plusieurs variables.

Boîte de dialogue Utilitaire de combinaison

Les variables de combinaison se définissent dans la boîte de dialogue **Utilitaire de combinaison** :



L'ensemble de données sélectionné est toujours au format désempilé. Une ou plusieurs variables à la fois peuvent servir à la création d'une variable de combinaison.

Dans la boîte de dialogue Utilitaire de combinaison :

- **Type de combinaison** désigne l'opération mathématique à effectuer sur les variables sélectionnées pour la création de la variable d'interaction. Les options proposées sont : produit, somme, moyenne, minimum, maximum ou plage min-max (étendue).

Ensemble de données avec variable de combinaison

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	Pair	Mari	Femme	Produit(Mari,Femme)
10	1	6	3	18
11	2	7	8	56
12	3	8	5	40
13	4	6	4	24
14	5	8	5	40
15	6	7	6	42
16	7	8	5	40
17	8	6	7	42
18	9	7	8	56
19	10	6	5	30
20	11	5	3	15
21	12	8	4	32
22	13	7	5	35
23	14	7	8	56
24	15	7	5	35
25	16	6	6	36
26	17	5	5	25
27	18	6	4	24
28	19	9	5	45
29	20	7	10	70
30	21	9	9	81
31	22	6	6	36
32	23	6	5	30
33	24	6	4	24
34	25	8	5	40
35	26	9	5	45
36	27	7	7	49
37	28	8	5	40

Commande Variable nominale

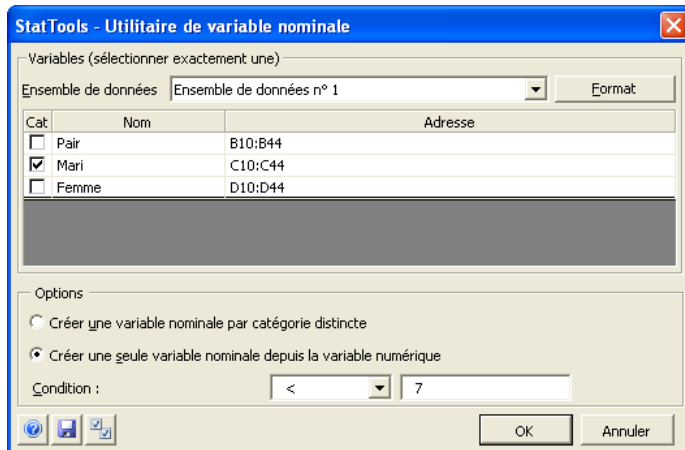
Crée des variables nominales (dummy 0-1) basées sur des variables existantes.

La commande Var. nominale crée des variables nominales (dummy 0-1) basées sur des variables existantes. Deux options sont possibles :

- 1) On crée une variable nominale (dummy) pour chaque catégorie d'une variable catégorielle. Par exemple, pour une variable catégorielle *Origine* (d'automobiles) à catégories *USA*, *Europe* et *Asie*, l'option crée trois variables nominales sous les noms *Origine_USA*, *Origine_Europe* et *Origine_Asie*.
- 2) On crée une seule variable nominale (dummy) au départ d'une variable numérique, en fonction d'une valeur limite sélectionnée. Par exemple, pour une variable *Poids*, on pourrait créer des variables nominales sous la condition $Poids \leq 60$. Dans ce cas, une nouvelle variable serait ajoutée, à valeur 0 quand $Poids > 60$ et 1 quand $Poids \leq 60$.

Boîte de dialogue Utilitaire de variable nominale

Les variables nominales se créent dans la boîte de dialogue **Utilitaire de variable nominale** :



L'ensemble de données sélectionné peut être au format empilé ou déempilé. Une seule variable peut être utilisée à la fois pour créer des variables nominales.

Dans la boîte de dialogue Utilitaire de variable nominale :

- Le mode de création des variables nominales (dummy) se sélectionne sous **Options** : soit 1) une variable nominale se crée pour chaque catégorie d'une variable catégorielle, ou 2) une seule variable nominale se crée depuis une variable numérique. La **condition** spécifie la valeur limite à utiliser lors de l'affectation d'une variable numérique à une variable nominale (dummy) 0-1.

Ensemble de données avec variable nominale

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	Pair	Mari	Femme	Mari < 7
9				
10	1	6	3	1
11	2	7	8	0
12	3	8	5	0
13	4	6	4	1
14	5	8	5	0
15	6	7	6	0
16	7	8	5	0
17	8	6	7	1
18	9	7	8	0
19	10	6	5	1
20	11	5	3	1
21	12	8	4	0
22	13	7	5	0
23	14	7	8	0
24	15	7	5	0
25	16	6	6	1
26	17	5	5	1
27	18	6	4	1
28	19	9	5	0
29	20	7	10	0
30	21	9	9	0
31	22	6	6	1
32	23	6	5	1
33	24	6	4	1
34	25	8	5	0
35	26	9	5	0
36	27	7	7	0
37	28	8	5	0

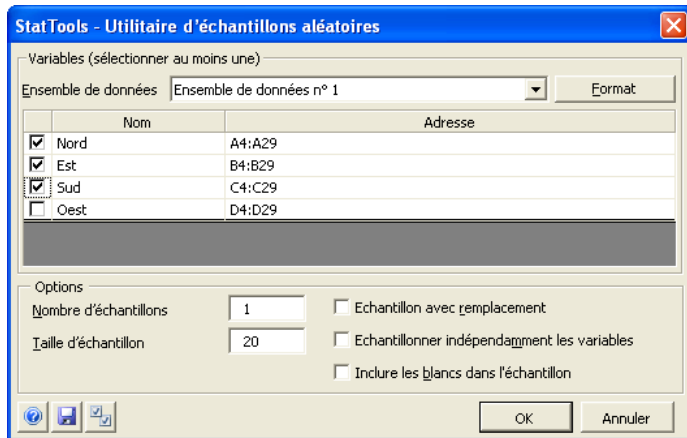
Commande Échantillons aléatoires

Génère des échantillons aléatoires au départ des variables sélectionnées.

La commande Échantillons aléatoires permet de générer un nombre quelconque d'échantillons aléatoires au départ d'une ou de plusieurs variables sélectionnées. On spécifie le nombre d'échantillons et la taille de chacun, et StatTools génère les échantillons au départ des variables sélectionnées. Plusieurs variables peuvent être échantillonnées de manière indépendante ou dépendante. L'échantillonnage peut s'effectuer avec ou sans remplacement.

Boîte de dialogue Utilitaire d'échantillons aléatoires

Les échantillons aléatoires se configurent dans la boîte de dialogue **Utilitaire d'échantillons aléatoires** :



L'ensemble de données sélectionné est toujours au format désempilé. Une ou plusieurs variables à la fois peuvent servir à la génération d'échantillons aléatoires.

Dans la boîte de dialogue Utilitaire d'échantillons aléatoires :

- **Nombre d'échantillons et Taille d'échantillon** : Le nombre d'échantillons indiqué se génère pour chaque variable sélectionnée. Chaque échantillon compte le nombre d'éléments spécifié par la taille d'échantillon.
- **Échantillon avec remplacement** indique qu'une valeur échantillonnée « retourne » à la population originale après l'échantillonnage et peut être à nouveau échantillonnée. Si cette option n'est pas sélectionnée, les valeurs échantillonnées ne peuvent l'être qu'une seule fois.
- **Échantillonner indépendamment les variables multiples** configure un prélèvement indépendant pour chaque valeur échantillonnée pour chaque variable. Si l'option n'est pas sélectionnée, pour chaque valeur échantillonnée, le même indice (nombre compris entre 1 et le nombre de valeurs de la variable) est utilisé pour toutes les variables.

Échantillons aléatoires générés

	A	B	C	D	E	F	G	H	I	J
1	Nord(1)	Est(1)	Sud(1)							
2	15975	8241	92347							
3	28375	56665	28375							
4	92347	28375	92347							
5	27065	88298	69574							
6	53883	69574	94943							
7	69574	54889	69574							
8	45812	4524	35747							
9	46809	43108	73115							
10	56984	3055	92347							
11	17569	92347	3055							
12	716	52476	59968							
13	3129	58842	88441							
14	94943	28078	94943							
15	3055	86583	3055							
16	22550	38457	234325							
17	59968	94943	59968							
18	39503	80831	113469							
19	49132	55628	129974							
20	47886	26313	176853							
21	15459	46488	34920							
22										

Menu Statistiques de synthèse

Les commandes du menu Statistiques de synthèse permettent de calculer différentes mesures de synthèse numérique pour variables simples ou paires de variables. On notera qu'il n'y a pas de tableaux de contingence dans les procédures StatTools. Excel assure déjà cette fonctionnalité dans les tableaux croisés dynamiques.

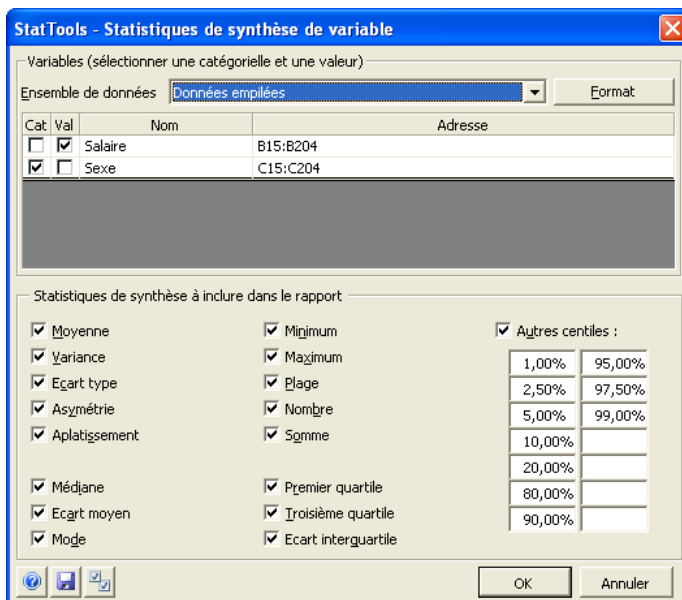
Commande Synthèse de variable

Calcule les statistiques de synthèse des variables.

La commande **Synthèse de variable** présente les données de synthèse des variables numériques sélectionnées. Parmi les statistiques proposées : moyenne, médiane, écart type, variance, minimum, maximum, étendue, premier quartile, troisième quartile, écart interquartile, écart moyen absolu, asymétrie, aplatissement, nombre, somme et centiles sélectionnés.

Boîte de dialogue Statistiques de synthèse de variable

L'analyse se configure dans la boîte de dialogue **Statistiques de synthèse de variable** :



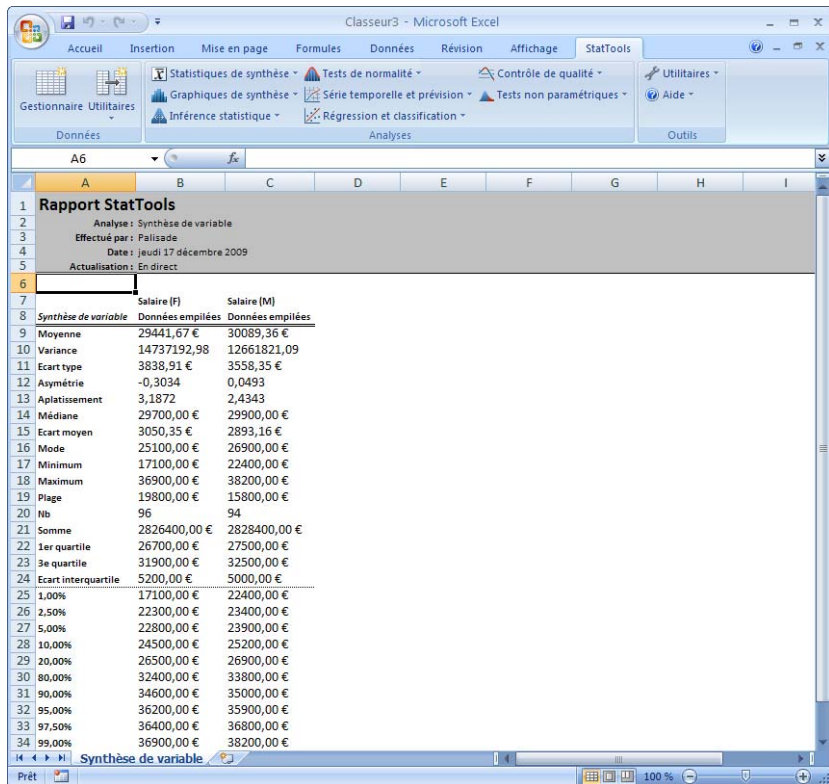
Une ou plusieurs variables peuvent être sélectionnées pour l'analyse. L'ensemble de données sélectionné peut être au format empilé ou désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue Statistiques de synthèse de variable :

- Le volet **Statistiques de synthèse à inclure dans le rapport** permet de sélectionner les statistiques désirées dans le rapport. Pour ajouter des centiles particuliers, il suffit d'en taper les valeurs désirées.

Rapport de synthèse de variable

Le rapport de synthèse de variable utilise les fonctions statistiques de StatTools (telles que **StatSkewness**) pour permettre la liaison automatique aux données. Le rapport se place à l'endroit spécifié sous la commande Paramètres.



The screenshot shows the Microsoft Excel interface with the StatTools ribbon active. The report generated is titled 'Rapport StatTools' and provides a summary of statistical analysis for 'Synthèse de variable'. The report includes the following data:

Synthèse de variable	Données empiquées	Données empiquées
Moyenne	29441,67 €	30089,36 €
Variance	14737192,98	12661821,09
Ecart type	3838,91 €	3558,35 €
Asymétrie	-0,3034	0,0493
Aplatissement	3,1872	2,4343
Médiane	29700,00 €	29900,00 €
Ecart moyen	3050,35 €	2893,16 €
Mode	25100,00 €	26900,00 €
Minimum	17100,00 €	22400,00 €
Maximum	36900,00 €	38200,00 €
Plage	19800,00 €	15800,00 €
Nb	96	94
Somme	2826400,00 €	2828400,00 €
1er quartile	26700,00 €	27500,00 €
3e quartile	31900,00 €	32500,00 €
Ecart interquartile	5200,00 €	5000,00 €
1,00%	17100,00 €	22400,00 €
2,50%	22300,00 €	23400,00 €
5,00%	22800,00 €	23900,00 €
10,00%	24500,00 €	25200,00 €
20,00%	26500,00 €	26900,00 €
80,00%	32400,00 €	33800,00 €
90,00%	34600,00 €	35000,00 €
95,00%	36200,00 €	35900,00 €
97,50%	36400,00 €	36800,00 €
99,00%	36900,00 €	38200,00 €

**Données
manquantes et
liaison aux
données**

- **Données manquantes** – Cette procédure admet les données manquantes au cas par cas. Autrement dit, pour chaque variable, les données manquantes de cette variable sont omises lors du calcul des mesures de synthèse. (Il s'agit d'ailleurs là de la méthode Excel par défaut : si la fonction MOYENNE est appliquée à une plage, par exemple, elle ne s'applique qu'aux valeurs *numériques* de la plage.)
- **Liaison aux données** – Toutes les mesures de synthèse se calculent au moyen de formules liées aux données. Si les données changent, les mesures de synthèse changent donc automatiquement.

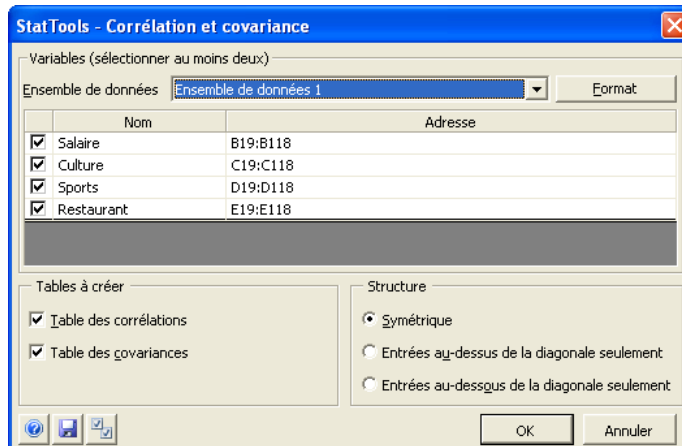
Commande Corrélation et covariance

Produit une table des corrélations et/ou des covariances entre les variables.

La commande **Corrélation et covariance** produit une table des corrélations et/ou des covariances entre un ensemble de variables numériques sélectionnées. Comme ces deux tables sont symétriques (la corrélation entre X et Y est identique à celle entre Y et X), on peut choisir d'afficher (1) seulement les corrélations (ou covariances) *au-dessous* de la diagonale, (2) seulement celles *au-dessus* de la diagonale ou (3) *les deux*.

Boîte de dialogue Corrélation et covariance

Cette analyse se configure dans la boîte de dialogue **Corrélation et covariance** :



Au moins deux variables doivent être sélectionnées pour l'analyse. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue Corrélation et covariance :

- Le volet **Tables à créer** désigne la ou les tables désirées : corrélations et/ou covariances.
- Le volet **Structure** spécifie la structure de la ou des tables à générer :
 - **Symétrique** affiche les corrélations (ou covariances) au-dessus *et* au-dessous de la diagonale.
 - **Entrées au-dessus de la diagonale seulement** affiche les corrélations (ou covariances) *au-dessus* de la diagonale.
 - **Entrées au-dessous de la diagonale seulement** affiche les corrélations (ou covariances) *au-dessous* de la diagonale.

Rapport de corrélation et covariance

Le rapport de corrélation et covariance utilise les fonctions statistiques de StatTools (telles que **StatCorrelationCoeff**) pour permettre la liaison automatique aux données. Le rapport se place à l'endroit spécifié sous la commande Paramètres.

Rapport StatTools				
Analyse : Corrélation et covariance				
Effectué par : Palisade				
Date : jeudi 17 décembre 2009				
Actualisation : En direct				
Table de corrélation				
	Salaire	Culture	Sports	Restaurant
	Ensemble de données 1	Ensemble de données 1	Ensemble de données 1	Ensemble de données 1
Salaire	1,000	0,506	-0,081	0,558
Culture	0,506	1,000	-0,520	0,170
Sports	-0,081	-0,520	1,000	0,266
Restaurant	0,558	0,170	0,266	1,000
Table de covariance				
	Salaire	Culture	Sports	Restaurant
	Ensemble de données 1	Ensemble de données 1	Ensemble de données 1	Ensemble de données 1
Salaire	91130278,79	1105845,25	-221238,79	2590600,81
Culture	1105845,25	52315,39	-33946,99	18938,02
Sports	-221238,79	-33946,99	81427,23	36829,58
Restaurant	2590600,81	18938,02	36829,58	236187,67

**Données
manquantes et
liaison aux
données**

- **Données manquantes** – Les données manquantes sont admises. Elles sont traitées à la paire : pour obtenir la corrélation (ou covariance) entre une paire quelconque de variables, tous les cas présentant des données manquantes au niveau de *l'une ou l'autre* des deux variables sont omis.
- **Liaison aux données** – Toutes les corrélations et covariances se calculent au moyen de formules liées aux données. Si les données changent, ces mesures de synthèse s'actualisent donc automatiquement.

Menu Graphiques de synthèse

Les commandes du menu Graphiques de synthèse servent à la création de graphiques fort utiles à l'analyse statistique mais tout aussi difficiles et parfois même impossibles à produire à l'aide de l'assistant graphique Excel. Étant donné les vastes capacités graphiques déjà proposées dans Excel, StatTools évite d'en dupliquer les qualités.

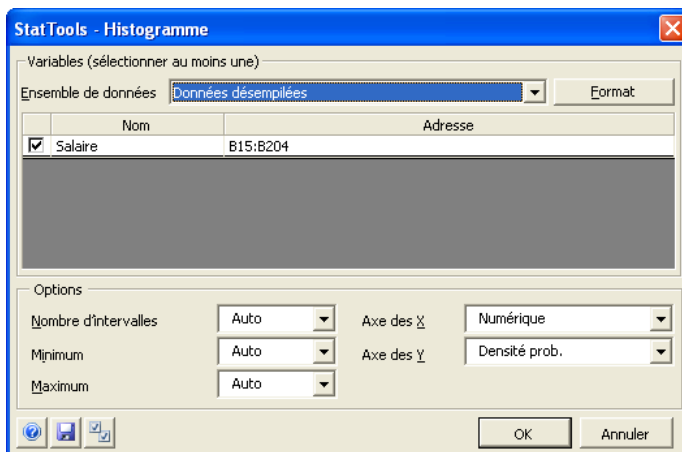
Commande Histogramme

Crée l'histogramme des variables.

La commande Histogramme crée un histogramme pour chaque variable sélectionnée. Elle offre l'option de définir les catégories ou « intervalles » de l'histogramme, qu'elle indique ensuite clairement dans le graphique. La commande produit aussi la table de fréquence à la base de chaque histogramme.

Boîte de dialogue Histogramme

Ce type de graphique se configure dans la boîte de dialogue **Histogramme** :

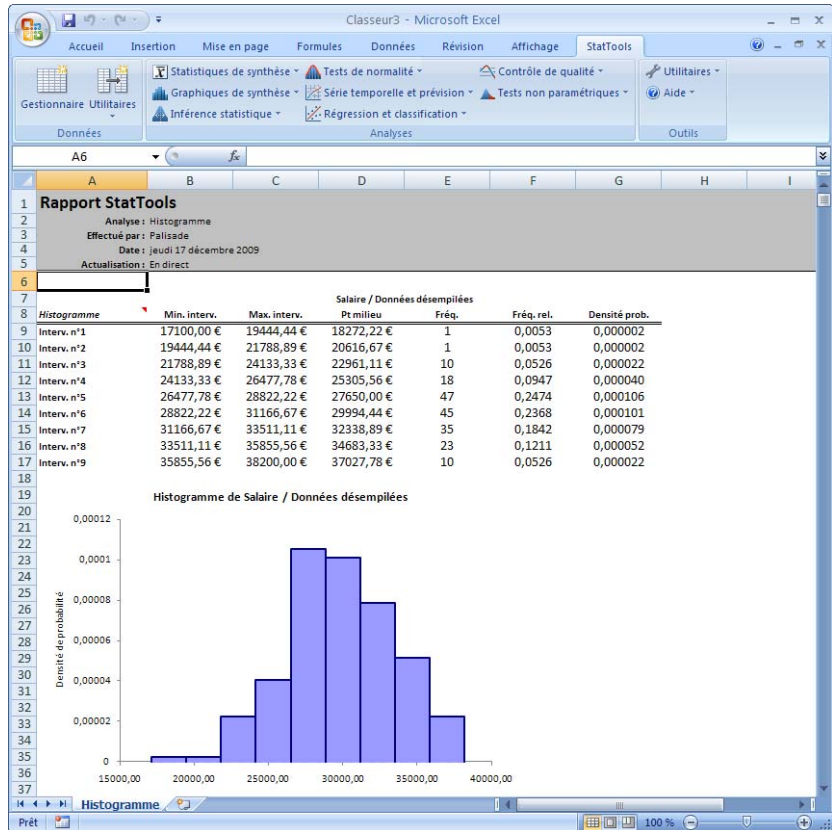


Une ou plusieurs variables peuvent être sélectionnées pour la représentation graphique. L'ensemble de données sélectionné peut être au format empilé ou désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Histogramme** :

- **Nombre d'intervalles** définit le nombre d'intervalles d'histogramme calculés sur l'étendue d'un graphique. La valeur entrée doit être comprise entre 1 et 200. Sous l'option **Auto**, le nombre optimal d'intervalles est calculé par méthode heuristique interne.
- **Minimum** définit la valeur minimum appelée à marquer le point de départ de l'histogramme. Sous **Auto**, StatTools se réfère à la valeur minimum des données représentées graphiquement.
- **Maximum** définit la valeur maximum appelée à marquer la fin de l'histogramme. Sous **Auto**, StatTools se réfère à la valeur maximum des données représentées graphiquement.
- **Axe des X** désigne un axe catégoriel ou numérique. Sous l'option **Catégorie**, l'axe étiquette simplement chaque intervalle de son point milieu. Sous l'option **Numérique**, l'axe présente un minimum et un maximum « lisibles » et les options d'échelle Excel standard sont applicables.
- **Axe des Y** désigne **Fréquence**, **Fréq. relative** ou **Densité prob.** comme unité de mesure de l'axe des Y. La fréquence est le nombre effectif d'observations dans l'intervalle. La fréquence relative représente la probabilité d'une valeur comprise dans la plage d'un intervalle (observations d'intervalle/observations totales). La densité représente la valeur de fréquence relative divisée par la largeur de l'intervalle, assurant la constance des valeurs de l'axe Y tandis que le nombre d'intervalles varie.

Histogramme de variable unique



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Les histogrammes sont partiellement liés aux données. Quand les données changent et que les nouvelles valeurs tombent dans la plage des intervalles de l’histogramme original, le graphique s’actualise automatiquement. Les changements de données qui exigent une révision des intervalles ne s’accompagnent pas d’une actualisation automatique du graphique.

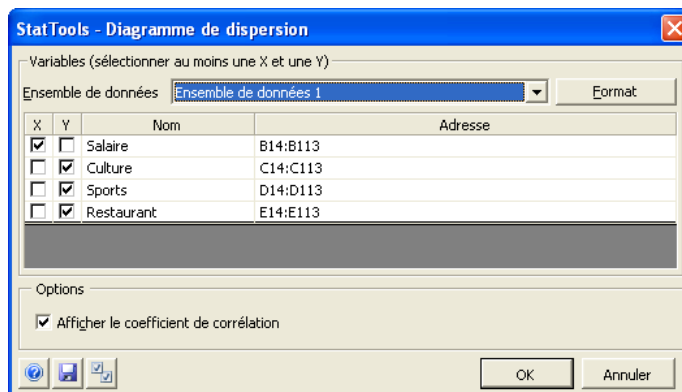
Commande Diagramme de dispersion

Crée des diagrammes de dispersion entre paires de variables.

La commande Diagramme de dispersion crée un diagramme de dispersion pour chaque paire de variables sélectionnée. L'option de tracé XY Excel permet bien la création de diagrammes de dispersion, mais le tableur place automatiquement la première variable (celle à la plus à gauche) sur l'axe horizontal. Plus souple, StatTools permet de choisir la variable à disposer sur l'axe horizontal. Chaque diagramme indique la corrélation entre les deux variables sur le tracé correspondant.

Boîte de dialogue Diagramme de dispersion

Ce type de graphique se configure dans la boîte de dialogue **Diagramme de dispersion** :

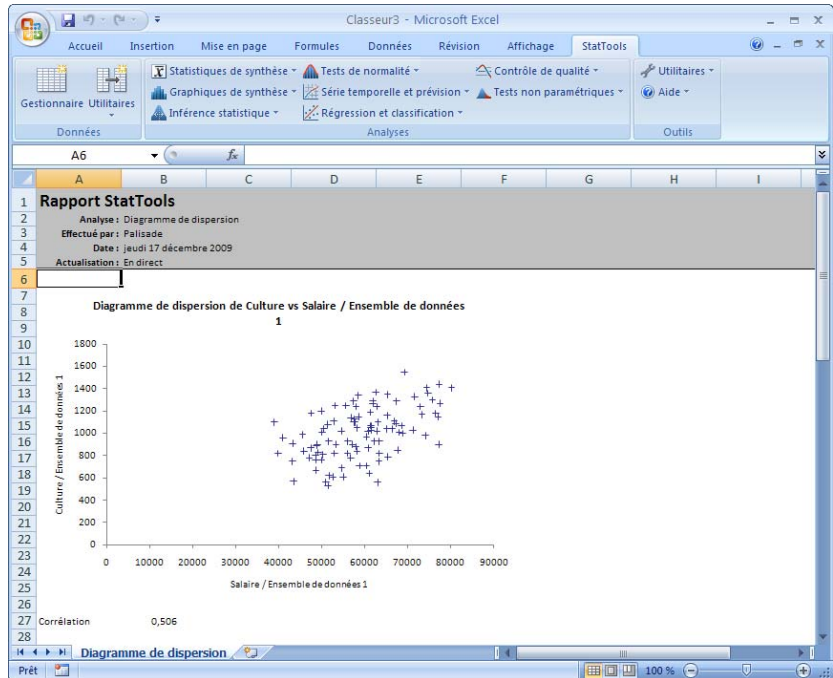


Deux variables ou plus peuvent être sélectionnées pour la représentation graphique. Au moins une variable doit être sélectionnée pour chaque axe, X et Y. Si plus de deux variables sont sélectionnées, plusieurs diagrammes de dispersion se tracent. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Diagramme de dispersion** :

- L'option **Afficher le coefficient de corrélation** spécifie que le coefficient de corrélation entre les variables représentées doit être affiché.

Exemple de diagramme de dispersion



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour l'une ou l'autre des deux variables sélectionnées d'une paire sont omises.
- **Liaison aux données** – Les diagrammes de dispersion sont liés aux données originales. Si les données changent, les diagrammes changent aussi. L'actualisation manuelle de l'échelle des axes peut cependant être nécessaire en cas de changement significatif des plages des variables sélectionnées.

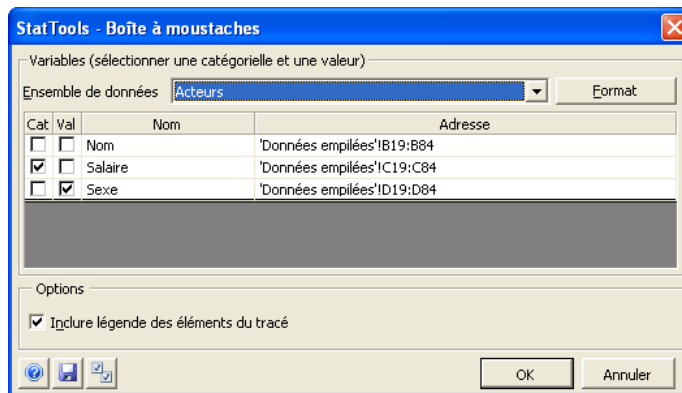
Commande Boîte à moustaches

Crée des diagrammes de type « boîte à moustaches » pour les variables.

La commande Boîte à moustaches crée une simple boîte à moustaches (si une seule variable est sélectionnée) ou plusieurs boîtes l'une à côté de l'autre (si plusieurs variables sont sélectionnées). Elle produit aussi une feuille des statistiques de synthèse (quartiles, écart interquartile, etc.) utilisées pour tracer la ou les boîtes à moustaches.

**Boîte de dialogue
Boîte à
moustaches**

Ce type de graphique se configure dans la boîte de dialogue **Boîte à moustaches** :

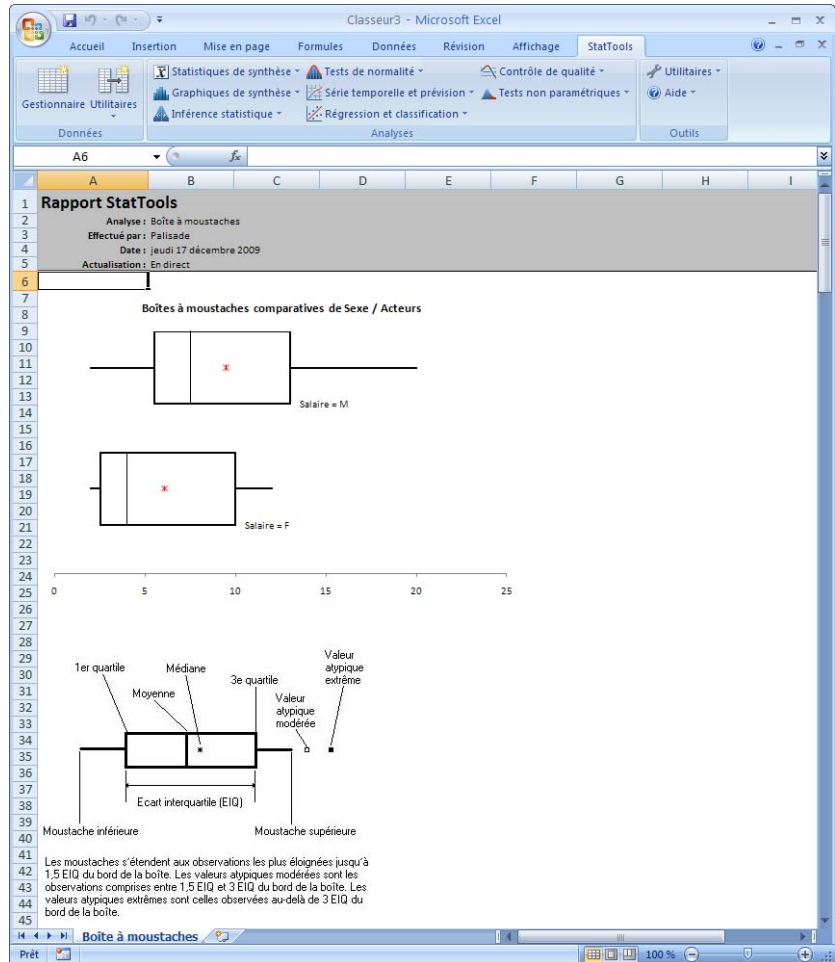


Une ou plusieurs variables peuvent être sélectionnées pour la représentation graphique. L'ensemble de données sélectionné peut être au format empilé ou déempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Boîte à moustaches** :

- L'option **Inclure [la] légende des éléments du tracé** spécifie l'affichage d'une légende décrivant les éléments du graphique sous le tracé.

Exemple de boîte à moustaches



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour l'une quelconque des variables sélectionnées sont omises.
- **Liaison aux données** – Les tracés de boîte à moustaches sont liés aux données originales : si ces données changent, les diagrammes changent aussi. L'actualisation manuelle de l'échelle de l'axe horizontal peut cependant être nécessaire en cas de changement significatif de l'échelle des données.

Menu Inférence statistique

Les commandes du menu Inférence statistique exécutent les analyses d'inférence statistique les plus courantes : intervalles de confiance, tests d'hypothèse et analyse de variance (ANOVA) simple et double.

Commande Intervalle de confiance – Moyenne/Écart type

Calcule les intervalles de confiance pour la moyenne et l'écart type des variables.

La commande Intervalle de confiance – Moyenne/Écart type calcule un intervalle de confiance pour la moyenne et l'écart type de variables simples, ou la différence entre les moyennes pour les paires de variables. Les intervalles de confiance se calculent par analyse à **un échantillon**, à **deux échantillons** ou à **échantillons appariés**.

Cette analyse se configure dans la boîte de dialogue **Intervalle de confiance pour moyenne/écart type** :

**Boîte de dialogue
Intervalle de
confiance pour
moyenne/écart
type**

Cat	Val	Nom	Adresse
<input type="checkbox"/>	<input type="checkbox"/>	Nom	B18:B83
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Sexe	C18:C83
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Salaire	D18:D83

Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à un échantillon requiert au moins une variable ; l'analyse à deux échantillons ou à échantillons appariés en requiert deux. L'ensemble de données sélectionné peut être au format empilé ou déempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Intervalle de confiance** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - **Un échantillon.** Ce type d'analyse calcule les intervalles de confiance pour une simple variable numérique.
 - **Deux échantillons.** Ce type d'analyse calcule l'intervalle de confiance pour la différence entre les moyennes de deux populations indépendantes.
 - **Échantillons appariés.** Ce type d'analyse est fondamentalement identique à l'analyse à deux échantillons. Il convient lorsque les deux variables sont naturellement appariées. Il s'agit essentiellement d'une analyse à un échantillon sur la différence entre les paires.
- Dans le volet **Intervalles de confiance à calculer**, on spécifie les intervalles de confiance à calculer sur les variables sélectionnées. Les options proposées varient suivant le type d'analyse sélectionné :
 - Pour le type d'analyse **Un échantillon**, on sélectionne le calcul sur la moyenne et/ou sur l'écart type et on précise le niveau de confiance (0 à 100 %) pour chacun.
 - Pour les types **Deux échantillons** et **Échantillons appariés**, on sélectionne le calcul pour la différence entre les moyennes de deux variables et on précise le niveau de confiance (0 à 100 %) désiré.

Rapport d'intervalle de confiance

Analyse : Intervalle de confiance		
Effectué par : Palisade		
Date : jeudi 17 décembre 2009		
Actualisation : En direct		
Synthèse des échantillons		
	Salaire (F)	Salaire (M)
	Ensemble de données 1	Ensemble de données 1
Taille	18	48
Moyenne	6,056	9,479
Ecart type	3,737	5,509
Intervalles de conf. (différence des moyennes)		
	Egale	Inégale
	Variances	Variances
Niveau de confiance	95,0%	95,0%
Différence des moyennes	-3,424	-3,424
Erreur type de la différence	1,409296073	1,18671913
Degrés de liberté	64	45
Limite inférieure	-6,239003668	-5,813786132
Limite supérieure	-0,608218554	-1,03343609
Test d'égalité des variances		
Rapport des variances	0,4602	
Valeur p	0,0827	

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Commande Intervalle de confiance – Proportion

Calcule les intervalles de confiance de proportions.

La commande Intervalle de confiance – Proportion permet d'analyser la proportion d'éléments d'un échantillon qui appartient à une catégorie donnée (analyse à **un échantillon**) ou de comparer deux échantillons concernant la proportion d'éléments compris dans une catégorie donnée (analyse à **deux échantillons**). Cette procédure gère trois types de données : **Échantillon de population**, **Table de synthèse avec nombres** et **Table de synthèse avec proportions**.

**Boîte de dialogue
Intervalle de
confiance pour
proportion**

Cette analyse se configure dans la boîte de dialogue **Intervalle de confiance pour proportion** :

StatTools - Intervalle de confiance pour proportion

Type d'analyse: Un échantillon

Type de données: Table de synthèse avec nombres

Sélectionner une variable à catégories et au moins une variable à nombres

Ensemble de données: Nombres simples

Cat	Nb	Nom	Adresse
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Catégorie	B18:B20
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Nombre	C18:C20

Catégories à analyser (sélectionner au moins une)

- Autre
- Démocrate
- Républicain

3 catégories dans la variable sélectionnée.

Options

Niveau de confiance: 95%

Premier échantillon: 150

Deuxième échantillon: 150

OK Annuler

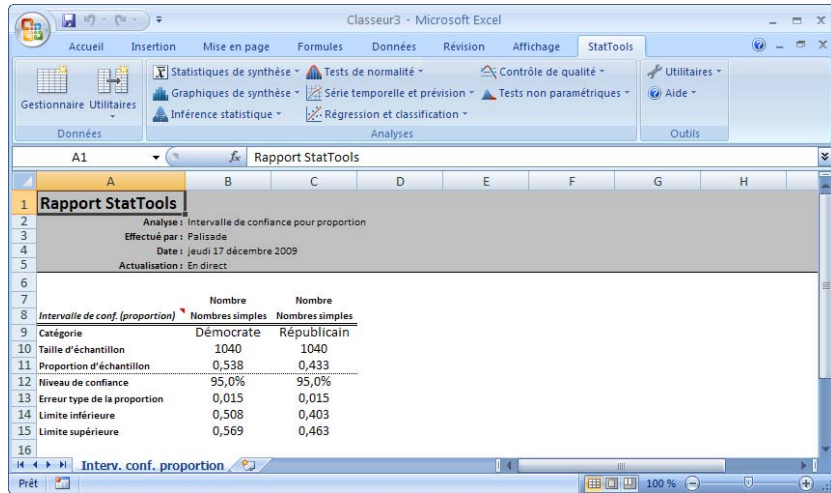
Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à **un échantillon** requiert un ou plusieurs échantillons ; l'analyse à **deux échantillons** en requiert deux. On sélectionne les variables dans la colonne **Nb** (Nombre), % (Proportion) ou dans la colonne sans étiquette pour le type de données **Échantillon de population**. (Si les données d'échantillon de population sont empilées, les échantillons se sélectionnent dans les colonnes **C1** et **C2**. **C1** contient les catégories empilées.) Si les données se présentent sous forme de table avec nombres ou proportions, la colonne **Cat** permet la sélection d'une variable à noms de catégorie.

Sous le type de données **Échantillon de population**, les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Intervalle de confiance** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - Pour le type d'analyse **Un échantillon**, on calcule les intervalles de confiance pour la proportion d'éléments d'un échantillon qui appartiennent à une catégorie donnée.
 - Pour le type **Deux échantillons**, on calcule l'intervalle de confiance de deux échantillons quant à la proportion d'éléments appartenant à une catégorie donnée.
- **Type de données** spécifie le type de données à analyser : **Échantillon de population**, **Table de synthèse avec nombres** ou **Table de synthèse avec proportions**.
- Au bas de la fenêtre, les **options** proposées varient en fonction du Type d'analyse et du Type de données :
 - **Niveau de confiance** indique le niveau de confiance (0 à 100 %) désiré pour l'analyse.
 - **Premier échantillon** et **Deuxième échantillon** déterminent, pour le **type de données Table de synthèse avec proportions**, la taille du premier et du deuxième échantillons (analyse à deux échantillons seulement).

Rapport d'intervalle de confiance



Classeur3 - Microsoft Excel

Statistiques de synthèse Tests de normalité Contrôle de qualité
Graphiques de synthèse Série temporelle et prévision Tests non paramétriques
Inférence statistique Régression et classification

Rapport StatTools

1 **Rapport StatTools**

2 Analyse : Intervalle de confiance pour proportion

3 Effectué par : Palisade

4 Date : jeudi 17 décembre 2009

5 Actualisation : En direct

6

7

Intervalle de conf. (proportion)	Nombre	
	Nombres simples	Nombres simples
Catégorie	Démocrate	Républicain
Taille d'échantillon	1040	1040
Proportion d'échantillon	0,538	0,433
Niveau de confiance	95,0%	95,0%
Erreur type de la proportion	0,015	0,015
Limite inférieure	0,508	0,403
Limite supérieure	0,569	0,463

Interv. conf. proportion

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Commande Test d'hypothèse – Moyenne / Écart type

Effectue un test d'hypothèse pour la moyenne et l'écart type de variables.

La commande Test d'hypothèse – Moyenne/Écart type effectue des tests d'hypothèse pour la moyenne et l'écart type de variables simples, ou la différence entre les moyennes de paires de variables. Les tests s'effectuent par analyse à **un échantillon**, à **deux échantillons** ou à **échantillons appariés**.

**Boîte de dialogue
Test d'hypothèse
pour moyenne/
écart type**

Cette analyse se configure dans la boîte de dialogue **Test d'hypothèse pour moyenne/écart type** :

Nom	Adresse
<input type="checkbox"/> Pièce	B18:B107
<input checked="" type="checkbox"/> Largeur	C18:C107

Tests d'hypothèse à réaliser

Moyenne
Valeur d'hypothèse nulle: 10
Type d'hypothèse rivale: Non égale à la valeur nulle (test bilatéral)

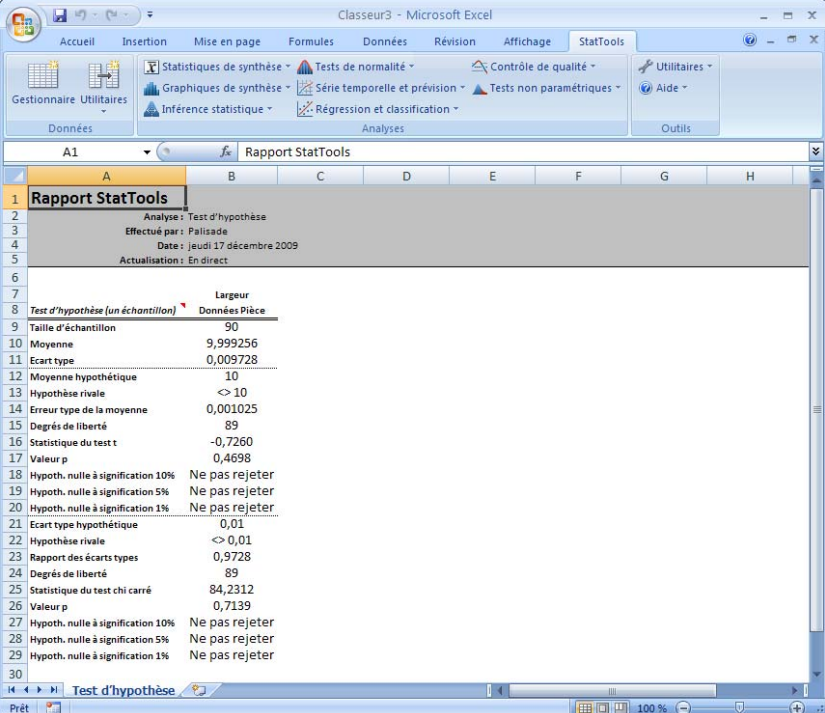
Ecart type
Valeur d'hypothèse nulle: 0,01
Type d'hypothèse rivale: Non égale à la valeur nulle (test bilatéral)

Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à un échantillon requiert au moins une variable ; l'analyse à deux échantillons ou à échantillons appariés en requiert deux. L'ensemble de données sélectionné peut être au format empilé ou désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Test d'hypothèse** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - Le type **Un échantillon** effectue les tests d'hypothèse pour une simple variable numérique.
 - **Deux échantillons** effectue les tests d'hypothèse pour la différence entre les moyennes de deux populations indépendantes.
 - Le type **Échantillons appariés** est fondamentalement identique à l'analyse à deux échantillons. Il convient quand les deux variables sont naturellement appariées. Il s'agit essentiellement d'une analyse à un échantillon sur la différence entre les paires.
- Dans le volet **Tests d'hypothèse à réaliser**, on spécifie les tests à effectuer sur les variables sélectionnées. Les options proposées varient suivant le type d'analyse sélectionné : Pour le type d'analyse à **un échantillon**, on sélectionne la réalisation des tests sur la moyenne et/ou sur l'écart type. Pour les types à **deux échantillons** et à **échantillons appariés**, on les sélectionne pour la différence entre les moyennes de deux variables. Pour chaque type de test sélectionné :
 - **Valeur d'hypothèse nulle** désigne la valeur du paramètre de population sous hypothèse nulle.
 - **Type d'hypothèse rivale** désigne la rivale de la valeur d'hypothèse nulle à évaluer pendant l'analyse. Le type d'hypothèse rivale peut être « **unilatéral** » (valeur supérieure ou inférieure à l'hypothèse nulle) ou « **bilatéral** » (valeur non égale à l'hypothèse nulle).

Rapport de tests d'hypothèse



Classeur3 - Microsoft Excel

Accueil Insertion Mise en page Formules Données Révision Affichage StatTools

Gestionnaire Utilitaires

Statistiques de synthèse Tests de normalité Contrôle de qualité

Graphiques de synthèse Série temporelle et prévision Tests non paramétriques

Inférence statistique Régression et classification

Analyses

Utilitaires Aide Outils

A1 Rapport StatTools

Analyse : Test d'hypothèse	
Effectué par : Palisade	
Date : jeudi 17 décembre 2009	
Actualisation : En direct	
Largeur	
Test d'hypothèse (un échantillon) Données Pièce	
Taille d'échantillon	90
Moyenne	9,999256
Ecart type	0,009728
Moyenne hypothétique	10
Hypothèse rivale	$< > 10$
Erreur type de la moyenne	0,001025
Degrés de liberté	89
Statistique du test t	-0,7260
Valeur p	0,4698
Hypoth. nulle à signification 10%	Ne pas rejeter
Hypoth. nulle à signification 5%	Ne pas rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter
Ecart type hypothétique	0,01
Hypothèse rivale	$< > 0,01$
Rapport des écarts types	0,9728
Degrés de liberté	89
Statistique du test chi carré	84,2312
Valeur p	0,7139
Hypoth. nulle à signification 10%	Ne pas rejeter
Hypoth. nulle à signification 5%	Ne pas rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter

Prêt

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Commande Test d'hypothèse – Proportion

Effectue un test d'hypothèse relatif à des proportions.

La commande Test d'hypothèse – Proportion analyse la proportion d'éléments d'un échantillon qui appartiennent à une catégorie donnée (**analyse à un échantillon**) ou compare deux échantillons concernant la proportion d'éléments compris dans une catégorie donnée (**analyse à deux échantillons**). Cette procédure gère trois types de données : **Échantillon de population**, **Table de synthèse avec nombres** et **Table de synthèse avec proportions**.

**Boîte de dialogue
Test d'hypothèse
pour proportion**

Cette analyse se configure dans la boîte de dialogue **Test d'hypothèse pour proportion** :

StatTools - Test d'hypothèse pour proportion

Type d'analyse: Un échantillon

Type de données: Table de synthèse avec nombres

Sélectionner une variable à catégories et au moins une variable à nombres

Ensemble de données: Tableau avec nombres [Format]

Cat	Nb	Nom	Adresse
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Catégorie	B18:B19
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Nombre	C18:C19

Catégories à analyser (sélectionner au moins une)

Insatisfaits

Satisfaits

2 catégories dans la variable sélectionnée.

Hypothèses sur la proportion

Valeur d'hypothèse nulle: 0,075

Type d'hypothèse rivale: Inférieure à la valeur nulle (test unilatéral)

Options

Premier échantillon: 400

Deuxième échantillon: 100

[OK] [Annuler]

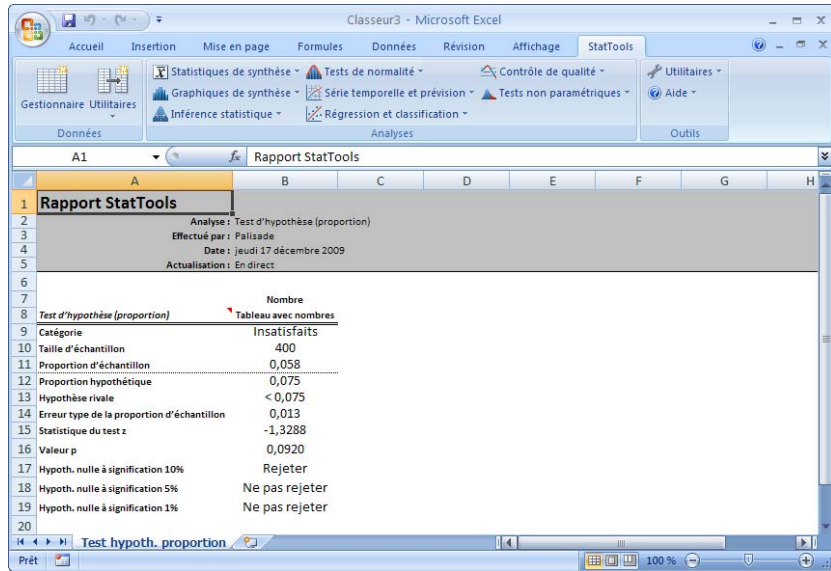
Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à **un échantillon** requiert un ou plusieurs échantillons ; l'analyse à **deux échantillons** en requiert deux. On sélectionne les variables dans la colonne **Nb** (Nombre), % (Proportion) ou dans la colonne sans étiquette pour le type de données **Échantillon de population**. (Si les données d'échantillon de population sont empilées, les échantillons se sélectionnent dans les colonnes **C1** et **C2**. **C1** contient les catégories empilées.) Si les données se présentent sous forme de table avec nombres ou proportions, la colonne **Cat** permet la sélection d'une variable à noms de catégorie.

Sous le type de données **Échantillon de population**, les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Test d'hypothèse pour proportion** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - Pour le type **Un échantillon**, le test d'hypothèse s'effectue pour la proportion d'éléments d'un échantillon qui appartiennent à une catégorie donnée.
 - Pour le type **Deux échantillons**, le test d'hypothèse s'effectue pour deux échantillons quant à la proportion d'éléments appartenant à une catégorie donnée.
- **Type de données** spécifie le type de données à analyser : **Échantillon de population**, **Table de synthèse avec nombres** ou **Table de synthèse avec proportions**.
- Dans le volet **Tests d'hypothèse à réaliser**, on spécifie les tests à effectuer sur la proportion sélectionnée. Les options proposées varient suivant le type d'analyse sélectionné :
 - **Valeur d'hypothèse nulle** désigne la valeur du paramètre de population sous hypothèse nulle.
 - **Type d'hypothèse rivale** désigne la rivale de la valeur d'hypothèse nulle à évaluer pendant l'analyse. Le type d'hypothèse rivale peut être « **unilatéral** » (valeur supérieure ou inférieure à l'hypothèse nulle) ou « **bilatéral** » (valeur non égale à l'hypothèse nulle).
- **Premier échantillon** et **Deuxième échantillon** déterminent, pour le **type de données Table de synthèse avec proportions**, la taille du premier et du deuxième échantillons (analyse à deux échantillons seulement).

Rapport de tests d'hypothèse



The screenshot shows the Microsoft Excel interface with the StatTools ribbon active. The report is titled 'Rapport StatTools' and provides the following details:

- Analyse: Test d'hypothèse (proportion)
- Effectué par: Palisade
- Date: jeudi 17 décembre 2009
- Actualisation: En direct

Test d'hypothèse (proportion)	Nombre
Catégorie	Insatisfaits
Taille d'échantillon	400
Proportion d'échantillon	0,058
Proportion hypothétique	0,075
Hypothèse rivale	< 0,075
Erreur type de la proportion d'échantillon	0,013
Statistique du test z	-1,3288
Valeur p	0,0920
Hypoth. nulle à signification 10%	Rejeter
Hypoth. nulle à signification 5%	Ne pas rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

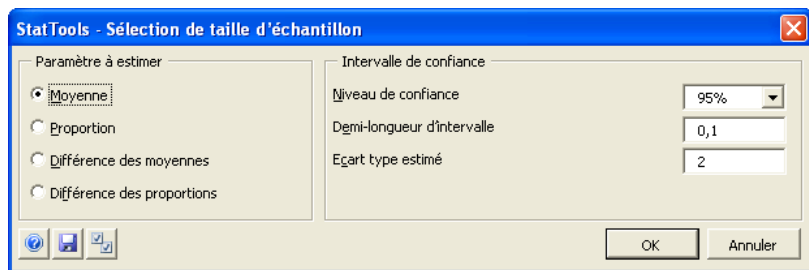
Commande Sélection de taille d'échantillon

Détermine la taille d'échantillon nécessaire au calcul des intervalles de confiance.

La commande Sélection de taille d'échantillon détermine la ou les tailles d'échantillon nécessaires à l'obtention d'un intervalle de confiance à demi-longueur prescrite. La procédure s'effectue pour les intervalles de confiance d'une moyenne, d'une proportion, de la différence entre deux moyennes et de la différence entre deux proportions. Aucun ensemble de données et aucune variable n'est nécessaire : la taille d'échantillon est une information généralement requise *avant* la collecte des données. Pour la détermination de la taille d'un échantillon, on spécifie, entre autres paramètres éventuellement nécessaires, le niveau de confiance et la demi-longueur désirée.

**Boîte de dialogue
Sélection de taille
d'échantillon**

Cette analyse se configure dans la boîte de dialogue **Sélection de taille d'échantillon** :

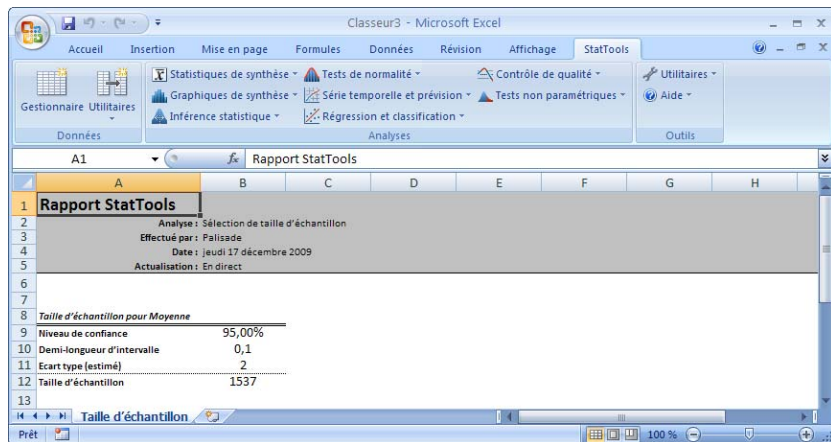


Dans la boîte de dialogue **Sélection de taille d'échantillon** :

- Le volet **Paramètre à estimer** configure le type de paramètre à estimer au départ de l'échantillon dont on cherche à déterminer la taille. Les options **Moyenne**, **Proportion** (valeurs comprises entre 0 et 1), **Différence des moyennes** et **Différence des proportions** sont proposées.

- Les options du volet **Intervalle de confiance** varient selon le paramètre à estimer sélectionné :
 - Pour **Moyenne** et **Différence des moyennes**, on entre le **niveau de confiance** désiré (généralement entre 90 et 100 %), la **demi-longueur d'intervalle** (élément « plus ou moins » de l'intervalle) et l'**écart type estimé** de la population. Remarque : Le niveau de confiance et la longueur d'intervalle sont liés : plus le niveau de confiance est élevé, plus l'intervalle doit être long.
 - Pour **Proportion** et **Différence des proportions**, on entre le **niveau de confiance** désiré (généralement entre 90 et 100 %), la **demi-longueur d'intervalle** (élément « plus ou moins » de l'intervalle) et la **proportion estimée** de la population. Si la **différence des proportions** est estimée, on indique la **proportion estimée** pour chaque population.

Rapport de sélection de taille d'échantillon



Données manquantes et liaison aux données

- **Données manquantes** – Considération sans objet.
- **Liaison aux données** – Considération sans objet.

Commande ANOVA simple

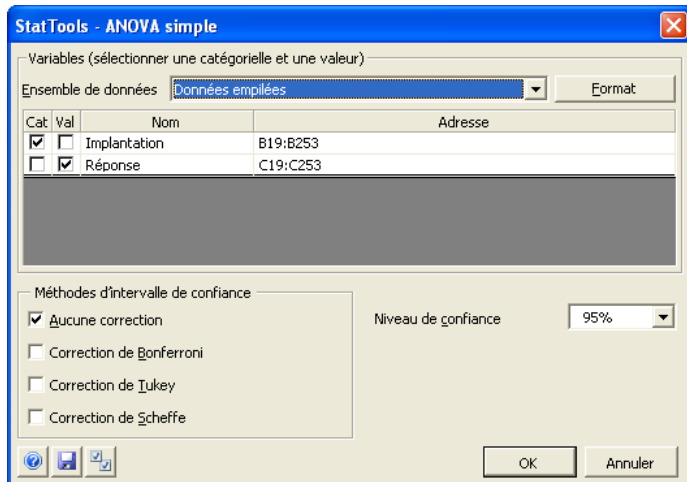
Effectue une analyse de variance « ANOVA » simple sur les variables.

La commande ANOVA simple généralise la procédure à deux échantillons pour la comparaison des moyennes de deux populations. Sous l'analyse ANOVA simple, les moyennes d'*au moins deux* (généralement plus de deux) populations sont comparées. La procédure s'effectue au moyen d'une table ANOVA (analyse de variance). Deux sources de variation y sont comparées : la variation *interne* à chaque population par rapport à la variation *entre* les moyennes d'échantillon de différentes populations. Une variation entre les moyennes supérieure à la variation interne, telle que mesurée par un test F, est signe de différences entre les moyennes des populations.

La valeur clé de la table ANOVA est la valeur *p*. Une valeur *p* faible est signe de moyennes de population divergentes. Outre la table ANOVA, les intervalles de confiance relatifs à toutes les différences entre les paires de moyennes peuvent être révélateurs. Les intervalles qui ne couvrent *pas* 0 sont signes de moyennes *non* égales. StatTools propose différents types d'intervalles de confiance basés, chacun, sur une méthode légèrement différente.

Boîte de dialogue ANOVA simple

Cette analyse se configure dans la boîte de dialogue ANOVA simple :



Au moins deux variables doivent être sélectionnées pour cette analyse. Les ensembles de données sélectionnés peuvent être au format empilé ou déempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue ANOVA simple :

- Le volet **Méthodes d'intervalle de confiance** permet de sélectionner une ou plusieurs méthodes de correction des intervalles de confiance pour les variables individuelles : **Aucune correction** ou méthode de **Bonferroni, Tukey** et/ou **Scheffe**.
- **Niveau de confiance** désigne un niveau de confiance « simultané » pour les résultats de toutes les variables : il s'agit du niveau de confiance que l'on désire avoir que *tous* les intervalles de confiance contiennent leurs différences de moyennes de population respectives. Pour des raisons techniques, le niveau de confiance global *réel* est généralement inférieur au niveau spécifié ici en l'absence de correction des intervalles de confiance, d'où les méthodes de « correction » proposées plus haut : ces méthodes corrigent (en l'étendant) la longueur des intervalles de confiance pour que le niveau de confiance général soit tel que spécifié.

Rapport d'ANOVA simple

Classeur3 - Microsoft Excel

Statistiques de synthèse Tests de normalité Contrôle de qualité
 Graphiques de synthèse Série temporelle et prévision Tests non paramétriques
 Inférence statistique Régression et classification

StatTools Utilitaires Aide

Rapport StatTools

Rapport StatTools

Analyse : ANOVA simple
 Effectué par : Palisade
 Date : jeudi 17 décembre 2009
 Actualisation : En direct

Synthèse ANOVA

Taille totale d'échantillon	235
Moyenne générale	5,383
Ecart type pondéré	1,976
Variance pondérée	3,904
Nombre d'échantillons	5
Niveau de confiance	95,00%

Statistiques d'échantillon ANOVA

	Réponse (Midwest)	Réponse (Nord-Est)	Réponse (Ouest)	Réponse (Sud)	Réponse (Sud-Ouest)
	Données emplies				
Taille d'échantillon	55	43	50	40	47
Moyenne d'échantillon	5,400	4,140	4,980	5,600	6,745
Ecart type d'échantillon	2,469	1,820	1,635	2,073	1,687
Variance d'échantillon	6,096	3,313	2,673	4,297	2,846
Pondération	0,2348	0,1826	0,2130	0,1696	0,2000

Table ANOVA simple

	Somme des carrés	Degrés de liberté	Carrés moyens	Rapport F	Valeur p
Variation entre les groupes	163,653	4	40,913	10,480	< 0,0001
Variation entre les groupes	897,879	230	3,904		
Variation totale	1061,532	234			

Tests d'intervalle de confiance des moyennes

	Différence des moyennes	Sans correction	
		Inférieur	Supérieur
Réponse (Midwest)-Réponse (Nord-Est)	1,260	0,467996986	2,052933246
Réponse (Midwest)-Réponse (Ouest)	0,420	-0,340698646	1,180698646
Réponse (Midwest)-Réponse (Sud)	-0,200	-1,008974445	0,608974445
Réponse (Midwest)-Réponse (Sud-Ouest)	-1,345	-2,117991788	-0,571369914
Réponse (Nord-Est)-Réponse (Ouest)	-0,840	-1,650132388	-0,030797844
Réponse (Nord-Est)-Réponse (Sud)	-1,460	-2,315648321	-0,605281911
Réponse (Nord-Est)-Réponse (Sud-Ouest)	-2,605	-3,426674097	-1,783617838
Réponse (Ouest)-Réponse (Sud)	-0,620	-1,445829889	0,205829889
Réponse (Ouest)-Réponse (Sud-Ouest)	-1,765	-2,55607623	-0,973754079
Réponse (Sud)-Réponse (Sud-Ouest)	-1,145	-1,98214272	-0,307218982

Dans le rapport ANOVA simple illustré ci-dessus, les statistiques de synthèse de chaque population (en l'occurrence, chaque usine) sont présentées en premier. Ces statistiques sont suivies d'une table de statistiques d'échantillon relatives à chaque variable considérée. Vient ensuite la table ANOVA. Dans cet exemple, la très faible valeur p indique clairement que les cotes de moyenne des cinq usines ne sont pas toutes égales. Les intervalles de confiance, au bas du rapport, révèlent les différences. Les paires dont les valeurs sont affichées en caractères gras présentent une différence de moyennes significative.

Données manquantes et liaison aux données

- **Données manquantes** - Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** - Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Commande ANOVA double

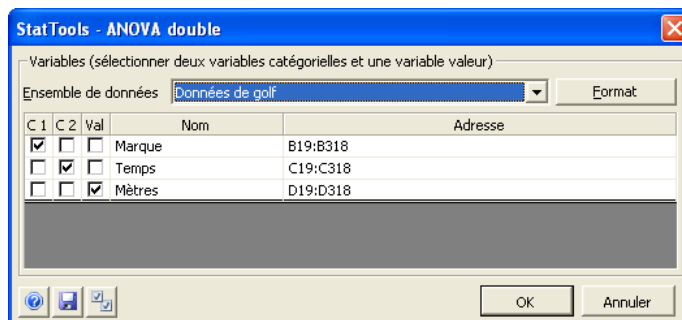
Effectue une analyse de variance « ANOVA » double sur les variables.

La commande ANOVA double effectue une analyse de variance double. Cette analyse s'effectue généralement dans le contexte d'un concept expérimental à deux « facteurs » réglés, chacun, à différents « niveaux de traitement ». Par exemple, dans une étude de performance de balles de golf, les deux facteurs considérés pourraient être la *marque* et le *temps* (température extérieure). Les niveaux de traitement de *Marque* sont ici « A » à « E », et ceux de *Temps* sont « Frais », « Doux » et « Chaud ». La variable de valeur est *Mètres*, et les observations relatives à cette variable sont recueillies pour un certain nombre de balles à chaque combinaison *Marque/Temps*. Le but de l'étude est de déterminer si les différences de moyenne sont significatives entre les différentes combinaisons de niveaux de traitement.

Les données de l'analyse ANOVA double *doivent* être au format empilé : il doit y avoir deux variables « catégorie » (*Marque* et *Temps*, dans l'exemple décrit) et une variable « valeur » (*Mètres*). Les données doivent aussi être « équilibrées » : il doit y avoir un nombre égal d'observations à chaque combinaison de niveaux de traitement. Il est certainement possible d'analyser un modèle non équilibré, mais il vaut alors mieux procéder par régression (avec variables nominales dummy).

Boîte de dialogue ANOVA double

Cette analyse se configure dans la boîte de dialogue ANOVA double :



Deux variables catégorielles (C1 et C2) et une variable valeur (Val) doivent être sélectionnées. L'ensemble de données sélectionné doit être au format empilé.

Rapport d'analyse ANOVA double

Classeur3 - Microsoft Excel

Statistiques de synthèse - Tests de normalité - Contrôle de qualité -
 Graphiques de synthèse - Série temporelle et prévision - Tests non paramétriques -
 Inférence statistique - Régression et classification - Analyses

Rapport StatTools

1 **Rapport StatTools**

2 Analyse : ANOVA double

3 Effectué par : Fallade

4 Date : jeudi 17 décembre 2009

5 Actualisation : En direct

6

7

Tailles d'échantillon ANOVA	Chaud	Doux	Frais	Totaux
9 A	20	20	20	60
10 B	20	20	20	60
11 C	20	20	20	60
12 D	20	20	20	60
13 E	20	20	20	60
14 Totaux	100	100	100	
15 Equilibre	VRAI			

Moyennes d'échantillon ANOVA	Chaud	Doux	Frais	Totaux
19 A	258,44	236,45	218,82	237,90
20 B	258,27	245,13	224,15	242,52
21 C	263,04	242,72	228,00	244,58
22 D	256,11	237,62	215,00	236,24
23 E	270,94	255,75	224,79	250,49
24 Totaux	261,36	243,53	222,15	

Écarts types d'échantillon ANOVA	Chaud	Doux	Frais	Totaux
27 A	11,01	8,83	10,90	19,22
28 B	8,93	9,80	11,70	17,36
30 C	7,08	14,25	10,85	18,15
31 D	12,13	10,18	13,64	20,69
32 E	9,05	10,96	10,67	21,84
33 Totaux	10,98	12,78	12,28	

Table ANOVA double	Somme des carrés	Degrés de liberté	Carrés moyens	Rapport F	Valeur p
37 Marque	7702,44	4	1925,61	16,47	< 0,0001
38 Temps	77086,00	2	38543,00	329,58	< 0,0001
39 Interaction	1999,97	8	250,00	2,14	0,0325
40 Erreur	33329,13	285	116,94		
41 Total	120117,53	299			

Les trois premiers éléments du rapport sont les mesures de synthèse (tailles, moyennes et écarts types des échantillons) aux différentes combinaisons de niveaux de traitement. Vient ensuite la table ANOVA. Trois valeurs p importantes y figurent : deux pour les « effets principaux » et une d'« interaction ». Les effets principaux indiquent l'existence ou non de différences de moyennes significatives sur les niveaux d'un facteur, en moyenne sur ceux de l'autre facteur. Par exemple, l'effet principal du Temps indique si les valeurs des cellules B17 à D17 sont significativement différentes. (Elles le sont, comme indiqué par la très faible valeur p de Temps dans la table ANOVA).

***Données
manquantes et
liaison aux
données***

- **Données manquantes** – Étant donné le critère d'équilibre, il ne devrait y avoir aucune donnée manquante.
- **Liaison aux données** – Toutes les formules ANOVA sont liées aux données. Si les données changent, les résultats changent automatiquement. L'équilibre de l'expérience est vérifié au moment de l'analyse. Les changements de données risquent d'affecter l'équilibre et de produire des résultats incorrects.

Commande Test d'indépendance chi carré

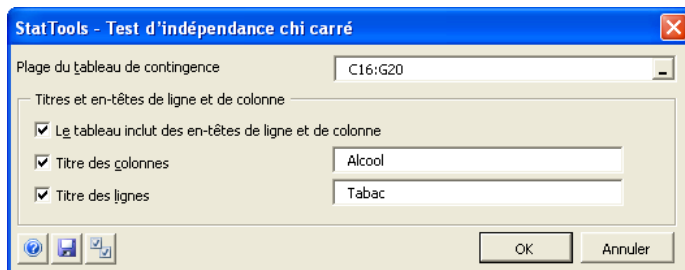
Teste l'indépendance entre les attributs de ligne et de colonne d'un tableau de contingence.

La commande Test d'indépendance chi carré teste l'indépendance entre les attributs de ligne et de colonne d'un tableau de contingence. Par exemple, si le tableau de contingence liste des nombres de personnes appartenant à différentes catégories de consommation d'alcool et de tabac, la procédure teste si les habitudes de tabagisme sont indépendantes de celles de consommation d'alcool. Le tableau de contingence (aussi appelé tableau à double entrée) pourrait bien être un tableau croisé dynamique Excel.

Cette procédure diffère quelque peu de la plupart des analyses StatTools. Pour cette analyse, la seule exigence est un tableau de contingence rectangulaire. Chaque cellule du tableau doit représenter un nombre d'observations pour une combinaison ligne/colonne particulière (non buveurs et grands fumeurs, par exemple). Les lignes et colonnes du tableau peuvent être étiquetées (en-têtes) et/ou totalisées. Utiles à la clarté des rapports de StatTools, ces indications ne sont pas obligatoires.

Boîte de dialogue Test d'indépendance chi carré

Cette analyse se configure dans la boîte de dialogue **Test d'indépendance chi carré** :



Dans la boîte de dialogue Test d'indépendance chi carré :

- Les options suivantes sont proposées dans le volet **Titres et en-têtes de ligne et de colonne** : **Le tableau inclut des en-têtes de ligne et de colonne** (colonne la plus à gauche et ligne supérieure du tableau) ; **Titre des colonnes** (titre à utiliser pour représenter les colonnes du tableau) et **Titre des lignes** (titre à utiliser pour représenter les lignes).

Rapport de test d'indépendance chi carré

Classeur3 - Microsoft Excel

Statistiques de synthèse Tests de normalité Contrôle de qualité
 Graphiques de synthèse Série temporelle et prévision Tests non paramétriques
 Inférence statistique Régression et classification

StatTools

Accueil Insertion Mise en page Formules Données Révision Affichage

Gestionnaire Utilitaires

Données

A6

1 **Rapport StatTools**

2 Analyse : Test d'indépendance chi carré

3 Effectué par : Pallsade

4 Date : jeudi 17 décembre 2009

5 Actualisation : En direct

6

7

8 Lignes : Tabac / Colonnes : Alcool

Nombres originaux	Non	Légère	Moderée	Forte	Total
Non	20	15	10	10	55
Consommation d'alcool	10	15	12	15	52
Moderée	8	12	15	20	55
Forte	6	10	15	25	56
Total	44	52	52	70	218

14

15 Lignes : Tabac / Colonnes : Alcool

Pourcentage des lignes	Non	Légère	Moderée	Forte	Total
Non	36,36%	27,27%	18,18%	18,18%	100,00%
Consommation d'alcool	19,23%	28,85%	23,08%	28,85%	100,00%
Moderée	14,55%	21,82%	27,27%	36,36%	100,00%
Forte	10,71%	17,86%	26,79%	44,64%	100,00%

21

22 Lignes : Tabac / Colonnes : Alcool

Pourcentage des colonnes	Non	Légère	Moderée	Forte	Total
Non	45,45%	28,85%	19,23%	14,29%	100,00%
Consommation d'alcool	22,73%	28,85%	23,08%	21,43%	100,00%
Moderée	18,18%	23,08%	28,85%	28,57%	100,00%
Forte	13,64%	19,23%	28,85%	35,71%	100,00%
Total	100,00%	100,00%	100,00%	100,00%	100,00%

29

30 Lignes : Tabac / Colonnes : Alcool

Nombres attendus	Non	Légère	Moderée	Forte	Total
Non	11,1009	13,1193	13,1193	17,6606	55
Consommation d'alcool	10,4954	12,4037	12,4037	16,6972	52
Moderée	11,1009	13,1193	13,1193	17,6606	55
Forte	11,3028	13,3578	13,3578	17,9817	56

36

37 Lignes : Tabac / Colonnes : Alcool

Distance	Non	Légère	Moderée	Forte	Total
Non	7,1340	0,2696	0,7416	3,3229	11,4681
Consommation d'alcool	0,0234	0,5435	0,0131	0,1725	0,7525
Moderée	0,8662	0,0955	0,2696	0,3099	1,5412
Forte	2,4878	0,8441	0,2019	2,7393	6,2721

43

44

45 Statistique chi carré

46 Chi carré 20,0349

47 Valeur p 0,0177

Prêt

Le rapport illustré ci-dessus présente le résultat fondamental du test : sa valeur p. Si cette valeur est faible (comme c'est le cas ici), on peut conclure que les attributs de ligne et de colonne ne sont *pas* indépendants. Un examen plus approfondi de cette feuille permet de mieux comprendre la relation entre boisson et tabac.

Données manquantes et liaison aux données

- **Données manquantes** – Il ne devrait pas y avoir de données manquantes dans les cellules du tableau de contingence.
- **Liaison aux données** – Les formules du rapport du test d'indépendance chi carré sont liées aux données. Si les nombres du tableau de contingence original changent, les sorties présentées sur cette feuille changent aussi.

Menu Tests de normalité

Beaucoup de procédures statistiques reposent sur l'hypothèse d'ensembles de données normalement distribués. Il est donc utile de disposer de méthodes aptes à vérifier cette hypothèse. StatTools propose trois tests de vérification courants, tels que décrits dans cette section.

Commande Test de normalité chi carré

Vérifie si les données observées pour une variable sont normalement distribuées.

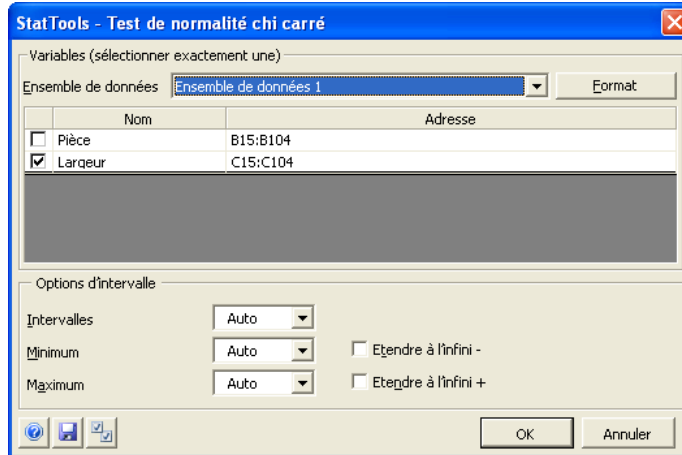
La procédure Test de normalité chi carré recourt au test de qualité d'ajustement chi carré pour vérifier si les données observées d'une variable spécifiée proviennent effectivement d'une distribution normale. Ce test crée un histogramme de la variable selon les catégories spécifiées, puis y superpose un histogramme de distribution normale. Si les deux histogrammes ont essentiellement la même forme, l'hypothèse nulle d'ajustement normal ne peut pas être rejetée.

Ce test formel s'effectue par comparaison des nombres observés dans les différentes catégories avec ceux attendus sur la base d'une supposition de normalité. La procédure permet de tester (séparément) la normalité de plusieurs variables. Un histogramme se crée pour chaque variable sélectionnée et le test chi carré s'exécute sur chacun.

La seule condition nécessaire à l'exécution du test de normalité chi carré est la présence d'au moins une variable numérique. La plupart des analystes recommandent sinon au moins 100 observations : plus le nombre d'observations est grand, plus le test est fiable.

**Boîte de dialogue
Test de normalité
chi carré**

Cette analyse se configure dans la boîte de dialogue **Test de normalité chi carré** :

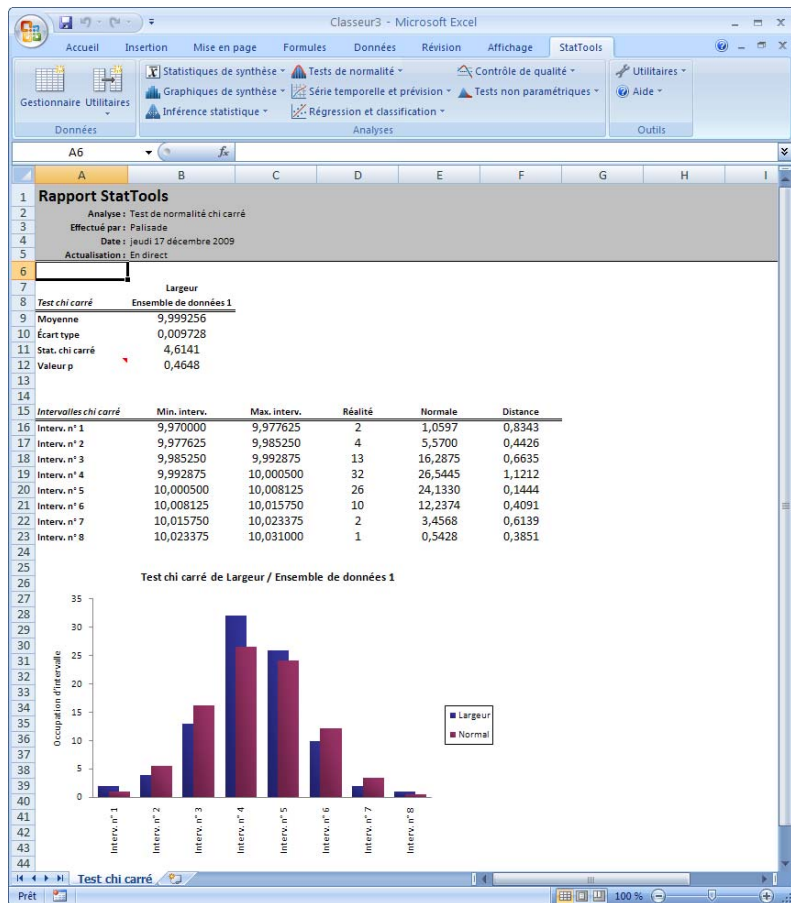


Le test requiert la sélection d'une variable. L'ensemble de données sélectionné peut être au format empilé ou déempilé.

Dans la boîte de dialogue **Test de normalité chi carré** :

- **Intervalles** spécifie un nombre d'intervalles fixe ou, automatiquement, le nombre calculé par le système.
- **Minimum** et **Maximum - auto** spécifient le calcul des valeurs minimum et maximum d'intervalles égaux en fonction des valeurs minimum et maximum de l'ensemble de données. Le premier et le dernier intervalles peuvent cependant être définis à l'aide des options **Étendre à l'infini -** et **Étendre à l'infini +**. Si l'option de calcul automatique n'est pas sélectionnée, on peut entrer directement les valeurs **Minimum** et **Maximum** devant marquer le début et la fin des intervalles et définir ainsi une plage précise de répartition, indépendante des valeurs minimum et maximum de l'ensemble de données.
- **Étendre à l'infini -** spécifie un premier intervalle s'étendant du minimum spécifié à l'infini négatif. Tous les autres intervalles sont de longueur égale. Dans certaines circonstances, cette option améliore le test d'ensembles de données dont la limite inférieure est inconnue.
- **Étendre à l'infini +** spécifie un dernier intervalle s'étendant du maximum spécifié à l'infini positif. Tous les autres intervalles sont de longueur égale. Dans certaines circonstances, cette option améliore le test d'ensembles de données dont la limite supérieure est inconnue.

Rapport de test de normalité chi carré



Le rapport ci-dessus illustre les résultats du test. La valeur p de 0,4776 reflète bien la distribution normale des nombres. Les histogrammes et les données de fréquence la confirment aussi. Deux remarques importantes s'imposent cependant. D'abord, si le nombre d'observations est insuffisant (largement inférieur à 100, par exemple), le test chi carré ne distingue pas bien la normalité de la non-normalité. La valeur p n'est en effet généralement pas suffisamment faible pour rejeter l'hypothèse de normalité. Pratiquement *tout* semble normal quand les ensembles de données sont réduits. En revanche, si l'ensemble de données est extrêmement vaste (plusieurs centaines d'observations, par exemple), la valeur p est généralement faible, signe de non-normalité. La raison en est qu'en présence de tels ensembles de données, chaque petite « excroissance » au niveau de la courbe tend à produire une valeur p faible. Le véritable test est alors de nature *pratique*, en ce qu'il revient à déterminer, à toutes fins utiles, si les histogrammes diffèrent vraiment l'un de l'autre.

***Données
manquantes et
liaison aux
données***

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – L’histogramme et toutes les formules du test sont liés aux données originales. Si les données changent, l’histogramme et les résultats du test changent automatiquement.

Commande Test de Lilliefors

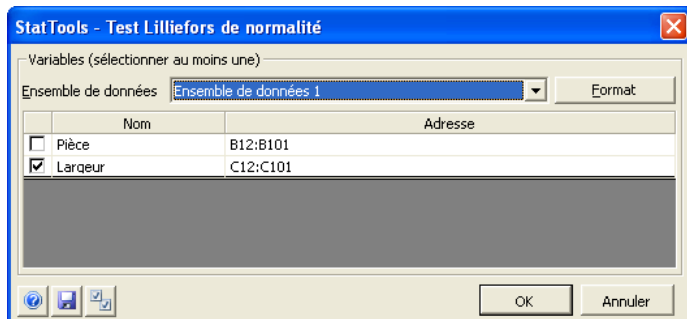
Vérifie si les données observées pour une variable sont normalement distribuées.

La procédure Lilliefors offre un test de normalité plus puissant que le test plus courant de qualité d'ajustement chi carré. (La puissance du test tient au fait qu'il est plus susceptible de détecter la non-normalité.) Le test de Lilliefors repose sur une comparaison de la « fdc empirique » et d'une fdc normale (fdc = fonction de distribution cumulative), indiquant la probabilité d'une situation d'infériorité ou égalité à une valeur particulière.

La fdc empirique repose sur les données. Par exemple, s'il y a 100 observations et que la 13^e, en partant de la plus faible, représente une valeur de 137, la fdc évaluée à 137 est 0,13. Le test de Lilliefors recherche la distance verticale maximale entre la fdc empirique et la fdc normale, puis compare ce maximum aux valeurs d'un tableau (basées sur la taille de l'échantillon). Une distance verticale maximale observée suffisamment grande est signe de données *non* originaires d'une distribution normale.

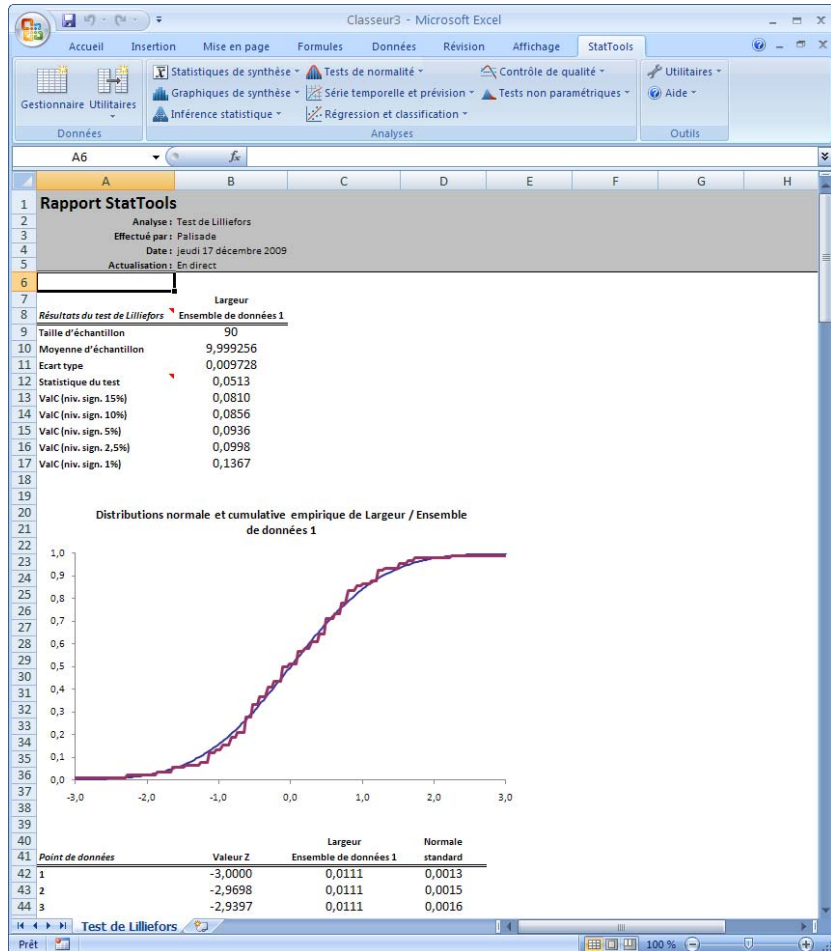
Boîte de dialogue Test Lilliefors de normalité

Cette analyse se configure dans la boîte de dialogue **Test Lilliefors de normalité** :



Une ou plusieurs variables peuvent être sélectionnées pour le test. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Rapport de test de Lilliefors



Le rapport ci-dessus illustre les résultats du test. Bien qu'il n'y ait pas de valeur p (comme dans la plupart des tests d'hypothèse), la distance verticale maximale semble suffisamment large pour mettre en doute l'hypothèse de normalité. Les fdc représentées dans le graphique en témoignent aussi. L'ajustement des deux courbes est cependant « relativement bon » et pourrait bien être suffisant à toutes fins utiles. On pourrait donc conclure que ces données sont « suffisamment proches » d'une distribution normale aux fins poursuivies.

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Les FDC et toutes les formules du test sont liées aux données originales. Si les données changent, le graphique et les résultats du test changent automatiquement.

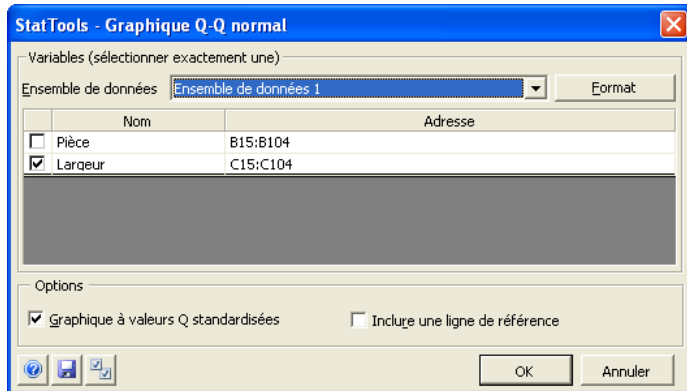
Commande Graphique Q-Q normal

Vérifie si les données observées pour une variable sont normalement distribuées.

La commande Graphique Q-Q normal crée un tracé double quantile (Q-Q) pour une variable. Elle offre ainsi un test informel de normalité. Si les détails en sont assez complexes, l'objectif est relativement simple : il s'agit de comparer les quantiles (ou centiles) des données à ceux d'une distribution normale. Si les données sont essentiellement normales, les points du tracé Q-Q sont proches d'une ligne à 45 degrés. La courbure évidente du tracé est cependant indicatrice d'un certain degré de non-normalité (asymétrie, par exemple).

Boîte de dialogue Graphique Q-Q normal

Cette analyse se configure dans la boîte de dialogue **Graphique Q-Q normal** :

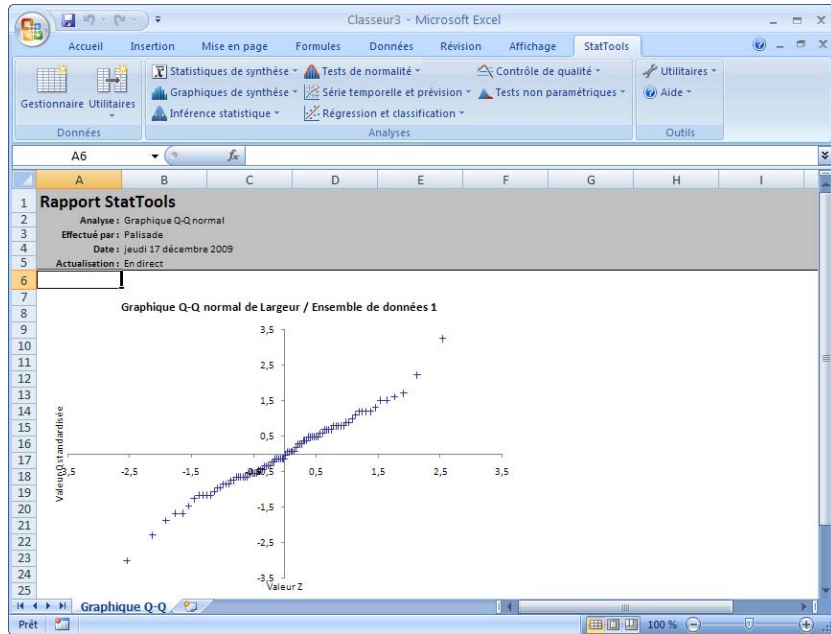


Une variable peut être sélectionnée pour le tracé. L'ensemble de données sélectionné doit être au format désempilé.

Dans la boîte de dialogue **Graphique Q-Q normal** :

- **Graphique à valeurs Q standardisées** spécifie l'usage d'une valeur Q standardisée, plutôt que de données Q-Q, sur l'axe Y du graphique. Cette option permet la comparaison des valeurs d'axe Y entre graphiques Q-Q normaux.

Rapport de graphique Q-Q normal



Comme indiqué plus haut, il s'agit ici d'un test *informel* de normalité. Il est difficile de déterminer la « proximité » de 45 degrés nécessaire à l'acceptation de l'hypothèse de normalité. On considère généralement plutôt la courbure évidente du tracé, non apparente dans l'exemple illustré ici.

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Les tracés et toutes les formules sont liés aux données originales. Si les données changent, les tracés changent automatiquement.

Menu Série temporelle et prévision

Les procédures proposées dans le menu Série temporelle et prévision concernent l'analyse de données collectées au fil du temps, applicable aux situations de prévision et de contrôle de qualité. Les méthodes de prévision considérées sont celles des moyennes mobiles, de lissage exponentiel simple, de lissage exponentiel de Holt pour la capture de tendance et de lissage exponentiel de Winters pour la capture de saisonnalité.

Commande Chronogramme

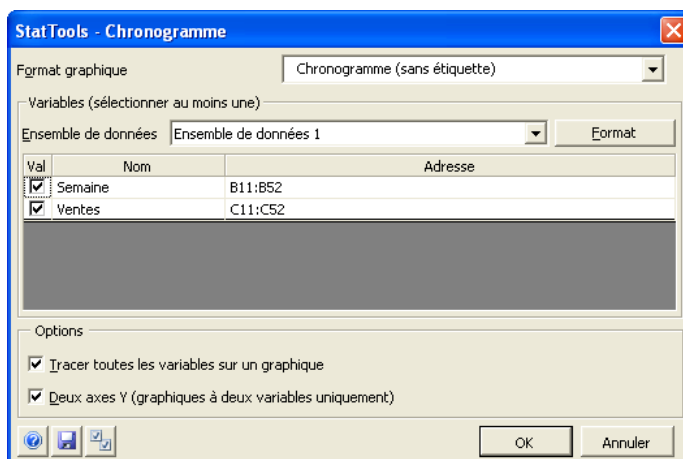
Trace un chronogramme des variables.

La commande Chronogramme trace le chronogramme d'une ou plusieurs variables de série temporelle sur un même diagramme. Si deux variables sont sélectionnées, l'échelle peut être identique ou différente, pour les deux variables, sur l'axe Y. Les échelles différentes sont utiles quand les plages de valeurs des deux variables diffèrent largement. Pour plus de deux variables toutefois, une même échelle verticale doit être partagée.

Il doit y avoir au moins une variable numérique dans l'ensemble de données. Une variable « date » est aussi admise, mais si cette variable doit servir d'étiquette à l'axe horizontal du graphique, elle doit être sélectionnée en tant que telle (variable « étiquette »).

Boîte de dialogue Chronogramme

Ce type de graphique se configure dans la boîte de dialogue Chronogramme :

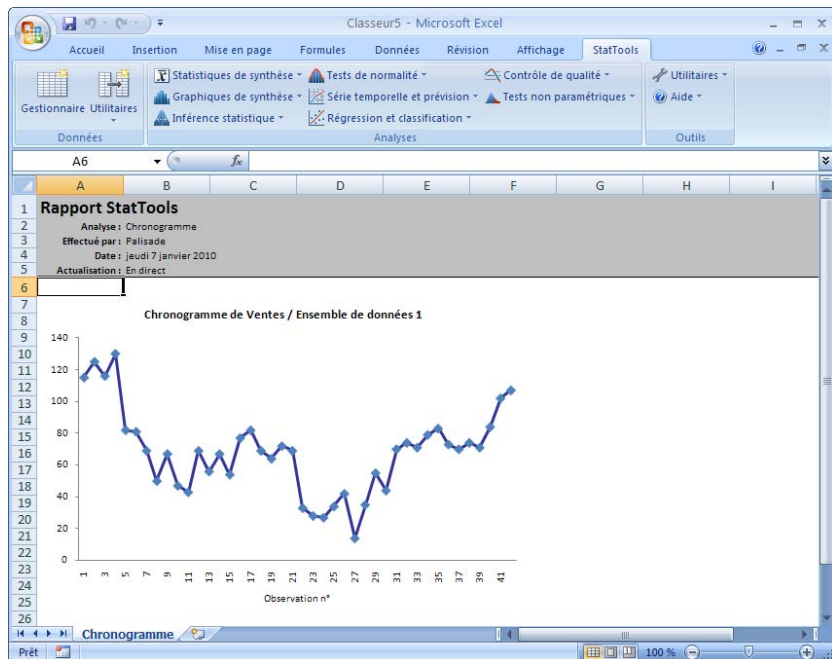


Une ou plusieurs variables peuvent être sélectionnées pour la représentation graphique. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents. La variable Étiquette (case Etq) figure sur l'axe des X.

Dans la boîte de dialogue **Chronogramme** :

- **Tracer toutes les variables sur un graphique** permet de représenter toutes les variables sur un même graphique.
- **Deux axes Y** affiche un axe Y différent pour chaque variable sur un graphique à deux variables. Les unités et valeurs de chaque variable peuvent ainsi s'afficher sur le graphique.

Chronogramme de deux variables



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Les graphiques sont liés aux données, de sorte que si les données changent, le graphique s'actualise automatiquement.

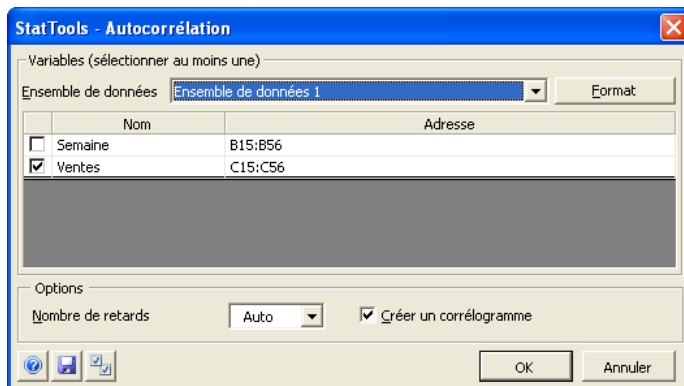
Commande Autocorrélation

Calcule les autocorrélations des variables.

La commande Autocorrélation calcule les autocorrélations d'une variable numérique sélectionnée. Il s'agit généralement d'une variable de série temporelle, mais StatTools calcule aussi les corrélations d'autres types. Le nombre d'autocorrélations (« retards ») désiré peut être sélectionné dans la boîte de dialogue. Leur graphique, appelé « corrélogramme », peut aussi être demandé. La sortie indique, le cas échéant, les autocorrélations significativement différentes de 0.

Boîte de dialogue Autocorrélation

Ce type de graphique se configure dans la boîte de dialogue **Autocorrélation** :

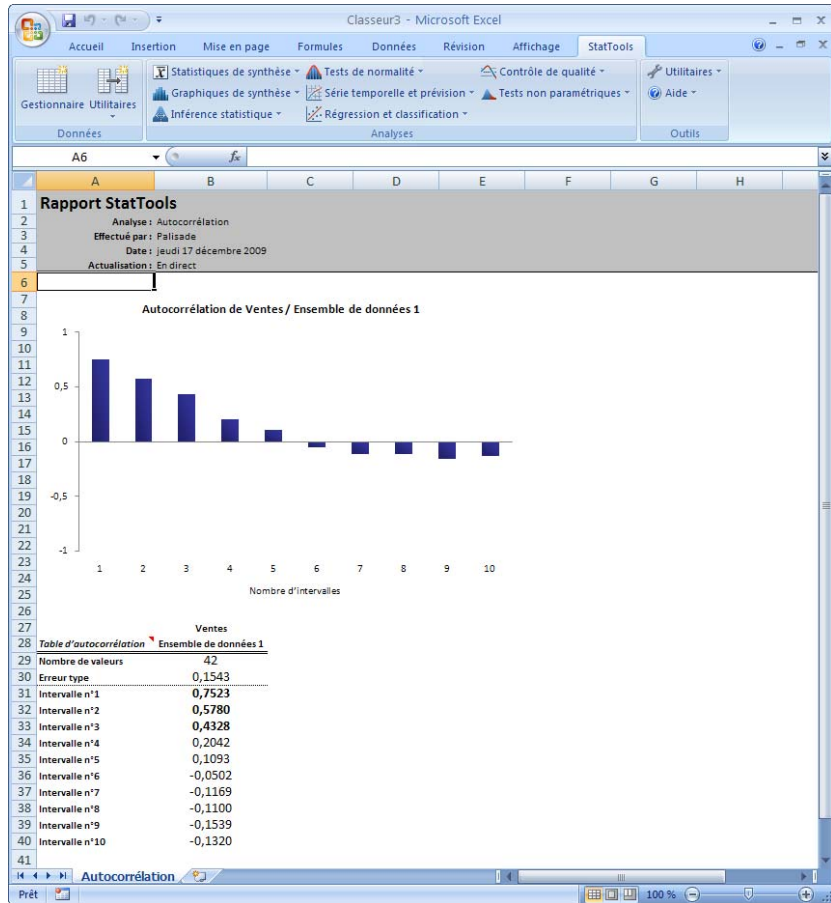


Une ou plusieurs variables peuvent être sélectionnées pour l'analyse. L'ensemble de données sélectionné doit être au format désempilé.

Dans la boîte de dialogue **Autocorrélation** :

- **Nombre de retards** représente le nombre de périodes de retard à appliquer lors du calcul des autocorrélations. Si l'option **Auto** est sélectionnée, StatTools détermine le nombre de retard à tester. Pour un nombre de retards spécifique, le nombre maximum admis est limité à 25 % du nombre d'observations de la série. Ainsi, pour 80 valeurs mensuelles, on peut demander un maximum de 20 retards.
- **Créer un corrélogramme** trace un graphique à barres dans lequel la hauteur de chaque barre représente l'autocorrélation correspondante.

Rapport d'autocorrélation



Un rapport d'autocorrélation est illustré ci-dessus. L'autocorrélation correspondant à chaque retard est représentée, avec une erreur type *approximative*.

Données manquantes et liaison aux données

- **Données manquantes** - Cette procédure admet les données manquantes en début de série temporelle, mais pas au milieu ni à la fin.
- **Liaison aux données** - StatTools lie la sortie aux données. Si les données changent, les autocorrélations (et le corrélogramme) changent automatiquement.

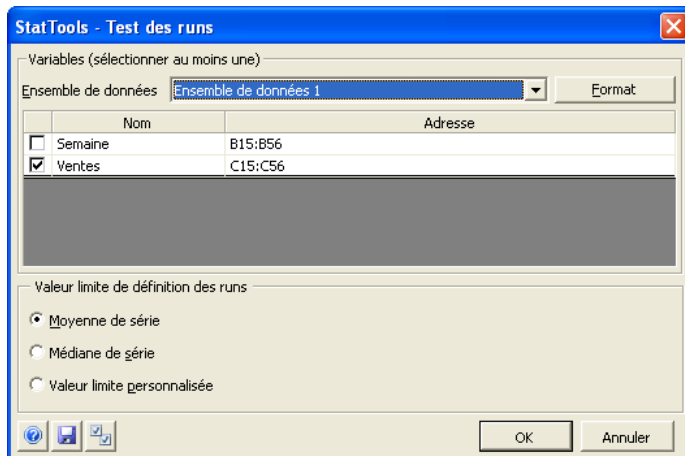
Commande Test des runs

Effectue un test des runs pour vérifier si une variable est aléatoire.

La commande Test des runs permet de vérifier le caractère « aléatoire » d'une séquence de valeurs dans une variable, généralement de série temporelle. Elle indique le nombre de « runs » présents dans une séquence, chaque « run » représentant un nombre consécutif de valeurs d'un côté ou de l'autre d'un point limite donné (la moyenne ou la médiane de la séquence, par exemple). Pour une séquence aléatoire, le nombre de runs ne doit être ni trop faible, ni trop élevé. Le test des runs détermine le nombre de runs et produit la valeur p du test. Une valeur p faible est signe de séquence probablement non aléatoire (trop ou trop peu de runs).

Boîte de dialogue Test des runs

Cette analyse se configure dans la boîte de dialogue **Test des runs** :

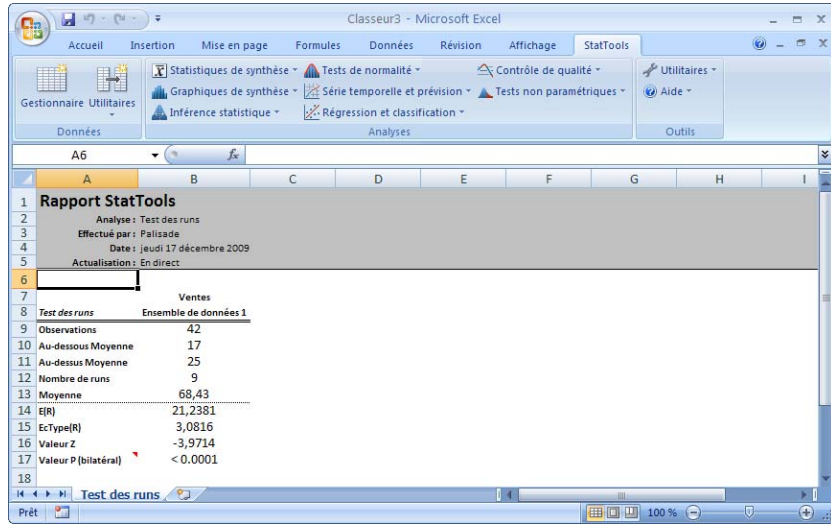


Une ou plusieurs variables peuvent être sélectionnées pour l'analyse. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Test des runs** :

- **Valeur limite de définition des runs** – Un test des runs se définit toujours en fonction des runs observés au-dessus ou au-dessous d'une valeur limite. Cette valeur peut être la **moyenne de série**, la **médiane de série** ou la **valeur limite personnalisée** spécifiée sous cette option.

Rapport de test des runs



The screenshot shows the Microsoft Excel interface with the StatTools ribbon active. The 'Test des runs' report is displayed in the worksheet. The report title is 'Rapport StatTools'. The analysis is 'Analyse : Test des runs', performed by 'Palisade' on 'jeudi 17 décembre 2009'. The data is updated 'En direct'. The report includes a table of statistics for 'Ensemble de données 1'.

Test des runs	Ensemble de données 1
Observations	42
Au-dessous Moyenne	17
Au-dessus Moyenne	25
Nombre de runs	9
Moyenne	68,43
E(R)	21,2381
EcType(R)	3,0816
Valeur Z	-3,9714
Valeur P (bilatéral)	< 0.0001

Le rapport illustré ci-dessus indique le nombre de runs et le nombre probable sous variable aléatoire $E(R)$. La valeur 20 est significativement inférieure à 24,8333. Cette série n'est donc pas entièrement aléatoire : les ventes ne « zigzaguerait » par autant dans une série aléatoire.

Données manquantes et liaison aux données

- **Données manquantes** – Cette procédure admet les données manquantes au début et à la fin de la série temporelle, mais pas au milieu.
- **Liaison aux données** – StatTools lie la sortie aux données. Si les données changent, les rapports changent automatiquement.

Commande Prévision

Génère les prévisions de variables de série temporelle.

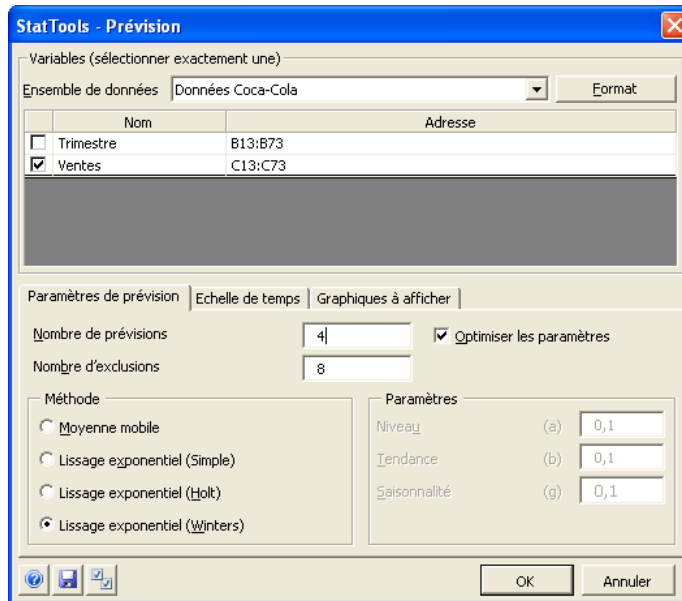
La commande Prévision propose différentes méthodes de prévision d'une variable de série temporelle. Les méthodes considérées sont celles des moyennes mobiles, de lissage exponentiel simple, de lissage exponentiel de Holt pour la capture de tendance et de lissage exponentiel de Winters pour la capture de saisonnalité. La commande Prévision permet aussi de désaisonnaliser les données d'abord, selon la méthode de rapport aux moyennes mobiles et un modèle de saisonnalité multiplicative. Toutes les méthodes de prévision (autres que celle de Winters) peuvent ensuite servir à la prévision des données désaisonnalisées. Cela fait, les prévisions peuvent être « resaisonnalisées » pour revenir aux unités originales.

Les rapports de prévision incluent une série de colonnes de présentation des calculs (niveaux lissés et tendances de la méthode Holt, facteurs saisonniers de celle des moyennes mobiles, etc.), des prévisions et des erreurs de prévision. Les mesures de synthèse sont également incluses (EAM, EMQ et EPAM) pour le suivi de correspondance du modèle aux données observées. (Sous les méthodes de lissage exponentielles, l'option est donnée de recourir à l'optimisation pour identifier la ou les constantes de lissage aptes à minimiser l'EMQ).

Enfin, plusieurs tracés graphiques peuvent être configurés, y compris un tracé de la série originale, de la série avec superposition des prévisions et des erreurs de prévision. En cas de désaisonnalisation, les graphiques sont proposés pour la série originale et pour celle désaisonnalisée.

Boîte de dialogue Prévision

Les prévisions se configurent dans la boîte de dialogue **Prévision** :



Une variable peut être sélectionnée pour l'analyse. L'ensemble de données sélectionné doit être au format désempilé.

Paramètres de prévision – Boîte de dialogue Prévision

L'onglet **Paramètres de prévision** de la boîte de dialogue permet de spécifier la méthode de prévision à utiliser et ses paramètres :

- **Nombre de prévisions** spécifie le nombre de périodes *futures* pour lesquelles des prévisions doivent être effectuées.
- **Nombre d'exclusions** spécifie le nombre d'observations à « exclure » du modèle de prévision. Toutes les observations peuvent être utilisées pour l'estimation du modèle de prévision (0 exclusion) ou quelques-unes peuvent être retenues à des fins de validation. Estimé sur la base des observations non exclues, le modèle sert alors à la prévision des observations exclues.
- **Optimiser les paramètres** (méthodes de lissage exponentiel uniquement) permet de rechercher la constante de lissage apte à minimiser l'EMQ (pour la période sans exclusion). L'optimisation exige des paramètres réglés entre 0 et 1 dans le volet Paramètres de l'onglet. En cas de modification directe des valeurs de ces paramètres dans un rapport de prévision, on veillera à respecter ces limites.

- **Désaisonnaliser** permet la désaisonnalisation des données avant la prévision. Pour les données saisonnières (dont on suppose qu'elles suivent une tendance saisonnière), deux options sont possibles : la méthode de Winters traite directement la saisonnalité, ou la sélection de **Désaisonnaliser** permet de désaisonnaliser d'abord les données selon la méthode de rapport aux moyennes mobiles. *N'importe quelle* méthode peut ensuite être sélectionnée pour la prévision de la série désaisonnalisée.
- Le volet **Méthode** sert à sélectionner la méthode de prévision à utiliser : **Moyenne mobile** ou méthode de lissage exponentiel **simple**, de **Holt** ou de **Winters**.
- Dans le volet **Paramètres**, on configure les paramètres à utiliser pour la méthode de prévision sélectionnée :
 - **Portée** (pour la méthode Moyenne mobile) définit le nombre d'observations consécutives de chaque moyenne mobile.
 - **Niveau** (toutes méthodes de lissage exponentiel) est un paramètre de lissage dont la valeur doit être comprise entre 0 et 1 (la valeur par défaut est 0,1).
 - **Tendance** (méthodes de lissage exponentiel Holt et Winters) est un deuxième paramètre de lissage dont la valeur doit être comprise entre 0 et 1 (la valeur par défaut est 0,1).
 - **Saisonnalité** (méthode de lissage exponentiel Winters) est un troisième paramètre de lissage dont la valeur doit être comprise entre 0 et 1 (la valeur par défaut est 0,1).

Remarque : Si l'option Optimiser les paramètres est sélectionnée, les paramètres Niveau, Tendance et Saisonnalité ne peuvent pas être réglés ici puisqu'il s'agit des paramètres dont les valeurs sont optimisées.

Onglet Échelle de temps – Boîte de dialogue Prévision

L'onglet Échelle de temps et ses options servent à spécifier les étiquettes d'échelle de moment et de temps de la variable analysée :

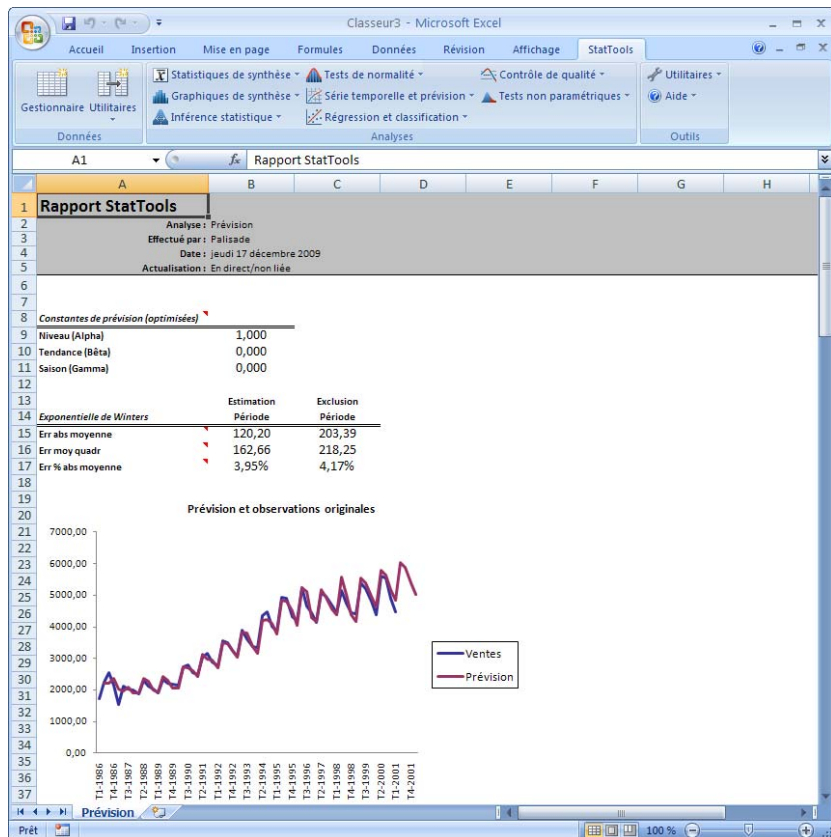
- **Période saisonnière** spécifie le type de données de série temporelle : Annuelle, Trimestrielle, Mensuelle, Hebdomadaire, Journalière ou Aucune. La sélection opérée ici sert à la saisonnalisation des données et à l'étiquetage.
- **Style d'étiquette** spécifie l'étiquetage de l'échelle de temps sur les graphiques générés.
- **Étiquette de départ** spécifie l'entrée de la première étiquette d'échelle de temps sur le graphique.

**Onglet
Graphiques –
Boîte de dialogue
Prévision**

L'onglet Graphiques et ses options configurent les graphiques de prévision à générer :

- 1) **Superposition de prévision** trace le chronogramme des valeurs de données générées par la prévision.
- 2) **Série originale** trace le chronogramme des données réelles.
- 3) **Erreurs de prévision** représente l'erreur entre la prévision et les valeurs réelles.
- 4) **Superposition de prévision désaisonnalisée** trace le chronogramme des valeurs de données générées par la prévision après désaisonnalisation des données originales.
- 5) **Série originale désaisonnalisée** trace le chronogramme des données réelles après leur désaisonnalisation.
- 6) **Erreurs de prévision désaisonnalisées** représente l'erreur entre la prévision et les valeurs réelles après désaisonnalisation des données originales.

**Exemple de
rapport de
prévision**



***Données
manquantes et
liaison aux
données***

- **Données manquantes** – Les données manquantes sont admises en début de série temporelle, mais pas au milieu ni à la fin.
- **Liaison aux données** – Étant donné la longueur des calculs, les prévisions ne sont pas liées aux données. La procédure doit être réexécutée en cas de changement affectant les données originales.

Menu Régression et classification

Les commandes du menu Régression et classification effectuent des analyses de régression et classification. Différents types d'analyse de régression sont proposés : multiple, en escalier, avant, arrière et par bloc. Le menu propose également des commandes d'analyse discriminante et de régression logistique.

Commande Régression

Effectue des analyses de régression sur un ensemble de variables.

La commande Régression exécute différents types d'analyse de régression : multiple, en escalier, avant, arrière et par bloc. Les rapports de chaque analyse incluent les mesures de synthèse de chaque équation de régression exécutée, une table d'analyse de variance ANOVA pour chaque régression et une table des coefficients de régression estimés, leurs écarts types, leurs valeurs t, leurs valeurs p et leurs intervalles de confiance à 95 % pour chaque régression.

De plus, deux nouvelles variables peuvent être créées, valeurs ajustées et résidus, de même que des diagrammes de dispersion diagnostiques.

Types de régression proposés

Les types de régression suivants sont proposés : **Multiple**, **Escalier**, **Avant**, **Arrière** et **Bloc**. La procédure de régression Multiple élabore une équation en une fois, selon les variables explicatives sélectionnées. Les autres procédures permettent l'entrée ou la sortie séquentielle des variables (ou blocs de variables). Plus précisément, la procédure en escalier permet l'entrée d'une variable à la fois. La variable suivante est celle la plus hautement corrélée à la partie inexpliquée de la variable réponse. L'option Escalier permet cependant aussi la sortie de variables entrées si leur contribution n'est plus significative. La procédure Avant est identique à l'escalier, si ce n'est que la sortie des variables entrées n'est pas admise. La procédure Arrière commence par considérer toutes les variables explicatives potentielles dans l'équation, puis les supprime une à une si leur contribution n'est pas significative. Enfin, la procédure Bloc permet aux blocs de variables explicatives d'entrer ou non, *en bloc*, selon un ordre spécifié. Si un bloc n'est pas significatif et n'entre pas, aucun bloc ultérieur n'est considéré à l'entrée.

Boîte de dialogue Régression

Les analyses se configurent dans la boîte de dialogue **Régression** :

I	D	Nom	Adresse
<input type="checkbox"/>	<input type="checkbox"/>	Mois	B20:B55
<input checked="" type="checkbox"/>	<input type="checkbox"/>	HrsMach	C20:C55
<input checked="" type="checkbox"/>	<input type="checkbox"/>	LotsProd	D20:D55
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Frais généraux	E20:E55

Une variable dépendante (D) et une ou plusieurs variables indépendantes (I) doivent être sélectionnées, sauf dans le cas de la régression de type Bloc. Pour la régression Bloc, une variable dépendante (D) et un à sept blocs (B1 à B7) doivent être sélectionnés. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Régression** :

- **Type de régression** désigne le type de régression à effectuer : Multiple, Escalier, Avant, Arrière ou Bloc.

Les **paramètres** à configurer varient suivant le type d'analyse sélectionné :

- **Valeurs p** affecte la manière dont les variables sont ajoutées à l'équation de régression ou en sont supprimées. Sous sélection de ce paramètre, on spécifie une **valeur p entrée** et/ou **valeur p sortie**, suivant la méthode de régression applicable. Plus la valeur p est faible, plus une variable doit être significative pour entrer dans l'équation de régression ou en sortir. Les valeurs proposées par défaut sont généralement acceptables. Deux choses à garder à l'esprit : D'abord, la valeur p d'entrée ne peut pas être supérieure à celle de sortie. Ensuite, la configuration de valeurs p *plus élevées* facilite l'entrée des variables (et en rend la sortie plus difficile), tandis que celle valeurs p *moins grandes* en rend l'entrée plus difficile (et la sortie plus facile). Les valeurs p sont généralement comprises entre 0,01 et 0,1.
- Comme le paramètre Valeurs p, **Valeurs F** affecte la manière dont les variables sont ajoutées à l'équation de régression ou en sont supprimées. Sous sélection de ce paramètre, on spécifie une **valeur F entrée** et/ou **valeur F sortie**, suivant la méthode de régression applicable. Les valeurs F sont généralement comprises entre 2,5 et 4.

StatTools propose différents diagrammes de dispersion facultatifs, tels que listés dans le volet **Graphiques** :

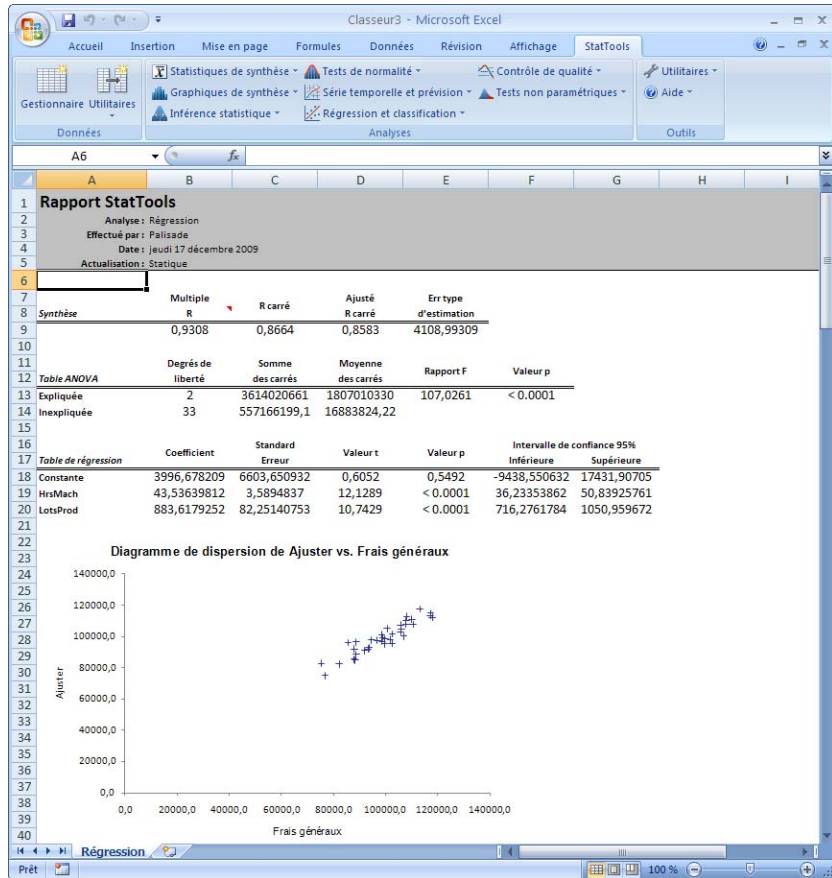
- **Valeurs ajustées vs valeurs Y réelles**
- **Valeurs ajustées vs valeurs X**
- **Résidus vs valeurs ajustées**
- **Résidus vs valeurs X**

Ces diagrammes servent généralement à l'« analyse résiduelle », pour vérifier la satisfaction ou non des hypothèses de régression. Le tracé le plus utile est probablement celui des résidus (sur l'axe vertical) par rapport aux valeurs ajustées (ou prédites) de la variable réponse.

Options avancées de l'analyse de régression :

- **Inclure l'info de pas détaillée** rapporte les statistiques R carré et d'erreur type de la variable indépendante à chaque pas intermédiaire de la régression.
- Sous **Inclure une prédiction**, les valeurs prédites pour la variable dépendante sont générées pour les valeurs variables indépendantes dans un deuxième ensemble de données. Cet **ensemble de données de prédiction** doit avoir les mêmes noms de variables que l'ensemble original soumis à l'analyse de régression. Dans l'ensemble de données de prédiction, il y a généralement des ensembles de valeurs de variables indépendantes pour lesquelles on désire **prédire** la valeur de la variable dépendante. L'équation de régression calculée sur la base du premier ensemble de données sert à effectuer les prédictions. Les valeurs prédites pour la variable dépendante s'introduisent directement dans l'ensemble de données de prédiction, où elles remplissent la colonne (ou la ligne) de la variable dépendante. Le **niveau de confiance** spécifie les limites inférieure et supérieure à générer pour les valeurs prédites.

Rapport de régression en escalier



Données manquantes et liaison aux données

- **Données manquantes** - S'il y a des données manquantes, toute ligne affectée pour l'une quelconque des variables sélectionnées est omise.
- **Liaison aux données** - Il n'y a pas de lien aux données originales. Si les données changent, l'analyse doit être réexécutée.

Commande Régression logistique

Effectue une analyse de régression logistique sur un ensemble de variables.

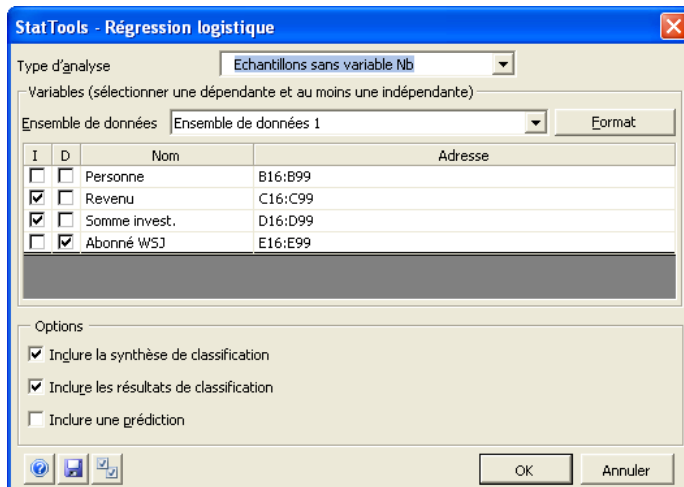
La commande Régression logistique effectue une analyse de régression logistique sur un ensemble de variables. Il s'agit essentiellement d'un type non linéaire d'analyse de régression où la variable réponse est binaire : 0 ou 1. Il doit y avoir une variable réponse 0-1 qui spécifie si chaque observation est un « succès » ou un « échec », ainsi qu'au moins une variable explicative pouvant servir à estimer la probabilité de succès.

Sous une seconde option de régression logistique, une variable « Nb » spécifie le nombre d'« essais » observés à chaque combinaison de variables explicatives. La variable réponse doit alors indiquer le nombre d'essais couronnés de « succès ». Le résultat de la régression logistique est une équation de régression similaire à l'équation de régression ordinaire multiple. Son interprétation est cependant légèrement différente, comme expliqué ci-dessous.

La procédure de régression logistique StatTools recherche l'équation de régression par optimisation. Cette optimisation repose sur un algorithme non linéaire complexe et la procédure peut être longue, suivant la vitesse du PC.

Boîte de dialogue Régression logistique

Cette analyse se configure dans la boîte de dialogue **Régression logistique** :



Une variable réponse ou dépendante (D) et une ou plusieurs variables indépendantes (I) doivent être sélectionnées pour l'analyse. Les données doivent

- 1) être au format « empilé » et il doit y avoir une variable réponse 0-1 spécifiant le « succès » ou l'« échec » de chaque observation. Il s'agit dans ce cas du type d'analyse **Échantillons sans variable Nb** ;
- 2) avoir une variable « Nb » et une variable réponse entière. Il s'agit alors du type d'analyse **Synthèse d'échantillons à variable Nb**.

Les variables peuvent provenir d'ensembles de données différents. Si le type d'analyse **Synthèse d'échantillons à variable Nb** est sélectionné, une variable Nb doit aussi être sélectionnée.

Dans la boîte de dialogue **Régression logistique** :

- **Type d'analyse** désigne le type de régression logistique à effectuer : **Échantillons sans variable Nb** ou **Synthèse d'échantillons à variable Nb**.

Le type **Échantillons sans variable Nb** comporte une variable réponse 0-1 qui spécifie si chaque observation est un « succès » ou un « échec », ainsi qu'au moins une variable explicative pouvant servir à estimer la probabilité de succès.

Sous le type **Synthèse d'échantillons à variable Nb**, une variable « Nb » spécifie le nombre d'« essais » observés à chaque combinaison de variables explicatives. La variable réponse doit alors indiquer le nombre d'essais couronnés de « succès ». Pour ce type d'analyse, une troisième colonne apparaît dans le sélecteur de variable, pour la sélection de la variable Nb.

- **Inclure la synthèse de classification** fait figurer la synthèse dans le rapport de régression.
- **Inclure les résultats de classification** fait figurer les résultats dans le rapport de régression.

- Sous **Inclure une prédiction**, les valeurs prédites pour la variable dépendante sont générées pour les valeurs variables indépendantes dans un deuxième ensemble de données. Cet **ensemble de données de prédiction** doit avoir les mêmes noms de variables que l'ensemble original soumis à l'analyse de régression. Dans l'ensemble de données de prédiction, il y a généralement des ensembles de valeurs de variables indépendantes pour lesquelles on désire **prédire** la valeur de la variable dépendante. L'équation de régression calculée sur la base du premier ensemble de données sert à effectuer les prédictions. Les valeurs prédites pour la variable dépendante s'introduisent directement dans l'ensemble de données de prédiction, où elles remplissent la colonne (ou la ligne) de la variable dépendante.

Rapport de régression logistique

Rapport StatTools

Analyse : Régression logistique
 Effectué par : Palisade
 Date : jeudi 17 décembre 2009
 Actualisation : Statique

Mesures de synthèse

Déviance nulle	105,4941869
Déviance du modèle	26,83613587
Amélioration	78,65805103
Valeur p	< 0,0001

Coefficients de régression

	Coefficient	Erreur type	Valeur Wald	Valeur p	Limite inférieure	Limite supérieure	Exp(Coeff)
Constante	-7,930438797	3,10005134	-2,558163697	0,0105	-14,00653942	-1,854338171	0,000359629
Revenu	-9,19596E-05	5,30141E-05	-1,734625165	0,0828	-0,000195867	1,1948E-05	0,999908045
Somme invest.	0,000351829	9,60677E-05	3,662306752	0,0002	0,000163537	0,000540122	1,000351891

Matrice de classification

	1	0	Pourcentage Correcte
1	23	4	85,19%
0	3	54	94,74%

Synthèse de classification

	Pourcentage
Correcte	91,67%
Base	67,86%
Amélioration	74,07%

Probabilités et classifications

	Revenu	Somme invest.	Probabilité	Classe d'analyse	Classe originale
1	66400	26900	1,02%	0	0
2	68000	7100	0,00%	0	0
3	54900	21500	0,44%	0	0
4	50600	19300	0,30%	0	0
5	54100	16700	0,09%	0	0
6	78200	31900	1,99%	0	0

Le rapport illustré ci-dessus inclut les données originales plus celles utilisées pour la classification. Les classifications prédites, dans la colonne E, varient suivant que les probabilités de « succès » estimées dans la colonne D sont supérieures ou inférieures à une valeur limite de 0,5 ou 50 %. Le rapport présente les statistiques de synthèse de la régression (un peu comme R carré pour la régression multiple), le détail de l'équation de régression et la synthèse des résultats de la procédure de classification. (Les commentaires de cellules aident à interpréter les résultats.) Dans cet exemple, 90,5 % des observations sont classifiées correctement. On espère bien sûr toujours atteindre un pourcentage aussi élevé que possible. En général, le seul moyen d'améliorer les résultats est d'accroître le nombre (ou la qualité) des variables explicatives. Les valeurs de la colonne H **Exp (Coef)** servent généralement à interpréter l'équation de régression. Ces valeurs indiquent le changement estimé de la probabilité de « succès » quand une variable explicative augmente d'une unité.

**Données
manquantes et
liaison aux
données**

- **Données manquantes** - S'il y a des données manquantes, toute ligne affectée pour *l'une quelconque* des variables sélectionnées est omise.
- **Liaison aux données** - Il n'y a pas de lien aux données originales. Si les données changent, l'analyse doit être réexécutée.

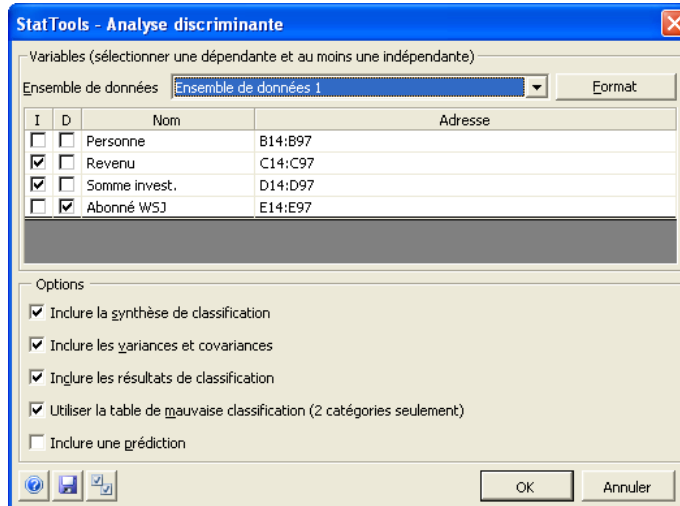
Commande Analyse discriminante

Effectue une analyse discriminante sur un ensemble de variables.

La commande Analyse discriminante effectue une analyse discriminante sur un ensemble de données. Dans cette analyse, une variable « catégorielle » spécifie, parmi deux groupes ou davantage, celui auquel une observation appartient. Une ou plusieurs variables explicatives peuvent aussi servir à prédire cette appartenance. La prédiction d'appartenance peut s'effectuer de deux manières. La première, plus générale et valable pour un nombre quelconque de groupes, consiste à calculer la « distance statistique » de chaque observation par rapport à la moyenne de chaque groupe et à classer l'observation suivant la distance statistique la plus faible. La seconde méthode, en présence de deux groupes, consiste à calculer une fonction discriminante (expression linéaire des variables explicatives) et à classer chaque observation suivant que sa valeur discriminante est inférieure ou supérieure à une valeur limite. Cette seconde méthode permet aussi de spécifier les probabilités d'appartenance antérieures, de même que les coûts de mauvaise classification. La procédure de classification revient alors à minimiser le coût probable de la mauvaise classification.

**Boîte de dialogue
Analyse
discriminante**

Cette analyse se configure dans la boîte de dialogue **Analyse discriminante** :



Une variable dépendante (D) et une ou plusieurs variables indépendantes (I) doivent être sélectionnées pour l'analyse. Les données doivent être au format « désempilé ». Les variables peuvent provenir d'ensembles de données différents.

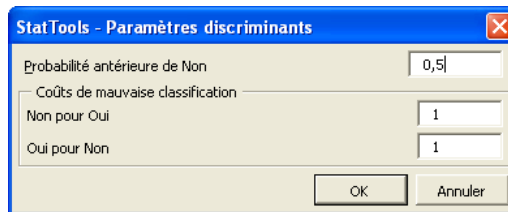
Dans la boîte de dialogue **Analyse discriminante** :

- **Inclure la synthèse de classification** fait figurer la synthèse dans le rapport de régression.
- **Inclure les variances et covariances** fait figurer celles-ci dans le rapport de régression.
- **Inclure les résultats de classification** fait figurer les résultats dans le rapport de régression.
- **Utiliser la table de mauvaise classification** permet de changer les probabilités antérieures ou les coûts de mauvaise classification.

- Sous **Inclure une prédiction**, les valeurs prédites pour la variable dépendante sont générées pour les valeurs variables indépendantes dans un deuxième ensemble de données. Cet **ensemble de données de prédiction** doit avoir les mêmes noms de variables que l'ensemble original soumis à l'analyse de régression. Dans l'ensemble de données de prédiction, il y a généralement des ensembles de valeurs de variables indépendantes pour lesquelles on désire **prédire** la valeur de la variable dépendante. L'équation de régression calculée sur la base du premier ensemble de données sert à effectuer les prédictions. Les valeurs prédites pour la variable dépendante s'introduisent directement dans l'ensemble de données de prédiction, où elles remplissent la colonne (ou la ligne) de la variable dépendante.

**Boîte de dialogue
des coûts de
mauvaise
classification**

En présence exacte de deux groupes possibles pour la variable catégorielle dépendante (comme dans l'exemple présenté ici), sous sélection de l'option **Utiliser la table de mauvaise classification**, une boîte de dialogue s'ouvre pour la spécification des probabilités antérieures et/ou des coûts de mauvaise classification. Sous les paramètres par défaut, chaque groupe présente une probabilité égale et les coûts de mauvaise classification sont égaux. Rien n'empêche cependant de changer ces paramètres.



Rapport d'analyse discriminante

Classeur3 - Microsoft Excel

Statistiques de synthèse Tests de normalité Contrôle de qualité
 Graphiques de synthèse Série temporelle et prévision Tests non paramétriques
 Inférence statistique Régression et classification

Données A6

Rapport StatTools

Analyse : Discriminante
 Effectué par : Palisade
 Date : jeudi 17 décembre 2009
 Actualisation : Statique

Variables de configuration	Antérieure Probabilité	Coûts de mauvaise classification	
		Non	Oui
Non	0,5	0	1
Oui	0,5	1	0

Synthèse des échantillons	Echantillon	Moyenne	Moyenne
	Taille	Revenu	Somme invest.
Non	57	66042,10526	24952,63158
Oui	27	80485,18519	53000

Fonction discriminante	Coefficient
Revenu	6,586E-05
Somme invest.	-0,000352735

Matrice de classification	Non	Oui	Correcte
Non	52	5	91,2%
Oui	2	25	92,6%

Synthèse de classification	
Correcte	91,7%
Base	67,9%
Amélioration	74,1%

Matrice des var. et covar.	Revenu	Somme invest.
Non		
Revenu	209308552,6	79303458,65
Somme invest.	79303458,65	99485394,74
Oui		
Revenu	118875156,7	64698846,15
Somme invest.	64698846,15	80470769,23
Combinées		
Revenu	180634549	74672727,86
Somme invest.	74672727,86	93456367,14

Distance statistique	Revenu	Somme invest.	Distance	Distance	Discrim	Originale Classe	Analyse Classe
			Non	Oui			
1	66400	26900	0,229005023	2,769086119	-5,115462701	Non	Non
2	68000	7100	2,363424601	5,232611294	1,974062536	Non	Non
3	54900	21500	0,84174592	3,258632671	-3,968084673	Non	Non

Discriminante

Le rapport d'analyse discriminante inclut les données originales plus celles utilisées pour la classification. Les classifications prédites varient suivant que les valeurs discriminantes indiquées sont inférieures ou supérieures à une valeur limite. Si les probabilités antérieures et les coûts de mauvaise classification sont maintenus à leurs valeurs par défaut, cette procédure de classification revient à baser la classification sur la plus faible des deux distances statistiques. De fait, s'il y avait plus que deux groupes, les valeurs discriminantes n'apparaîtraient pas et la classification reposerait sur la plus faible des distances statistiques.

Le rapport indique aussi les statistiques descriptives des groupes et les coefficients de la fonction discriminante (en présence de deux groupes seulement), les probabilités antérieures, les coûts de mauvaise classification et la valeur limite de mauvaise classification (ici aussi en présence de deux groupes seulement), ainsi que la synthèse des résultats de la procédure de classification (avec commentaires de cellule utiles à l'interprétation). Dans cet exemple, 89% des observations sont classifiées correctement. On espère bien sûr toujours atteindre un pourcentage aussi élevé que possible. En général, le seul moyen d'améliorer les résultats est d'accroître le nombre (ou la qualité) des variables explicatives.

**Données
manquantes et
liaison aux
données**

- **Données manquantes** - S'il y a des données manquantes, toute ligne affectée pour *l'une quelconque* des variables sélectionnées est omise.
- **Liaison aux données** - Il n'y a pas de lien aux données originales. Si les données changent, l'analyse doit être réexécutée.

Menu Contrôle de qualité

Les procédures proposées dans le menu Contrôle de qualité concernent l'analyse de données collectées au fil du temps, applicable aux situations de contrôle de qualité.

Le diagramme de Pareto affiche l'importance relative de données classées par catégories.

Les quatre types de graphiques de contrôle représentent des données de série temporelle et permettent de voir si un processus est conforme à ses limites de contrôle statistique. Outre le maintien des données dans les limites de contrôle, on peut y vérifier d'autres comportements non aléatoires tels que les longues séquences par-dessus ou par-dessous la ligne centrale.

Commande Diagramme de Pareto

Crée un diagramme de Pareto pour une variable classée par catégories.

Les diagrammes de Pareto sont utiles à la détermination des éléments les plus significatifs d'un groupe de données classées par catégories, de même qu'à la représentation visuelle rapide de leur importance relative. Ces diagrammes s'utilisent généralement dans le domaine de l'assurance de la qualité, où ils aident à identifier les quelques facteurs les plus significatifs (règle du 80/20 de Pareto).

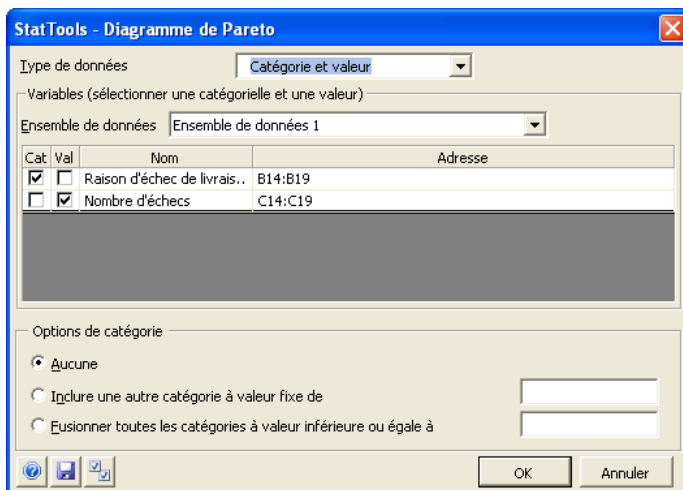
Imaginons, par exemple, un fabricant de pièces de machine cherchant à déterminer pourquoi ses clients rejettent un produit particulier. Au retour de chaque lot, il saisit une raison (« format incorrect », « fini incorrect », etc.) Au bout de quelques mois de collecte de données, un diagramme de Pareto est tracé. Les mesures nécessaires peuvent ainsi être prises pour résoudre les causes principales de problèmes.

StatTools permet la création de diagrammes de Pareto en fonction de données présentées sous deux formats : Catégorie seule ou Catégorie et valeur. Sous le type Catégorie seule, une variable contient généralement une entrée par observation. Dans l'exemple considéré ici, chaque cellule correspondrait à la raison de retour d'un lot de pièces. Une valeur de cellule pourrait être « fini incorrect » et il y aurait probablement de nombreuses cellules redoublées. StatTools compte le nombre d'apparition de chaque entrée dans la variable et trace le diagramme de Pareto correspondant. Sous l'option Catégorie et valeur, les variables spécifiées représentent les catégories et chaque nombre correspondant.

Les axes du diagramme de Pareto se construisent comme suit :

- Les catégories se placent le long de l'axe horizontal.
- La fréquence (ou le nombre) se place le long de l'axe vertical de gauche.
- Le pourcentage cumulatif se place le long de l'axe vertical de droite.

Boîte de dialogue Diagramme de Pareto

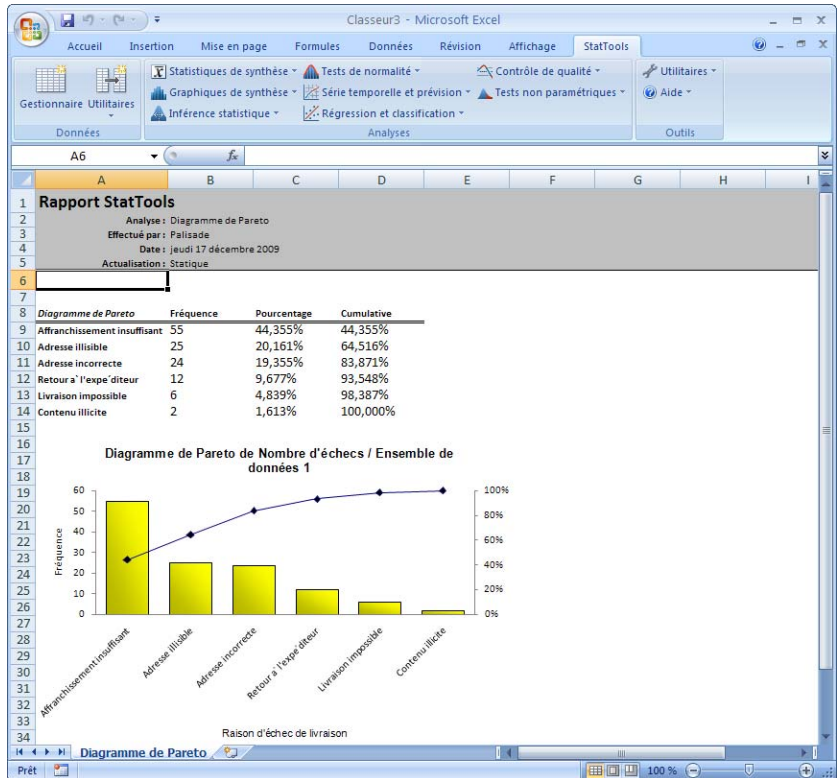


Dans la boîte de dialogue **Diagramme de Pareto** :

- **Type de données** désigne le type de données appelées à construire le diagramme de Pareto : Catégorie et valeur ou Catégorie seule.

Les **options de catégorie** suivantes sont proposées :

- **Aucune** – chaque catégorie distincte est représentée par une barre dans le diagramme de Pareto.
- **Inclure une autre catégorie à valeur fixe de** – une barre intitulée « Divers » s’ajoute à l’extrême droite du diagramme, à fréquence égale à la valeur spécifiée.
- **Fusionner toutes les catégories à valeur inférieure ou égale à** – toutes les catégories dont la fréquence est inférieure ou égale à la valeur spécifiée se combinent en une catégorie intitulée « Divers », à l’extrême droite du diagramme de Pareto.



Données manquantes et liaison aux données

- **Données manquantes** – S'il y a des données manquantes, toute ligne affectée pour l'une quelconque des variables sélectionnées est omise.
- **Liaison aux données** – Il n'y a pas de lien aux données originales. Si les données changent, l'analyse doit être réexécutée.

Commande Graphiques X/R

Trace des graphiques de contrôle X/R pour des variables de série temporelle.

Cette analyse produit des graphiques X barre/R pour des données de série temporelle. Les données doivent avoir été collectées en petits sous-échantillons, au fil du temps. Par exemple, un opérateur pourrait collecter les mesures de largeur de quatre pièces sélectionnées aléatoirement toutes les demi-heures. La taille du sous-échantillon serait donc 4. Si les données sont collectées sur des périodes de 50 demi-heures, elles peuvent être organisées en quatre colonnes et 50 lignes adjacentes, sous en-têtes de variable Sous-éch1 à Sous-éch4 par-dessus la première ligne de données.

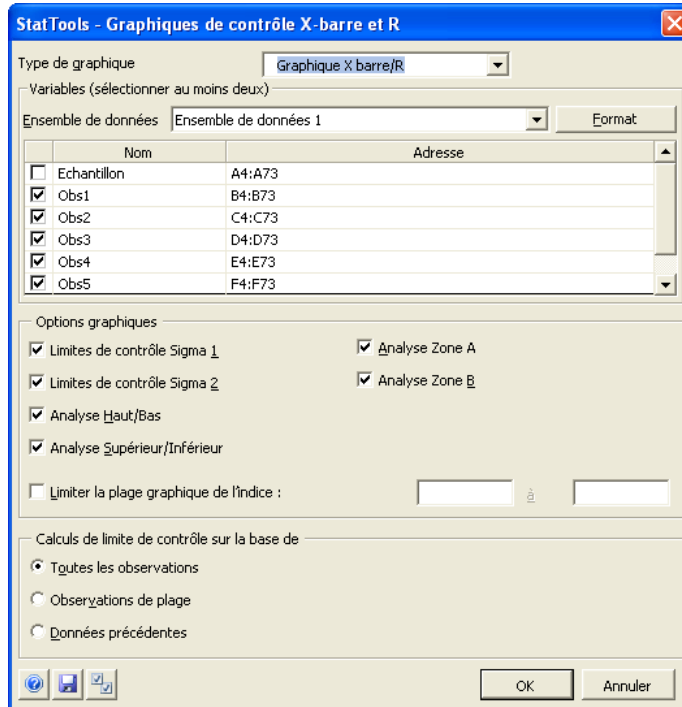
Le but de la procédure est de vérifier si le processus générateur des données est conforme au contrôle statistique. Pour ce faire, la procédure calcule d'abord les valeurs X barre et R de chaque ligne de l'ensemble de données. X barre représente la moyenne des observations de la ligne et R, l'étendue (le maximum moins le minimum) des observations de cette ligne.

Les points X barre et R se tracent sur des chronogrammes différents, de part et d'autre de la ligne centrale. Pour le graphique X barre, cette ligne représente la moyenne des valeurs X barre (parfois appelées X double barre) et pour le tracé R, il s'agit de la barre R, moyenne des valeurs R. Pour déterminer la conformité du processus, il suffit parfois de voir si l'une des valeurs X barre ou R dépasse sa limite de contrôle supérieure ou inférieure respective (LCS et LCI), équivalente à plus ou moins 3 écarts types, environ, de la ligne centrale. Les graphiques affichent ces limites de contrôle et facilitent ainsi le repérage des valeurs extrêmes.

Cette procédure permet aussi d'identifier d'autres comportements hors-contrôle éventuels : notamment, au moins 8 points consécutifs par-dessus ou par-dessous la ligne centrale ; au moins 8 points consécutifs à la hausse ou à la baisse ; au moins 4 points sur 5 consécutifs au-delà d'un écart type de la ligne centrale ; et au moins 2 points sur 3 consécutifs au-delà de deux écarts types de la ligne centrale.

**Boîte de dialogue
Graphiques de
contrôle X-barre
et R**

Ces graphiques se configurent dans la boîte de dialogue **Graphiques de contrôle X-barre et R** :



Deux variables ou plus peuvent être sélectionnées pour l'analyse. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

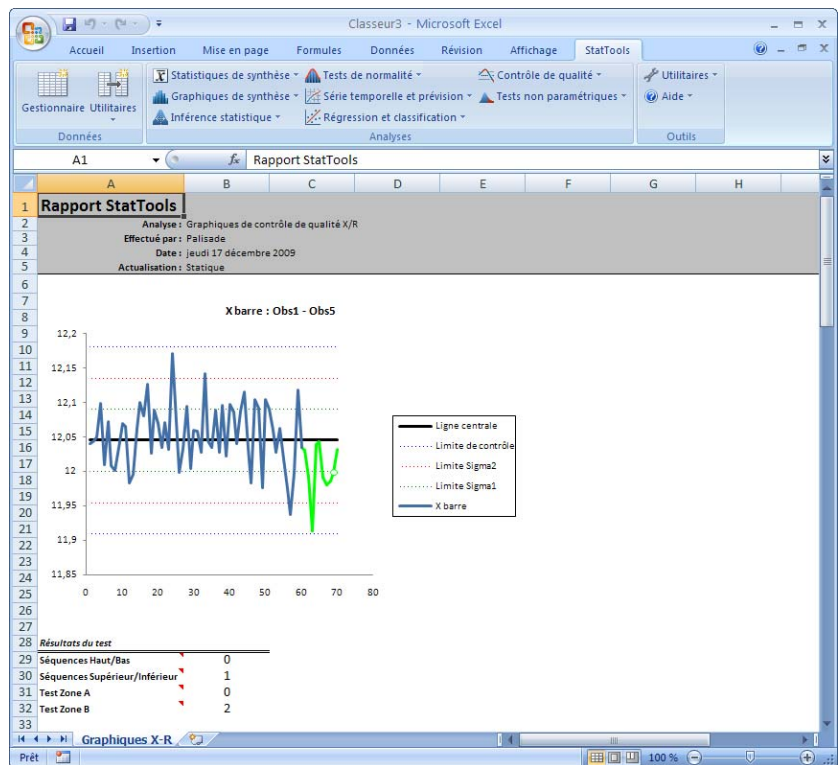
Options graphiques :

- **Limites de contrôle Sigma 1 et 2** ajoutent des lignes de limite de contrôle à un et/ou deux sigma de la ligne moyenne. Ces lignes supplémentaires permettent d'identifier d'autres types de comportements hors-contrôle (règles de « zone »).
- **Analyse Zone A** et **Analyse Zone B** indiquent le nombre de points au-delà de Zone A (2 sigma) et de Zone B (1 sigma).
- **Analyse Haut/Bas** et **Analyse Supérieur/Inférieur** identifient les mouvements à la hausse ou à la baisse de 8 points ou plus.
- **Limiter la plage graphique de l'indice** limite les points du graphique à une plage de points de donnée pour une variable (de l'indice de départ à l'indice de fin).

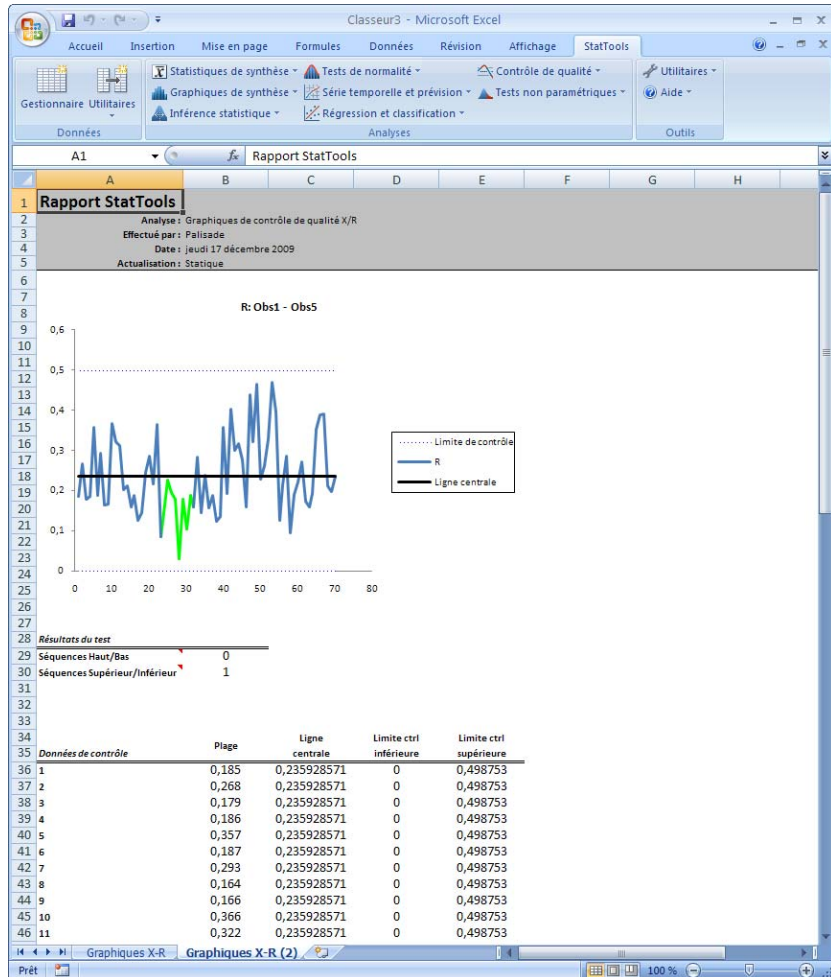
Les options du volet **Calculs de limite de contrôle sur la base de** de la boîte de dialogue Graphiques de contrôle X-barre et R déterminent les données sur lesquelles doivent reposer les calculs des limites de contrôle :

- **Toutes les observations** fait reposer les calculs de limite de contrôle sur toutes les données disponibles.
- **Observations de plage** limite les calculs aux données comprises entre l'**indice de départ** et l'**indice d'arrêt**.
- **Données précédentes** détermine les limites de contrôle en fonction des données observées précédemment. On entre alors la taille de **Sous-échantillon** et les valeurs **R moyen** et **X barre moyen** calculées pour les données antérieures.

Exemple de graphique X barre



Exemple de graphique R



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes ne sont pas admises.
- **Liaison aux données** – Les graphiques ne sont pas liés aux données.

Commande Graphique P

Trace des graphiques de contrôle P pour des variables de série temporelle.

Les graphiques P concernent les données d'« attribut ». Chaque observation indique dans ce cas le nombre (ou la fraction) d'éléments non conformes aux spécifications pour un échantillon de ces éléments. Par exemple, un processus peut produire un certain nombre d'éléments par demi-heure et certains de ces éléments sont non conformes. Un graphique P tracerait pour chaque demi-heure la fraction d'éléments non conformes. Ici encore, le but est de déterminer si le processus est conforme aux critères de contrôle.

Pour l'exécution de cette procédure, il faut un ensemble de données contenant au moins l'un des éléments suivants : une variable représentant le *nombre* d'éléments non conformes de chaque échantillon ou une variable représentant la *fraction* d'éléments non conformes de chaque échantillon. Facultativement, on peut avoir aussi une variable représentant les tailles d'échantillon. En l'absence de variable de taille d'échantillon, on entre une taille présumée commune à tous les échantillons. En présence d'une variable de taille, toutefois, les tailles d'échantillon ne doivent pas être égales.

**Boîte de dialogue
Graphiques de
contrôle P**

Ce type de graphique se configure dans la boîte de dialogue **Graphiques de contrôle P** :

Val	Nom	Adresse
<input type="checkbox"/>	Heure	A3:A27
<input type="checkbox"/>	Taille échant.	B3:B27
<input checked="" type="checkbox"/>	Non conforme	C3:C27

L'analyse requiert la sélection d'une variable de valeur et, facultativement, d'une variable de taille. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Options **Données en entrée** :

- **Nombres d'éléments non conformes** spécifie que la variable Val(eur) indique le nombre effectif d'éléments non conformes de l'échantillon total.
- **Fractions d'éléments non conformes** spécifie que la variable Val(eur) indique la fraction d'éléments non conformes de l'échantillon.

Options **Taille** :

- **Variable taille** spécifie l'emploi d'une variable de taille indiquant la taille totale de chaque échantillon.
- **Taille commune** spécifie, en l'absence de variable de taille, la taille commune à tous les échantillons.

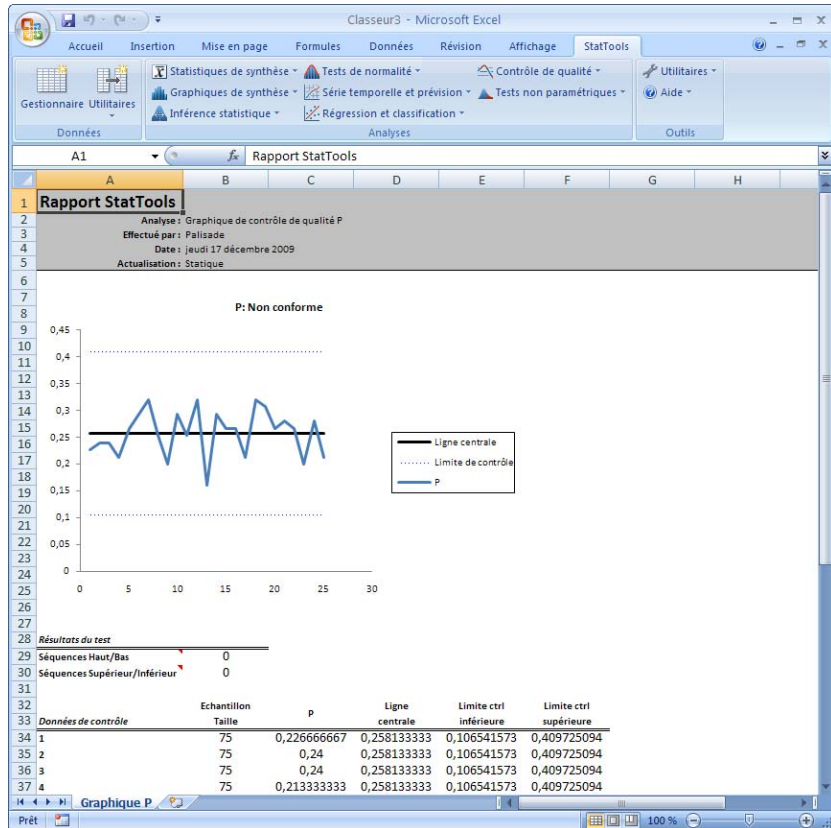
Options **graphiques** :

- **Limites de contrôle Sigma 1 et Sigma 2** ajoutent des lignes de limite de contrôle à un et/ou deux sigma de la ligne moyenne. Ces lignes supplémentaires permettent d'identifier d'autres types de comportements hors-contrôle (règles de « zone »).
- **Analyse Zone A et Analyse Zone B** indiquent le nombre de points au-delà de Zone A (2 sigma) et de Zone B (1 sigma).
- **Analyse Haut/Bas et Analyse Supérieur/Inférieur** identifient les mouvements à la hausse ou à la baisse de 8 points ou plus.
- **Limiter la plage graphique de l'indice** limite les points du graphique à une plage de points de donnée pour une variable (de l'indice de départ à l'indice de fin).

Les options du volet **Calculs de limite de contrôle sur la base de** de la boîte de dialogue **Graphiques de contrôle P** déterminent les données sur lesquelles doivent reposer les calculs des limites de contrôle :

- **Toutes les observations** fait reposer les calculs de limite de contrôle sur toutes les données disponibles.
- **Observations de plage** limite les calculs aux données comprises entre l'**indice de départ** et l'**indice d'arrêt**.
- **Données précédentes** détermine les limites de contrôle en fonction des données observées précédemment. On entre alors la taille de **Sous-échantillon** et la valeur **P moyen** calculée pour les données antérieures.

Exemple de graphique P



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes ne sont pas admises.
- **Liaison aux données** – Les graphiques ne sont pas liés aux données.

Commande Graphique C

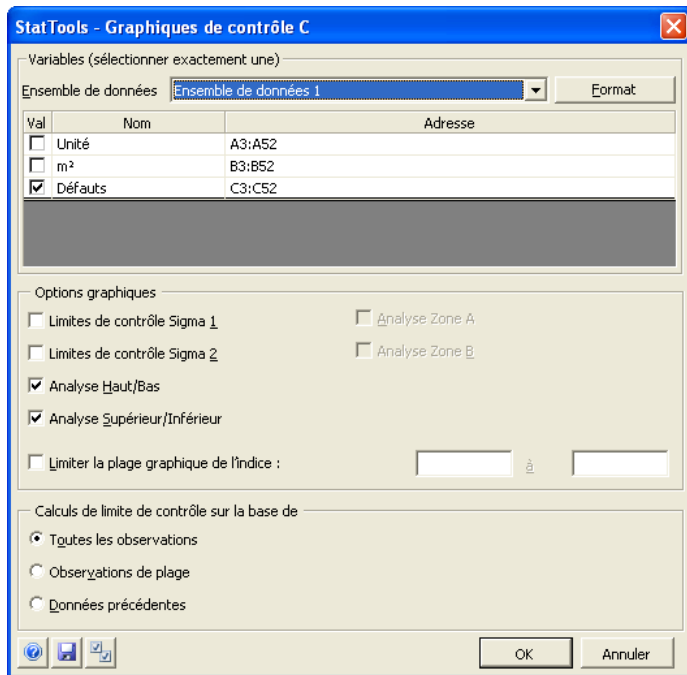
Trace des graphiques de contrôle C pour des variables de série temporelle.

Les graphiques de contrôle C servent à tracer le nombre de défauts d'éléments de taille constante. Dans le cas de portières de voiture produites par lots de 50, par exemple, on pourrait compter le nombre de défauts (de peinture ou autres) relevés par lot. Ces nombres se tracent ensuite sur un graphique. Ici encore, le but est de déterminer si le processus est conforme aux critères de contrôle.

L'ensemble de données nécessaire à la production d'un graphique de contrôle C doit inclure une variable représentant le nombre de défauts par élément. La taille de l'élément est supposée égale pour chaque observation. Ainsi, si l'« élément » considéré est un lot de portières, on suppose que chaque lot compte un nombre identique de portières.

Boîte de dialogue Graphiques de contrôle C

Ce type de graphique se configure dans la boîte de dialogue
Graphiques de contrôle C :



L'analyse requiert la sélection d'au moins une variable. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

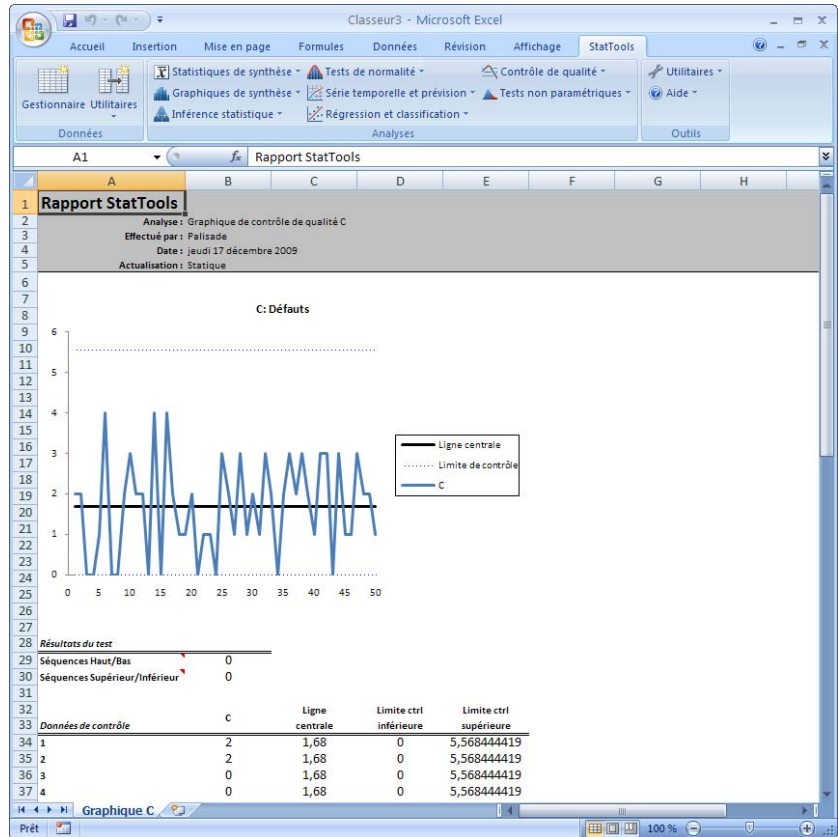
Options graphiques :

- **Limites de contrôle Sigma 1 et Sigma 2** ajoutent des lignes de limite de contrôle à un et/ou deux sigma de la ligne moyenne. Ces lignes supplémentaires permettent d'identifier d'autres types de comportements hors-contrôle (règles de « zone »).
- **Analyse Zone A et Analyse Zone B** indiquent le nombre de points au-delà de Zone A (2 sigma) et de Zone B (1 sigma).
- **Analyse Haut/Bas et Analyse Supérieur/Inférieur** identifient les mouvements à la hausse ou à la baisse de 8 points ou plus.
- **Limiter la plage graphique de l'indice** limite les points du graphique à une plage de points de donnée pour une variable (de l'indice de départ à l'indice de fin).

Les options du volet **Calculs de limite de contrôle sur la base de** de la boîte de dialogue Graphiques de contrôle C déterminent les données sur lesquelles doivent reposer les calculs des limites de contrôle :

- **Toutes les observations** fait reposer les calculs de limite de contrôle sur toutes les données disponibles.
- **Observations de plage** limite les calculs aux données comprises entre l'**indice de départ** et l'**indice d'arrêt**.
- **Données précédentes** détermine les limites de contrôle en fonction des données observées précédemment. On entre alors la valeur **C moyen** calculée pour les données antérieures.

Exemple de graphique C



Données manquantes et liaison aux données

- **Données manquantes** - Les données manquantes ne sont pas admises.
- **Liaison aux données** - Les graphiques ne sont pas liés aux données.

Commande de graphique U

Trace des graphiques de contrôle U pour des variables de série temporelle.

Les graphiques de contrôle U sont similaires aux graphiques C, si ce n'est qu'ils représentent les taux de défauts. Concernant notre exemple de portières de voiture, supposons maintenant que les tailles de lot ne soient pas nécessairement égales et que différents lots puissent donc compter différents nombres de portières. Dans un graphique de contrôle U, on tracerait le taux de défauts par portière, soit le nombre de défauts comptés dans un lot divisé par celui de portières comprises dans le lot. Ici encore, le but est de déterminer si le processus est conforme aux critères de contrôle.

Pour l'exécution de cette procédure, il faut un ensemble de données contenant au moins l'un des éléments suivants : une variable représentant le *nombre* de défauts par observation ou une variable représentant le *taux* de défauts par observation. Facultativement, on peut avoir aussi une variable représentant les tailles d'élément. En l'absence de variable de taille d'élément, on entre une taille présumée commune à toutes les observations. En présence d'une variable de taille, toutefois, les tailles d'élément ne doivent pas être égales.

**Boîte de dialogue
Graphiques de
contrôle U**

Ce type de graphique se configure dans la boîte de dialogue **Graphiques de contrôle U** :

Tail	Val	Nom	Adresse
<input type="checkbox"/>	<input type="checkbox"/>	Unité	A3:A52
<input checked="" type="checkbox"/>	<input type="checkbox"/>	m ²	B3:B52
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Défauts	C3:C52
<input type="checkbox"/>	<input type="checkbox"/>	Taux	D3:D52

L'analyse requiert la sélection d'une variable de valeur et, facultativement, d'une variable de taille. L'ensemble de données sélectionné doit être au format désempilé. Les variables peuvent provenir d'ensembles de données différents.

Options **Données en entrée** :

- **Nombres de défauts** spécifie que la variable Val(eur) indique le nombre effectif d'éléments non conformes de l'échantillon total.
- **Taux de défauts** spécifie que la variable Val(eur) indique la fraction d'éléments non conformes de l'échantillon.

Options **Taille d'échantillon** :

- **Variable taille** spécifie l'emploi d'une variable de taille indiquant la taille totale de chaque échantillon.
- **Taille commune** spécifie, en l'absence de variable de taille, la taille commune à tous les échantillons.

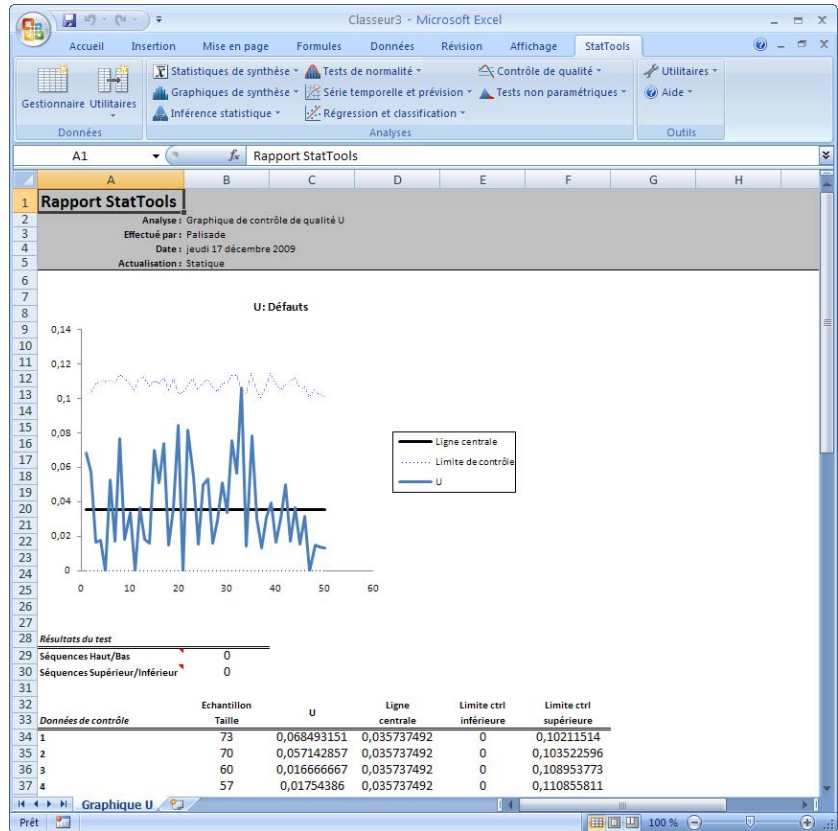
Options **graphiques** :

- **Limites de contrôle Sigma 1 et Sigma 2** ajoutent des lignes de limite de contrôle à un et/ou deux sigma de la ligne moyenne. Ces lignes supplémentaires permettent d'identifier d'autres types de comportements hors-contrôle (règles de « zone »).
- **Analyse Zone A et Analyse Zone B** indiquent le nombre de points au-delà de Zone A (2 sigma) et de Zone B (1 sigma).
- **Analyse Haut/Bas et Analyse Supérieur/Inférieur** identifient les mouvements à la hausse ou à la baisse de 8 points ou plus.
- **Limiter la plage graphique de l'indice** limite les points du graphique à une plage de points de donnée pour une variable (de l'indice de départ à l'indice de fin).

Les options du volet **Calculs de limite de contrôle sur la base de** de la boîte de dialogue Graphiques de contrôle U déterminent les données sur lesquelles doivent reposer les calculs des limites de contrôle :

- **Toutes les observations** fait reposer les calculs de limite de contrôle sur toutes les données disponibles.
- **Observations de plage** limite les calculs aux données comprises entre l'**indice de départ** et l'**indice d'arrêt**.
- **Données précédentes** détermine les limites de contrôle en fonction des données observées précédemment. On entre alors la taille de **Sous-échantillon** et la valeur **U moyen** calculée pour les données antérieures.

Exemple de graphique U



Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes ne sont pas admises.
- **Liaison aux données** – Les graphiques ne sont pas liés aux données.

Menu Tests non paramétriques

Les tests « non paramétriques » sont des procédures statistiques appliquées à des échantillons de données pour tester les hypothèses de distributions de probabilités sous-jacentes. Les tests d'hypothèse « paramétriques » sont mieux connus et plus généralement utilisés. L'approche non paramétrique offre cependant des avantages qui en font une solution mieux adaptée à de nombreuses situations.

Les tests d'hypothèse paramétriques concernent le type de distribution sous-jacente (généralement normale) et estiment les paramètres de ce type de distribution (généralement la moyenne et l'écart type). Dans de nombreuses applications, l'hypothèse de normalité serait incorrecte. Par exemple, le nombre d'appels par heure à un centre de service à la clientèle et le temps d'attente à la caisse d'un supermarché ne sont pas normalement distribués. Les tests non paramétriques ne présument pas le type de distribution sous-jacente. Certains reposent sur certaines suppositions générales quant à la forme de la distribution : dans les tests proposés ici, celui de rang de Wilcoxon présume une distribution symétrique. Pour les deux autres (Test des signes et Test de Mann-Whitney), aucune supposition n'est faite quant à la forme de la distribution.

En présence d'échantillons de taille faible, l'approche non paramétrique convient souvent mieux. Un échantillon de grande taille peut être soumis à un test de normalité. Si la présomption de distribution normale se confirme, un test paramétrique peut être effectué. Pour les faibles tailles d'échantillon, les tests de normalité sont cependant impuissants à distinguer les distributions normales des autres. Les tests non paramétriques sont alors utiles.

Données ordinales

Certains tests non paramétriques conviennent du reste à certains types de données sinon imperméables aux tests paramétriques : les **données ordinales**, notamment, où les observations sont décrites en termes de nombres exprimant des rangs de classement ; la différence entre ces nombres ne présente sinon aucune utilité. Ainsi les niveaux d'études peuvent être codés 0 (primaire), 1 (1^{er} cycle secondaire), 2 (2^e cycle secondaire), 3 (1^{er} cycle supérieur), 4 (2^e cycle supérieur) et 5 (3^e cycle supérieur). Cette échelle n'implique nullement que la différence de niveau d'études entre « 2^e cycle secondaire » et « primaire » est équivalente à la différence entre les catégories « 3^e cycle supérieur » et « 1^{er} cycle supérieur », même si, dans les deux cas, la différence entre les rangs est égale à 2. Les options Test des signes (type Un échantillon) et Test de Mann-Whitney proposées ici s'appliquent à ce type de données.

En résumé

En résumé, les tests non paramétriques s'appliquent aux situations suivantes, où l'approche paramétrique ne convient pas :

- quand l'information disponible ne suffit pas à déterminer la distribution de probabilités sous-jacente ;
- quand la taille d'échantillon est trop faible pour tester de manière fiable l'hypothèse de normalité ;
- quand les données sont de type ordinal.

Commande Test des signes

Effectue un test des signes sur les variables.

La commande Test des signes effectue des tests d'hypothèse pour la médiane d'une simple variable (type d'analyse **Un échantillon**) ou la médiane des différences pour une paire de variables (type d'analyse **Échantillons appariés**). Le test ne repose sur aucune supposition quant à la forme de la distribution (concernant la forme normale en particulier). Le type d'analyse à un échantillon convient aux **données ordinales**, comme décrit plus haut.

Boîte de dialogue Test des signes

Cette analyse se configure dans la boîte de dialogue **Test des signes** :

Variables (sélectionner exactement deux)	
Ensemble de données: Notes d'examen	
Nom	Adresse
<input checked="" type="checkbox"/> 2002	B19:B42
<input checked="" type="checkbox"/> 2003	C19:C42

Médiane des différences

Valeur d'hypothèse nulle: 0

Type d'hypothèse rivale: Inférieure à la valeur nulle (test unilatéral)

Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à un échantillon requiert une ou plusieurs variables ; celle à échantillons appariés en requiert deux. Pour l'analyse à un échantillon, les variables sélectionnées peuvent être des données empilées ou désempilées ; pour l'analyse à échantillons appariés, elles doivent être désempilées. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Test des signes** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - Le type **Un échantillon** effectue des tests d'hypothèse pour une simple variable numérique.
 - **Échantillons appariés** convient quand deux variables sont naturellement appariées. Il s'agit essentiellement d'une analyse à un échantillon sur la différence entre les paires.
- **Médiane (ou Médiane des différences)**.
 - **Valeur d'hypothèse nulle** désigne la valeur du paramètre de population sous l'hypothèse nulle.
 - **Type d'hypothèse rivale** désigne la rivale de la valeur d'hypothèse nulle à évaluer pendant l'analyse. Le type d'hypothèse rivale peut être « **unilatéral** » (valeur supérieure ou inférieure à l'hypothèse nulle) ou « **bilatéral** » (valeur non égale à l'hypothèse nulle).

Rapport de test des signes

Rapport StatTools	
Analyse : Test des signes	
Effectué par : Palisade	
Date : jeudi 17 décembre 2009	
Actualisation : En direct	
Test des signes (échantillons appariés) 2002 - 2003	
Statistiques d'échantillonnage	
Taille d'échantillon	24
Moyenne	-6,358333333
Ecart type	13,11745988
Médiane	-5,75
Hypothèses	
Médiane hypothétique [MH]	0
Hypothèse rivale	< 0
Ajustement de taille d'échantillon	
Nombre de valeurs = MH	0
Nbre de valeurs < ou = MH (taille ajustée)	24
Nombres de valeurs	
Nbre de valeurs < MH	17
Nbre de valeurs > MH (statistique du test)	7
Valeur p et niveaux de signification	
Valeur p	0,0320
Hypoth. nulle à signification 10%	Rejeter
Hypoth. nulle à signification 5%	Rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter

***Données
manquantes et
liaison aux
données***

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Commande Test de rang de Wilcoxon

Effectue un test de rang de Wilcoxon sur les variables.

La commande Test de rang de Wilcoxon effectue des tests d'hypothèse pour la médiane d'une simple variable (type d'analyse **Un échantillon**) ou la médiane des différences pour une paire de variables (type d'analyse **Échantillons appariés**). Ce test suppose que la distribution des probabilités est symétrique (sans présomption de normalité toutefois).

Cette analyse se configure dans la boîte de dialogue **Test de rang de Wilcoxon** :

**Boîte de dialogue
Test de rang de
Wilcoxon**

StatTools - Tests non paramétriques - Test de rang de Wilcoxon

Type d'analyse: Un échantillon

Variables (sélectionner au moins une)

Ensemble de données: Rang de Wilcoxon

Nom	Adresse
<input checked="" type="checkbox"/> Température	B20:B43

Médiane

Valeur d'hypothèse nulle: 0

Type d'hypothèse rivale: Non égale à la valeur nulle (test bilatéral)

Correction des égalités dans le calcul d'approximation normale (recommandé)

OK Annuler

Le nombre de variables sélectionnées dépend du type d'analyse. L'analyse à un échantillon requiert une ou plusieurs variables ; celle à échantillons appariés en requiert deux. Pour l'analyse à un échantillon, les variables sélectionnées peuvent être des données empilées ou désempilées ; pour l'analyse à échantillons appariés, elles doivent être désempilées. Les variables peuvent provenir d'ensembles de données différents.

Dans la boîte de dialogue **Test de rang de Wilcoxon** :

- **Type d'analyse** désigne le type d'analyse à effectuer. Les options suivantes sont proposées :
 - Le type **Un échantillon** effectue des tests d'hypothèse pour une simple variable numérique.
 - **Échantillons appariés** convient quand deux variables sont naturellement appariées. Il s'agit essentiellement d'une analyse à un échantillon sur la différence entre les paires.
- **Médiane (ou Médiane des différences)**.
 - **Valeur d'hypothèse nulle** désigne la valeur du paramètre de population sous l'hypothèse nulle.
 - **Type d'hypothèse rivale** désigne la rivale de la valeur d'hypothèse nulle à évaluer pendant l'analyse. Le type d'hypothèse rivale peut être « **unilatéral** » (valeur supérieure ou inférieure à l'hypothèse nulle) ou « **bilatéral** » (valeur non égale à l'hypothèse nulle).
- L'option **Correction des égalités** est recommandée pour corriger les rangs ex-æquo du test sous approximation normale seulement. La correction implique le dénombrement d'éléments dans les groupes de rangs ex-æquo et la réduction correspondante de la variance. La correction accroît toujours la valeur de la statistique z en présence de rangs ex-æquo. **(Remarque : La correction des égalités ne cause aucun changement de variance en l'absence de rangs ex-æquo.)**

Rapport de test de rang de Wilcoxon

Classeur3 - Microsoft Excel

StatTools

Rapport StatTools

Analyse : Test de rang de Wilcoxon.
Effectué par : Palisade
Date : jeudi 17 décembre 2009
Actualisation : En direct

Test de rang (un échantillon)	
Statistiques d'échantillonnage	Température
	Rang de Wilcoxon
Taille d'échantillon	24
Moyenne d'échantillon	-6,34
Ecart type d'échantillon	13,14
Médiane d'échantillon	-5,75
Hypothèses	
Médiane hypothétique (MH)	0
Hypothèse rivale	<> 0
Ajustement de taille d'échantillon	
Nombre de valeurs = MH	0
Nbre de valeurs < ou > MH (taille ajustée)	24
Rangs	
Nombre de valeurs égales	2
Somme des rangs négatifs	226,5
Somme des rangs positifs (statistique du test)	73,5
Calcul de la valeur p	
Approximation normale (AN) utilisée	Oui
Egalités, mais sans correction	Non
Moyenne pour AN	150
Ecart type pour AN avec correction des égalités	34,95711659
Statistique z pour AN avec correction des égalités	-2,1741
Valeur p	0,0297
Niveaux de signification	
Hypoth. nulle à signification 10%	Rejeter
Hypoth. nulle à signification 5%	Rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter

Remarque : Dans ce rapport, la valeur p se calcule sous approximation normale quand la taille d'échantillon est supérieure à 15.

Données manquantes et liaison aux données

- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

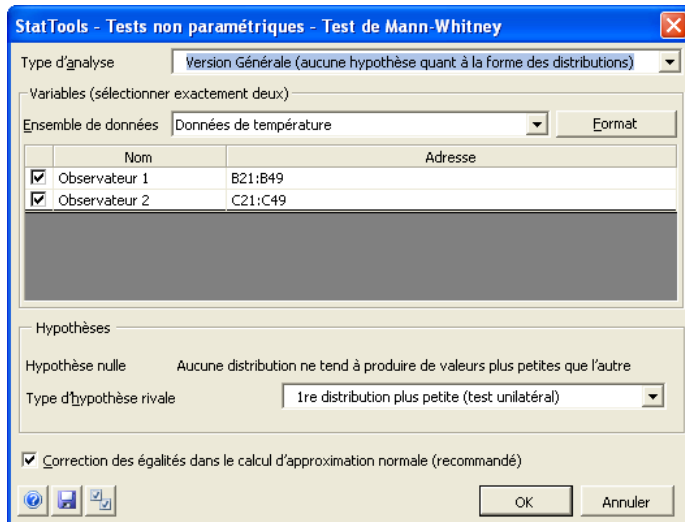
Commande Test de Mann-Whitney

Effectue un test de Mann-Whitney sur les variables.

La commande Test de Mann-Whitney effectue un test d'hypothèse sur deux échantillons. Dans une version du test (la version Médianes), l'hypothèse stipule que les médianes des deux populations sont identiques. Dans cette version, les distributions de probabilités sont présumées de même forme. L'autre version (Générale) ne repose pas sur ce principe. Sous l'hypothèse de cette version, aucune des distributions de probabilités ne tend à produire de valeurs plus petites que l'autre. Plus précisément, l'hypothèse stipule que $P[X1 > X2] = P[X2 > X1]$, où $P[X1 > X2]$ représente la probabilité qu'une observation de la population 1 soit supérieure à une observation de la population 2. On remarquera que le test de Mann-Whitney peut servir à rejeter l'hypothèse que deux échantillons soient générés par la même distribution de probabilités. Le test de Mann-Whitney est souvent désigné aussi sous l'appellation de test de somme ordinale.

Boîte de dialogue Test de Mann-Whitney

Cette analyse se configure dans la boîte de dialogue **Test de Mann-Whitney** :



L'analyse requiert deux variables. Elles peuvent être au format empilé ou désempilé, et peuvent provenir d'ensemble de données différents.

Dans la boîte de dialogue **Test de Mann-Whitney** :

- **Type d'analyse** désigne la formulation de l'hypothèse nulle et de ses rivales. Les options suivantes sont proposées :
 - **Version générale** effectue un test d'hypothèse pour déterminer si une distribution de probabilités tend à produire des valeurs plus petites que l'autre
 - **Hypothèses.**
 - Sous l'**hypothèse nulle**, aucune des distributions de probabilités ne tend à produire de valeurs plus petites que l'autre. Plus précisément, l'hypothèse stipule que $P[X_1 > X_2] = P[X_2 > X_1]$, où $P[X_1 > X_2]$ représente la probabilité qu'une observation de la population 1 soit supérieure à une observation de la population 2, et $P[X_2 > X_1]$ est interprété de même. Pour les distributions continues, cela revient à dire que les deux probabilités sont de 0,5 ($P[X_1 > X_2] = P[X_2 > X_1] = 0,5$)
 - L'**hypothèse rivale** peut être « unilatérale » (une probabilité est supérieure ou inférieure à l'autre) ou « bilatérale » (les deux probabilités ne sont pas égales).
 - **Version Médianes** effectue un test d'hypothèse pour déterminer si la médiane d'une population est identique à celle de l'autre ou si elle en est différente. Cette version présume deux distributions de forme identique..
 - **Hypothèses.**
 - L'**hypothèse nulle** stipule que les deux médianes sont égales.
 - L'**hypothèse rivale** peut être « unilatérale » (la médiane de la première population est supérieure ou inférieure à celle de la seconde) ou « bilatérale » (les deux médianes ne sont pas égales).

Remarque : Les calculs effectués sous les versions Générale et Médianes du test sont identiques : les deux versions ne diffèrent qu'en ce qui concerne les suppositions de formes de distribution égales et l'hypothèse nulle. Les deux versions sont proposées pour indiquer clairement que le test de Mann-Whitney s'applique même aux situations où l'on ne peut pas présumer que les distributions ont des formes plus ou moins identiques, du moment que l'on considère une hypothèse nulle appropriée. Plus spécifiquement, si on exécute le test en présence de deux distributions de formes clairement distinctes et que le test rejette l'hypothèse nulle, la raison peut en être la différence des médianes mais aussi celle des variances ou il peut s'agir d'autres raisons encore.

- L'option **Correction des égalités** est recommandée pour corriger les rangs ex-æquo du test sous approximation normale seulement. La correction implique le dénombrement d'éléments dans les groupes de rangs ex-æquo et la réduction correspondante de la variance. La correction accroît toujours la valeur de la statistique z en présence de rangs ex-æquo. (Remarque : La correction des égalités ne cause aucun changement de variance en l'absence de rangs ex-æquo.)

Rapport de test de Mann-Whitney

Statistiques d'échantillonnage		
	Observateur 1 Données de température	Observateur 2 Données de température
Taille d'échantillon	29	24
Moyenne	29,292	32,017
Ecart type	4,671	5,076
Médiane	29,400	31,195

Test de Mann-Whitney (Version Générale)	
Hypothèses	
Hypothèse nulle	Aucune dist. plus petite
Hypothèse rivale	Observateur 1 Plus petite
Rangs	
Nombre de valeurs ordinales	53
Nombre de valeurs égales	2
Somme ordinale de la variable 1 (statistique du test)	690,5
Somme ordinale de la variable 2	740,5
Calcul de la valeur p	
Approximation normale (AN) utilisée	Oui
Egalités, mais sans correction	Non
Moyenne pour AN	783
Ecart type pour AN avec correction des égalités	55,96314618
Statistique z pour AN avec correction des égalités	-1,6439
Valeur p	0,0501
Niveaux de signification	
Hypoth. nulle à signification 10%	Rejeter
Hypoth. nulle à signification 5%	Ne pas rejeter
Hypoth. nulle à signification 1%	Ne pas rejeter

Dans ce rapport, la *valeur p* se calcule sous approximation normale quand la taille de l'un des deux échantillons est supérieure à 10 (sauf quand l'une des tailles est 11 ou 12 et que l'autre est 3 ou 4).

Données manquantes et liaison aux données

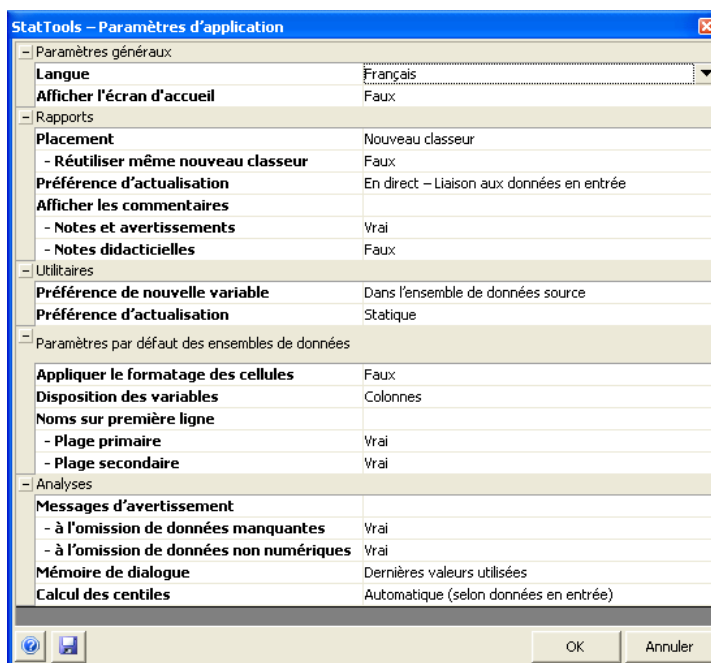
- **Données manquantes** – Les données manquantes sont admises. Toutes les lignes présentant des données manquantes pour les variables sélectionnées sont omises.
- **Liaison aux données** – Tous les rapports se calculent au moyen de formules liées aux données. Si les valeurs de la variable sélectionnée changent, les sorties changent automatiquement.

Menu Utilitaires

Commande Paramètres d'application

Spécifie les paramètres des rapports, graphiques, utilitaires, ensembles de données et analyses StatTools.

La commande **Paramètres d'application** permet de spécifier les paramètres généraux des rapports, graphiques, utilitaires, ensembles de données et analyses StatTools. Ces paramètres s'appliquent à toutes les analyses et tous les ensembles de données. D'autres paramètres spécifiques se configurent dans la boîte de dialogue propre à chaque analyse.



Rapports

Les paramètres des **Rapports** configurent les options par défaut des rapports et graphiques créés par les analyses StatTools.

- **Placement** spécifie l'emplacement Excel des nouveaux rapports et graphiques :
 - **Classeur actif** crée une nouvelle feuille de calcul pour chaque rapport.
 - **Nouveau classeur** crée (si nécessaire) un nouveau classeur de rapports StatTools appelé à recevoir chaque rapport sur une nouvelle feuille.
 - **Après la dernière colonne utilisée de la feuille active** place chaque rapport StatTools sur la feuille active, à droite de la dernière colonne utilisée.
 - **Invite de première cellule** invite l'utilisateur à sélectionner, après l'analyse, la cellule appelée à marquer le coin supérieur gauche du rapport ou graphique.
 - Sous l'option Nouveau classeur, **Réutiliser même nouveau classeur** configure l'utilisation de ce même classeur pour tous les rapports.
- **Préférence d'actualisation** spécifie l'actualisation ou non des résultats en cas de changement des données en entrée :
 - **En direct - Liaison aux données en entrée** actualise automatiquement les rapports quand les données en entrée changent.
 - **Statique - Valeurs fixes** n'actualise pas les rapports quand les données en entrée changent. Les statistiques restent fixes, en fonction des valeurs de données en entrée applicables au moment de l'exécution de la procédure.

Les résultats s'actualisent dans StatTools à travers l'usage de formules Excel et de fonctions StatTools. Par exemple, la formule

`=StatMean('Intervalle de confiance.xls'!Paire)`

calcule la moyenne de la variable *Paire* (basée sur les données de la plage Excel « Paire » située dans le classeur Intervalle de confiance.xls). La valeur renvoyée par la fonction StatTools s'actualise en fonction des changements qui affectent les données de la plage Paire.

Les rapports et graphiques de toutes les procédures StatTools peuvent être actualisés en direct, sauf pour les analyses suivantes :

- 1) Régression
- 2) Analyse discriminante
- 3) Régression logistique
- 4) Préviation

Ces procédures exigent en effet de longs et fastidieux recalculs, qui ralentiraient trop l'exécution d'Excel lors de l'actualisation.

- **Afficher les commentaires** spécifie les catégories de messages StatTools à inclure dans les rapports. Des notes, avertissements et notes didacticielles contextuelles peuvent s'afficher dans les rapports comme illustré ici :

Exemple StatTools : Test de Mann-Whitney (version générale, données déséquilibrées)

L'analyse Test de Mann-Whitney de StatTools sert à vérifier si un ensemble de données d'échantillon permet d'accepter ou de rejeter différentes hypothèses. Dans une version de test (la version Médiane), l'hypothèse stipule que les médians de deux populations sont identiques. Dans cette version, les distributions de probabilités sont présumées de même forme. L'autre version (Général) ne repose pas sur ce principe. Sous l'hypothèse de cette version, aucune des distributions de probabilités ne tend à produire de valeurs plus petites qu'une autre. Plus précisément, l'hypothèse stipule que $P(X_1 < X_2) = P(X_2 < X_1)$, ou $P(X_1 < X_2)$ représente la probabilité qu'une observation de la population 1 soit supérieure à une observation de la population 2.

Remarque : Cette procédure admet les données empliées comme déséquilibrées. Dans l'exemple présenté ici, elles sont déséquilibrées.

L'ensemble de données considéré ici représente les mesures de température de deux observateurs à une station météorologique locale. L'exécution du Test de Mann-Whitney (version générale) quand à savoir si les observations de l'observateur 1 tendent à être inférieures à celles de l'observateur 2 produit les résultats indiqués. Sous l'hypothèse nulle unilatérale, on voit que l'on peut rejeter l'hypothèse nulle à signification 10%, mais pas à signification 5% et 1%.

Exemple créé par Palisade Corporation.

Observateur 1	Observateur 2
29,4	29,73
26,26	37,88
34,1	32,59
36,81	33,54
26,54	29,08
24,55	29,03
25,33	29,64
32,8	23,55
21,24	30,47
31,84	38,34
19,75	35,69
31,35	39,33
27,74	29,24
35,85	34,75
35,52	25,27
27,48	24,54

Statistiques d'échantillon	Observateur 1	Observateur 2
	Données de température	Données de température
Taille d'échantillon	29	24
Moyenne	29,292	32,017
Ecart type	4,671	5,076
Médiane	29,400	31,195

Test de Mann-Whitney (Version Générale)

Hypothèses	Aucune dist. plus petite	Observateur 1 Plus petite
Hypothèse nulle		
Hypothèse rivale		

Range

Statistique	Valeur
Nombre de valeurs ordinales	53
Nombre de valeurs égales	2
Somme ordinale de la variable 1 (statistique)	690,5
Somme ordinale de la variable 2	740,5
Calcul de la valeur p	

Note StatTools :
Sous l'hypothèse de version générale, aucune des distributions ne tend à produire de valeurs plus petites que l'autre ($P(X_1 < X_2) = P(X_2 < X_1)$, ou $P(X_1 < X_2)$: probabilité que l'observation de la population 1 > obs. de pop. 2).

Utilitaires

Les paramètres des **Utilitaires** configurent les options par défaut applicables aux nouvelles variables créées par les Utilitaires de données StatTools. Le menu Utilitaires de données y donne accès. Ces utilitaires créent de nouvelles variables par transformation, combinaison ou autre traitement de variables existantes.

- **Préférence de nouvelle variable** spécifie l'emplacement désiré des nouvelles variables créées par les utilitaires :
 - **Dans l'ensemble de données source** insère chaque nouvelle variable créée à droite (ou au bas) de l'ensemble de données de la variable originale.
 - **Nouvel ensemble de données** place les nouvelles variables dans un nouvel ensemble de données.

La préférence sélectionnée ici n'est cependant pas toujours applicable : sous les utilitaires **Empiler** et **Désempiler** (qui placent toujours les nouvelles variables dans un nouvel ensemble de données), notamment, et quand **les données originales proviennent d'un ensemble de données multiples** (et que les nouvelles variables ne peuvent donc être introduites que dans l'ensemble de données source).

- **Préférence d'actualisation** spécifie l'actualisation ou non des nouvelles variables créées par un utilitaire en cas de changement des données de la variable originale.
 - **En direct - Liaison aux données en entrée** actualise automatiquement les valeurs des nouvelles variables quand les données en entrée changent.
 - **Statique - Valeurs fixes** n'actualise pas les valeurs des nouvelles variables quand les données en entrée changent. Les valeurs des nouvelles variables restent fixes, en fonction des valeurs de données en entrée applicables au moment de l'exécution de la procédure.

Dans certains cas, toutefois, la préférence d'actualisation sélectionnée ne s'applique pas et les nouvelles variables sont toujours soumises à la préférence **Statique - Valeurs fixes** : sous les utilitaires **Empiler**, **Désempiler** et **Échantillons aléatoires**, auxquels l'actualisation en direct ne s'applique pas.

Ensembles de données

Les **Paramètres par défaut des ensembles de données** configurent les options applicables aux nouveaux ensembles de données définis sous la commande Gestionnaires des ensembles de données. Il s'agit des paramètres proposés par défaut lors de la création d'un nouvel ensemble de données. Rien n'empêche de les changer dans la boîte de dialogue Gestionnaire des ensembles de données.

- **Appliquer le formatage des cellules** spécifie le formatage ou non de l'ensemble de données par StatTools.
- **Disposition** configure la disposition par défaut des variables (en colonnes ou lignes).
- **Noms dans première colonne/sur première ligne (Plage primaire)** spécifie si les noms doivent être entrés dans la première colonne ou sur la première ligne de la première plage définie.
- **Noms dans première colonne/sur première ligne (Plage secondaire)** spécifie si les noms doivent être entrés dans la première colonne ou sur la première ligne de la deuxième et de toutes les plages suivantes d'un ensemble de données multi-plages.

Analyses

Les paramètres des **Analyses** configurent les entrées affichées par défaut dans les boîtes de dialogue des analyses. Il s'agit simplement des entrées proposées par défaut à la première ouverture de la boîte de dialogue d'une analyse. Rien n'empêche de les changer dans chaque boîte de dialogue.

- Les options **Messages d'avertissement** spécifient l'affichage ou non de messages d'avertissement en cours d'analyse si StatTools détecte des données manquantes dans une variable ou des données non numériques.

- **Mémoire de dialogue** spécifie les entrées affichées par défaut dans les boîtes de dialogue utilisées pour la configuration des analyses :
 - **Dernières valeurs utilisées (par classeur)** spécifie l’affichage des dernières entrées saisies dans la boîte de dialogue lors de sa dernière ouverture dans le classeur actif. À la première ouverture, les boîtes de dialogue proposent les paramètres système par défaut enregistrés.
 - **Défauts système** propose toujours dans les boîtes de dialogue affichées les paramètres système enregistrés par défaut pour l’analyse.

- **Calcul des centiles** spécifie la méthode de calcul des centiles. Suivant la nature des données, différentes méthodes peuvent produire de meilleurs résultats. Les méthodes suivantes sont proposées, suivant le type de données auxquelles elles conviennent le mieux :
 - 1) Automatique (selon données en entrée)
 - 2) Interpolation – extrémités asymétriques (continues)
 - 3) Interpolation – extrémités symétriques (continues)
 - 4) Fonction de centiles Excel (continues)
 - 5) Observation la plus proche (discrètes)
 - 6) Fonction dist. empirique (discrètes)
 - 7) Fonction dist. empirique avec étalement (discrètes)

Commande Supprimer les ensembles de données

Supprime les ensembles de données StatTools du classeur actif.

La commande **Supprimer les ensembles de données** supprime tous les ensembles de données définis du classeur actif. Les données Excel en soi ne sont pas supprimées : l'opération ne concerne que la définition des ensembles de données.

Commande Effacer la mémoire de dialogue

Efface la mémoire des entrées saisies dans les boîtes de dialogue d'analyse.

La commande **Effacer la mémoire des entrées** efface le « souvenir » des entrées saisies dans les boîtes de dialogue d'analyse. Les boîtes de dialogue affichées ultérieurement proposeront initialement les paramètres par défaut enregistrés.

Commande Décharger StatTools

Décharge le compagnon StatTools.

La commande Décharger StatTools décharge StatTools et ferme toutes ses fenêtres.

Menu Aide

Aide StatTools

Ouvre le fichier d'aide en ligne de StatTools.

La commande Aide StatTools du menu Aide ouvre le fichier d'aide principal de StatTools. Toutes les fonctionnalités et commandes de StatTools y sont décrites.

Manuel en ligne

Ouvre le manuel en ligne de StatTools.

La commande Manuel en ligne du menu Aide ouvre l'exemplaire en ligne de ce manuel, au format PDF. Adobe Acrobat Reader doit être installé pour permettre la consultation de ce manuel en ligne.

Commande Activation de licence

Affiche les informations de licence de StatTools et permet l'autorisation des versions d'essai.

La commande Activation de licence du menu Aide affiche la boîte de dialogue Activation de licence, indiquant la version et les informations de licence de votre exemplaire de StatTools. La conversion des versions d'essai de StatTools en copie autorisée s'effectue aussi dans cette boîte de dialogue.

Pour plus de détails sur l'autorisation de votre exemplaire de StatTools, voir le **Chapitre 1 : Mise en route**.

Commande À propos

Affiche la version et les informations de copyright relatives à StatTools.

La commande À propos du menu Aide affiche la boîte de dialogue À propos, indiquant la version et les mentions relatives aux droits d'auteur de votre exemplaire de StatTools.

Référence : Fonctions StatTools

Introduction

StatTools utilise ses fonctions de feuille de calcul personnalisées pour renvoyer les statistiques calculées aux formules Excel. Ces fonctions permettent :

- 1) l'incorporation de calculs statistiques dans les formules de feuille de calcul, comme les fonctions Excel standard ;
- 2) l'actualisation « en direct » des statistiques : les résultats changent quand les données originales changent.

Les formules, dans les cellules d'un rapport StatTools, contiennent des fonctions de feuille de calcul StatTools ; Toutes les fonctions StatTools commencent par le préfixe « Stat » : **StatMean()** (pour la moyenne) ou **StatStdDev()** (pour l'écart type), par exemple. Toutes les fonctions StatTools figurent dans la boîte de dialogue Excel Insérer une fonction pour la simplicité de leur insertion.

Fonctions StatTools vs Fonctions Excel

Dans certains cas, StatTools remplace les fonctions statistiques intégrées d'Excel par ses propres calculs, robustes et rapides. La précision des calculs statistiques intégrés d'Excel laisse souvent à désirer. StatTools les évite totalement. Même les fonctions statistiques de feuille de calcul Excel telles qu'ECARTYPE() sont remplacées par de nouvelles versions StatTools robustes, telles que StatSTDEV(). Les calculs statistiques de StatTools sont conformes aux tests de précision les plus rigoureux, sous performance optimisée par .DLL C++, plutôt que par macro-calculs.

Les fonctions StatTools, par opposition aux fonctions Excel intégrées, gèrent les données empilées. La fonction StatTools **StatDestack** désempile automatiquement les données d'un ensemble de données empilé (pour la catégorie que l'utilisateur spécifie). Elle transmet ensuite les données à l'analyse d'une fonction statistique StatTools.

Les fonctions StatTools gèrent aussi l'analyse de données de différentes feuilles de calcul. Les ensembles de données multi-feuilles admettent plus de 65 535 points par variable. Ils se définissent à

travers le bouton **Multiple** de la boîte de dialogue Gestionnaire des ensembles de données.

Fonctions de distribution

StatTools inclut un ensemble de fonctions de distribution (telles que **StatBinomial**) utilisées en remplacement des fonctions Excel intégrées (en l'occurrence **loi.binomiale**). Contrairement à leurs homologues Excel, les fonctions de distribution StatTools peuvent renvoyer plusieurs valeurs distinctes d'une distribution de probabilités. L'argument **statistique** (avant dernier argument de la fonction) régit la valeur renvoyée. Cet argument peut être une valeur de 1 à 12 ou une chaîne indiquant la statistique à obtenir pour la distribution entrée :

Liste des statistiques possibles

Valeur ou chaîne entrée	Statistique renvoyée
1 ou « mean »	moyenne
2 ou « stddev »	écart type
3 ou « variance »	variance
4 ou « skewness »	asymétrie
5 ou « kurtosis »	aplatissement
6 ou « mode »	mode
7 ou « discrete mean »	moyenne discrète (ou valeur la plus proche de la vraie moyenne possible)
8 ou « x to y »	x à y (valeur y de la distribution pour une valeur x entrée)
9 ou « x to p »	x à p (valeur p de la distribution pour une valeur x entrée)
10 ou « p to x »	p à x (valeur x de la distribution pour une valeur p entrée)
11 ou « x to q »	x à q (valeur q de la distribution pour une valeur x entrée)
12 ou « q to x »	q à x (valeur x de la distribution pour une valeur q entrée)

Par exemple, la fonction de distribution StatTools

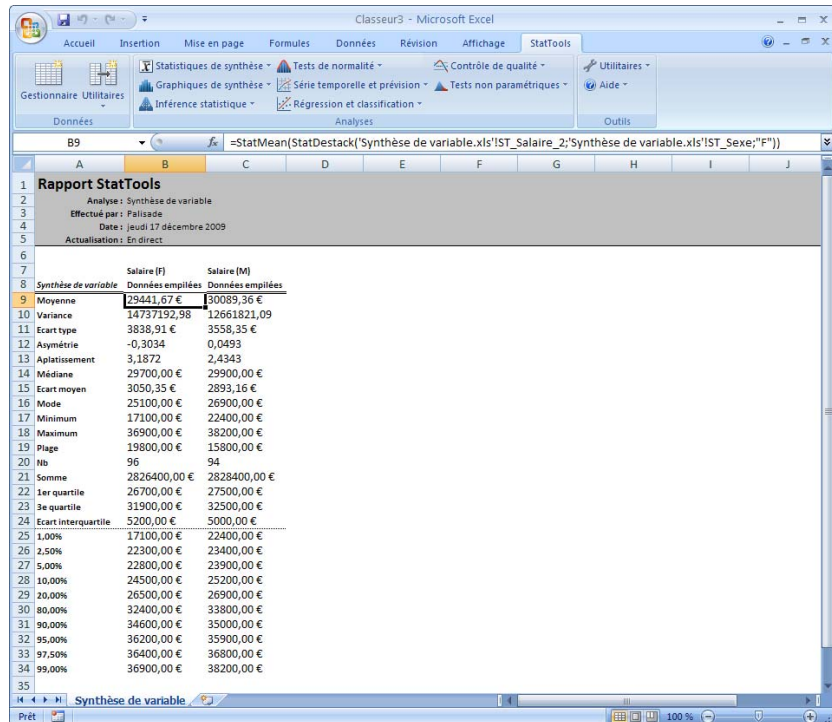
StatNormal(10;1;"x to p"; 9,5)

renvoie la valeur p associée à la valeur x 9,5 dans une distribution normale à moyenne 10 et écart type 1.

Rapports « en direct »

StatTools utilise ses fonctions personnalisées pour rendre les résultats aussi « actuels » que possible. Dans la mesure du possible, les rapports sont dotés de formules qui les relient aux données originales. Supposons par exemple une variable *Poids*, pour laquelle on désire des mesures de synthèse, telles que la moyenne et l'écart type. La procédure Statistiques de synthèse donne à la plage des poids le nom *Poids* et entre les formules dans les cellules de sortie :

=StatMean(Poids) et **=StatStdDev(Poids)**. StatMean et StatStdDev sont des fonctions StatTools intégrées de calcul de la moyenne et de l'écart type. Elles remplacent les fonctions Excel standard intégrées correspondantes. Ces fonctions présentent l'avantage qu'en cas de changement des données, les résultats changent automatiquement, sans qu'il soit nécessaire de réexécuter la procédure.



Classeur3 - Microsoft Excel

StatTools

Statistiques de synthèse - Tests de normalité - Contrôle de qualité -
Graphiques de synthèse - Série temporelle et prévision - Tests non paramétriques -
Inférence statistique - Régression et classification -

Données

B9 =StatMean(StatDestack('Synthèse de variable.xls!ST_Salaire_2';Synthèse de variable.xls!ST_Sexe;'F'))

Rapport StatTools

Analyse : Synthèse de variable
Effectué par : Pallsade
Date : jeudi 17 décembre 2009
Actualisation : En direct

Synthèse de variable	Données employés	Données employées
Moyenne	29441,67 €	30089,36 €
Variance	14737192,98	12661821,09
Ecart type	3838,91 €	3558,35 €
Asymétrie	-0,3034	0,0493
Aplatissement	3,1872	2,4343
Médiane	29700,00 €	29900,00 €
Ecart moyen	3050,35 €	2893,16 €
Mode	25100,00 €	26900,00 €
Minimum	17100,00 €	22400,00 €
Maximum	36900,00 €	38200,00 €
Plage	19800,00 €	15800,00 €
Nb	96	94
Somme	2826400,00 €	2828400,00 €
1er quartile	26700,00 €	27500,00 €
3e quartile	31900,00 €	32500,00 €
Ecart interquartile	5200,00 €	5000,00 €
1.00%	17100,00 €	22400,00 €
2.50%	22300,00 €	23400,00 €
5.00%	22800,00 €	23900,00 €
10.00%	24500,00 €	25200,00 €
20.00%	26500,00 €	26900,00 €
80.00%	32400,00 €	33800,00 €
90.00%	34600,00 €	35000,00 €
95.00%	36200,00 €	35900,00 €
97.50%	36400,00 €	36800,00 €
99.00%	36900,00 €	38200,00 €

Synthèse de variable

La liaison des résultats aux données n'est cependant pas toujours pratique. Dans le cas des régressions, notamment. StatTools ne fournit pas les formules utilisées pour créer une sortie de régression ; il n'en produit que les résultats numériques. Si les données changent, il faut donc réexécuter les procédures.

L'**option Statique des paramètres des Rapports** désactive la liaison des rapports aux données. Cette désactivation est utile si le temps de recalcul Excel est excessif lors du changement des données.

Référence : Liste des fonctions statistiques

Fonctions disponibles

Le tableau ci-dessous présente la liste des fonctions personnalisées que StatTools ajoute à Excel. Lorsqu'elles sont utilisées, toutes ces fonctions sont précédées du préfixe **Stat**.

Fonction	Description
AUTOCORRELATION(<i>données; nbreRetards</i>)	Calcule l'autocorrélation des valeurs d'un ensemble de données.
AVEDEV(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'écart moyen absolu des données par rapport à leur moyenne. Les arguments peuvent être des nombres, des matrices ou des plages.
BINOMIAL (<i>N; P; statistique; valeur</i>)	Calcule la <i>statistique</i> de la distribution binomiale entrée.
CATEGORYINDICES (<i>plage; nom_catégorie</i>)	Obtient les indices de cellule d'une catégorie spécifiée.
CATEGORYNAMES(<i>plage</i>)	Obtient les noms des catégories d'une plage.
CATEGORYOCCURRENCECOUNT(<i>plage; nom_catégorie</i>)	Calcule le nombre de cellules dans une plage de catégorie spécifiée.
ChiSQ (<i>deg_liberté; statistique; valeur</i>)	Calcule la <i>statistique</i> de la distribution chi carré unilatérale entrée.
CORRELATIONCOEFF(<i>données1; données2</i>)	Calcule le coefficient de corrélation entre 2 ensembles de données.
COUNT(<i>Données1; Données2; ...DonnéesN</i>)	Calcule le nombre d'éléments compris dans ses arguments, qui peuvent être des nombres, des matrices ou des plages.
COUNT CATEGORIES(<i>plage</i>)	Compte le nombre de catégories dans une plage.
COUNTCELLSBYTYPE (<i>plage; type</i>)	Calcule le nombre de cellules du type spécifié dans une plage.
COUNTRANGE(<i>Plage; ValMin; ValMax; InclureMin; InclureMax</i>)	Calcule le nombre de valeurs de la <i>Plage</i> comprises entre <i>ValMin</i> et <i>ValMax</i> .
COVARIANCE(<i>données1; données2</i>)	Calcule la covariance d'échantillon entre 2 ensembles de données.
CovarianceP (<i>données1; données2</i>)	Calcule la covariance de population entre 2 ensembles de données. Les nombres manquants renvoient un blanc.
DESTACK(<i>plage_données; plage_catégories_1; catégorie_1; plage_catégories_2; catégorie_2</i>)	Extrait les données d'une catégorie spécifiée de données empilées.
DURBINWATSON(<i>données</i>)	Calcule la statistique Durbin-Watson des valeurs d'un ensemble de données.

Fonction	Description
F (deg_liberté1; deg_liberté2; statistique; valeur)	Calcule la <i>statistique</i> pour la distribution F entrée pour 2 ensembles de données.
GETCELLVALUES(plage)	Obtient les valeurs de toutes les cellules d'un type spécifié dans la plage.
KURTOSIS(Données1; Données2; ... DonnéesN)	Calcule l'aplatissement d'échantillon de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
KURTOSISP(Données1; Données2; ... DonnéesN)	Calcule l'aplatissement de population de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
LN(x)	Calcule le logarithme népérien d'un nombre réel positif.
MAX(Données1; Données2; ...DonnéesN)	Calcule le maximum de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
MEAN(Données1; Données2; ...DonnéesN)	Calcule la moyenne arithmétique de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
MEANABS(Données1; Données2; ...DonnéesN)	Calcule la moyenne arithmétique des valeurs absolues de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
MEDIAN(données; drapeau discret)	Calcule la médiane d'un ensemble de données.
MIN(Données1; Données2; ...DonnéesN)	Calcule le minimum de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
NORMAL (moyenne; écart_type; statistique; valeur)	Calcule la <i>statistique</i> pour la distribution normale (de Gauss) entrée.
PAIRCOUNT(Données1; Données2)	Compte le nombre de paires de cellules pour lesquelles chaque cellule de la paire est numérique.
PairMean (Données1; Données2)	Calcule la moyenne des différences entre les paires de cellules.
PAIRMEDIAN(Données1; Données2)	Calcule la médiane des différences entre les paires de cellules.
PairStdDev (Données1; Données2)	Calcule l'écart type d'échantillon des différences entre les paires de cellules.
PERCENTILE(données; p; Drapeau_discret)	Calcule le p ^e centile d'un ensemble de données.
Product (Données1; Données2; ...DonnéesN)	Calcule le produit de ses arguments, qui peuvent être des nombres, des matrices ou des plages.

Fonction	Description
QUARTILE(<i>données; q; drapeau discret</i>)	Calcule le quartile spécifié d'un ensemble de données.
RAND()	Renvoie un nombre aléatoire compris entre 0 et 1.
RANGE(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'étendue (maximum – minimum) de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
RUNSTEST(<i>données; limite</i>)	Calcule la statistique des runs pour les valeurs d'un ensemble de données.
SKEWNESS(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'asymétrie d'échantillon de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
SKEWNESSP(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'asymétrie de population de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
Standardize (<i>x; moyenne; écart_type</i>)	Calcule une valeur normalisée d'une distribution à moyenne et écart type spécifiés.
STDDEV(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'écart type d'échantillon de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
STDDEVP(<i>Données1; Données2; ...DonnéesN</i>)	Calcule l'écart type de population de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
SUM(<i>Données1; Données2; ...DonnéesN</i>)	Calcule la somme de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
SumDEVSQ (<i>Données1; Données2; ...DonnéesN</i>)	Calcule la somme du carré de l'écart par rapport à la moyenne de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
SUMSQ(<i>Données1; Données2; ...DonnéesN</i>)	Calcule la somme du carré de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
STUDENT(<i>deg_liberté; statistique; valeur</i>)	Calcule la <i>statistique</i> pour la distribution de Student entrée.
VARIANCE(<i>Données1; Données2; ...DonnéesN</i>)	Calcule la variance d'échantillon de ses arguments, qui peuvent être des nombres, des matrices ou des plages.
VARIANCEP(<i>Données1; Données2; ...DonnéesN</i>)	Calcule la variance de population de ses arguments, qui peuvent être des nombres, des matrices ou des plages.

Description détaillée des fonctions

Les fonctions statistiques sont listées ici avec leurs arguments obligatoires.

AUTOCORRELATION

Description	AUTOCORRELATION (<i>données</i> ; <i>nbreRetards</i>) calcule l'autocorrélation des valeurs d'un ensemble de données selon le nombre de retards <i>nbreRetards</i> . L'argument <i>données</i> représente la matrice ou la plage de données pour lesquelles l'autocorrélation doit être calculée ; <i>nbreRetards</i> représente le nombre de retards à utiliser.
Exemples	StatAutocorrelation (C1:C100;1) renvoie l'autocorrélation pour les données de la plage C1:C100 avec 1 retard.
Directives	<i>nbreRetards</i> doit être supérieur ou égal à 1.

AVEDEV

Description	AVEDEV (<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i>) calcule l'écart absolu moyen de <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> par rapport à leur moyenne.
Exemples	StatAveDev (1;2;5) calcule l'écart absolu moyen de 1;2;5 par rapport à leur moyenne.
Directives	Les arguments <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

BINOMIAL

Description	BINOMIAL (<i>N</i> ; <i>P</i> ; <i>statistique</i> ; <i>valeur</i>) calcule la <i>statistique</i> pour la distribution binomiale à valeurs <i>N</i> et <i>P</i> spécifiées.
Exemples	StatBinomial (2;0,3;"x to p";1) calcule la valeur p pour une valeur x de 1 dans la distribution binomiale à N=2 et P =0,3
Directives	<i>N</i> représente le nombre d'essais ou d'événements et doit être un entier. <i>P</i> représente la probabilité et doit être ≥ 0 et ≤ 1 . <i>statistique</i> est une valeur entière (1 à 12) ou une chaîne indiquant la statistique à renvoyer. Pour plus de détails, voir la section intitulée Fonctions de distribution en début de chapitre. <i>valeur</i> doit être ≥ 0 (quand une valeur x est entrée) ou ≥ 0 et ≤ 1 (quand une valeur p est entrée).

CATEGORYINDICES

Description	CATEGORYINDICES (<i>plage</i> ;nom_ catég orie) obtient les indices des cellules contenant le <i>nom de catég orie</i> spécifié dans la <i>plage</i> Excel. Il s'agit d'une fonction matricielle et les indices renvoyés sont des nombres-indices (entre 1 et le nombre de cellules comprises dans la <i>plage</i>) indiquant la position des cellules contenant le <i>nom de catég orie</i> dans la <i>plage</i> .
Exemples	StatCategoryIndices (C1:C100;"Masculin") renvoie les indices (entre 1 et 100) des cellules qui contiennent la chaîne Masculin.
Directives	<i>plage</i> désigne une plage Excel correcte. <i>nom_ catég orie</i> est une chaîne, une valeur ou une référence de cellule indiquant la catég orie à rechercher.

CATEGORYNAMES

Description	CATEGORYNAMES (<i>plage</i>) obtient le nom des catég ories comprises dans la plage Excel spécifiée. Il s'agit d'une fonction matricielle, renvoyant un nombre de noms compris entre 1 et le nombre de cellules comprises dans la <i>plage</i> .
Exemples	StatCategoryNames (C1:C100) renvoie les noms des catég ories comprises dans la plage C1:C100.
Directives	<i>plage</i> désigne une plage Excel correcte.

CATEGORYOCCURRENCECOUNT

Description	CATEGORYOCCURRENCECOUNT (<i>plage</i> ;nom_ catég orie) renvoie le nombre de cellules contenant le <i>nom de catég orie</i> spécifié dans la <i>plage</i> Excel.
Exemples	StatCategoryOccurrenceCount (C1:C100;"Masculin") renvoie le nombre de cellules comprises dans la plage C1:C100 qui contiennent la chaîne « Masculin ».
Directives	<i>plage</i> désigne une plage Excel correcte. <i>nom_ catég orie</i> est une chaîne, une valeur ou une référence de cellule indiquant la catég orie à rechercher.

CHISQ

Description	CHISQ (<i>deg_liberté</i> ;statistique;valeur) calcule la <i>statistique</i> pour la distribution chi carré unilatérale sous les degrés de liberté indiqués.
Exemples	StatChisq (2;"x to p";5) calcule la distribution chi carré unilatérale à la valeur 5 à 2 degrés de liberté.
Directives	Le nombre de degrés de liberté <i>deg_liberté</i> doit être compris entre 1 et 32767. <i>statistique</i> est une valeur entière (1 à 12) ou une chaîne indiquant la statistique à renvoyer. Pour plus de détails, voir la section intitulée Fonctions de distribution en début de chapitre. <i>valeur</i> doit être ≥ 0 (quand une valeur x est entrée) ou ≥ 0 et ≤ 1 (quand une valeur p est entrée).

CORRELATIONCOEFF

Description	CORRELATIONCOEFF (<i>données1</i> ;données2) calcule le coefficient de corrélation entre deux ensembles de données, <i>données1</i> et <i>données2</i> .
Exemples	StatCorrelationCoeff (A1:A100;B1:B100) calcule le coefficient de corrélation entre les deux ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>données1</i> et <i>données2</i> doivent compter le même nombre d'éléments.

COUNT

Description	COUNT (<i>données1</i> ;données2;...donnéesN) calcule le nombre d'éléments compris dans <i>données1</i> , <i>données2</i> à <i>donnéesN</i> , qui peuvent être des nombres, des matrices ou des plages.
Exemples	StatCount (A1:A100;B1:B100) calcule le nombre d'éléments compris dans les deux ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>données1</i> ;données2;...donnéesN représentent 1 à 30 arguments, qui peuvent être des nombres, des matrices ou des plages.

COUNT CATEGORIES

Description	COUNTCATEGORIES (<i>plage</i>) renvoie le nombre de catégories comprises dans la <i>plage</i> Excel spécifiée.
Exemples	StatCountCategories (C1:C100) renvoie le nombre de catégories comprises dans la <i>plage</i> C1:C100.
Directives	<i>plage</i> désigne une <i>plage</i> Excel correcte.

COUNTCELLSBYTYPE

Description	COUNTCELLSBYTYPE (<i>plage</i> ; <i>type</i>) calcule le nombre d'éléments du <i>type</i> spécifié dans la <i>plage</i> entrée.
Exemples	StatCountCellsByType (A1:A100;1) calcule le nombre d'éléments numériques compris dans l'ensemble de données situé dans la <i>plage</i> A1:A100.
Directives	<i>plage</i> désigne une <i>plage</i> Excel correcte. Pour <i>type</i> , 1=numérique, 2=non vide, 3=non vide, non numérique, 4=vide. Remarque : StatTools considère vide une cellule ne contenant que des espaces.

COUNTRANGE

Description	COUNTRANGE (<i>plage</i> ; <i>ValMin</i> ; <i>ValMax</i> ; <i>inclureMin</i> ; <i>inclureMax</i>) calcule le nombre de valeurs tombant entre <i>ValMin</i> et <i>ValMax</i> dans la <i>plage</i> . Les valeurs <i>ValMin</i> et <i>ValMax</i> mêmes peuvent être incluses en réglant <i>InclureMin</i> et/ou <i>InclureMax</i> sur VRAI.
Exemples	StatCountRange (A1:A100;1;10;VRAI;VRAI) calcule le nombre de valeurs >=1 et <=10 comprises dans l'ensemble de données situé dans la <i>plage</i> A1:A100.
Directives	<i>plage</i> représente la <i>plage</i> de cellules dans laquelle les valeurs doivent être comptées. <i>ValMin</i> représente la valeur minimum considérée dans la <i>plage</i> . <i>ValeurMax</i> représente la valeur maximum considérée dans la <i>plage</i> . <i>InclureMin</i> est une valeur booléenne indiquant si la valeur minimum doit être incluse dans le nombre compté. Par défaut, cet argument est VRAI. <i>InclureMax</i> est une valeur booléenne indiquant si la valeur maximum doit être incluse dans le nombre compté. Par défaut, cet argument est VRAI.

COVARIANCE

Description	COVARIANCE (<i>données1;données2</i>) calcule la covariance d'échantillon entre les ensembles de données <i>données1</i> et <i>données2</i> .
Exemples	StatCovariance (A1:A100;B1:B100) calcule la covariance d'échantillon entre les deux ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>données1</i> et <i>données2</i> peuvent être des matrices ou des plages. <i>données1</i> et <i>données2</i> doivent compter le même nombre d'éléments.

COVARIANCEP

Description	COVARIANCEP (<i>données1;données2</i>) calcule la covariance de population entre les ensembles de données <i>données1</i> et <i>données2</i> .
Exemples	StatCovarianceP (A1:A100;B1:B100) calcule la covariance de population entre deux ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>données1</i> et <i>données2</i> peuvent être des matrices ou des plages.

DESTACK

Description	DESTACK (<i>plage_données</i> ; <i>plage_catégories_1</i> ; <i>catégorie_1</i> ; <i>plage_catégories_2</i> ; <i>catégorie_2</i>) extrait les données comprises dans une catégorie <i>catégorie_1</i> spécifiée des données empilées de <i>plage_données</i> . Cette fonction permet à d'autres fonctions statistiques StatTools d'accepter des données empilées en entrée. La fonction StatDestack ne s'utilise donc qu'« incorporée » dans d'autres fonctions, comme illustré dans l'exemple ci-dessous. Elle renvoie une matrice de données pour la catégorie spécifiée, extraite de <i>plage_données</i> .
Exemples	StatMean(StatDestack (B1:B100;A1:A100;"Masculin")) calcule la moyenne des valeurs de la plage B1:B100 pour lesquelles la plage de catégories correspondante A1:A100 contient la valeur « Masculin ».
Directives	<i>plage_données</i> représente la plage des données empilées. <i>plage_catégories_1</i> représente la plage contenant le nom de la première catégorie. <i>catégorie_1</i> représente la première catégorie pour laquelle les données doivent être obtenues. <i>plage_catégories_2</i> (argument facultatif) représente la plage contenant le nom de la deuxième catégorie. <i>catégorie_2</i> (facultatif) représente la deuxième catégorie pour laquelle les données doivent être obtenues.

DURBINWATSON

Description	DURBINWATSON (<i>données</i>) calcule la statistique Durbin-Watson des valeurs de l'ensemble de données <i>données</i> .
Exemples	StatDurbinWatson (A1:A100) calcule la statistique Durbin-Watson pour la plage de données située dans A1:A100.
Directives	<i>données</i> peut représenter une matrice ou une plage de données.

F

Description	F (<i>deg_liberté1</i> ; <i>deg_liberté2</i> ; <i>statistique</i> ; <i>valeur</i>) calcule la <i>statistique</i> pour la distribution F sous degrés de liberté <i>deg_liberté1</i> au numérateur et <i>deg_liberté2</i> au dénominateur.
Exemples	StatF (1;1;"x to p";1,5) calcule la valeur p d'une distribution F pour une valeur x de 1,5 à degrés de liberté = 1 au numérateur et degrés de liberté = 1 au dénominateur.
Directives	<i>deg_liberté1</i> et <i>deg_liberté2</i> doivent être des nombres entiers > 0. <i>statistique</i> est une valeur entière (1 à 12) ou une chaîne indiquant la statistique à renvoyer. Pour plus de détails, voir la section intitulée Fonctions de distribution en début de chapitre. <i>valeur</i> doit être >= 0.

GETCELLVALUES

Description	GETCELLVALUES (<i>plage</i> ; <i>typeCell</i>) obtient les valeurs de toutes les cellules du type spécifié dans la plage.
Exemples	StatGetCellValues (A1:A100;2) obtient les valeurs des cellules non vides de la plage située dans A1:A100.
Directives	<i>plage</i> représente la plage de cellules dont on veut obtenir les valeurs. <i>typeCell</i> 0=toutes les cellules, 1 =numériques, 2=non vides, 3=non vides, non numériques et 4=vides.

KURTOSIS

Description	KURTOSIS (<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i>) calcule l'aplatissement d'échantillon des données spécifiées dans <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> . Remarque : Calculée sur des données normalement distribuées, StatKurtosis renvoie la valeur 3.
Exemples	StatKurtosis (A1:A100;{1;2;3;2,4}) calcule l'aplatissement d'échantillon de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

KURTOSISP

Description	KURTOSISP (Données1;Données2;...DonnéesN) calcule l'aplatissement de population des données spécifiées dans Données1;Données2;...DonnéesN. Remarque : Calculée sur des données normalement distribuées, StatKurtosisP renvoie la valeur 3.
Exemples	StatKurtosisP (A1:A100;{1;2;3;2,4}) calcule l'aplatissement de population de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

LN

Description	LN (x) calcule le logarithme népérien.
Exemples	StatLN (4,5) calcule le logarithme népérien de 4,5.
Directives	x doit être un nombre réel positif.

MAX

Description	MAX (Données1;Données2;...DonnéesN) calcule le maximum des données spécifiées dans Données1;Données2;...DonnéesN.
Exemples	StatMax (A1:A100;{1;2;3;2,4}) calcule la valeur maximum de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

MEAN

Description	MEAN (Données1;Données2;...DonnéesN) calcule la moyenne des données spécifiées dans Données1;Données2;...DonnéesN.
Exemples	StatMean (A1:A100;{1;2;3;2,4}) calcule la valeur moyenne de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

MEANABS

Description	MEANABS (Données1;Données2;...DonnéesN) calcule la moyenne de la valeur absolue des données spécifiées dans Données1;Données2;...DonnéesN.
Exemples	StatMeanAbs (A1:A100;{1;2;3;2,4}) calcule la valeur moyenne des valeurs absolues de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

MEDIAN

Description	MEDIAN (données;drapCalc) calcule la médiane des valeurs comprises dans les données. Le calcul peut s'effectuer selon l'une de cinq méthodes, telle que facultativement spécifiée par drapCalc.
Exemples	StatMedian (A1:A100;1) calcule la valeur médiane de l'ensemble de données situé dans A1:A100. Les données sont continues.
Directives	données représente une plage Excel. drapCalc est un argument facultatif représenté par une valeur entière comprise entre -1 et 5. Cette valeur représente la méthode de calcul désirée du centile. -1 ou non spécifié = Automatique (selon données en entrée) 0 = Identique à la fonction de centiles Excel (continues) 1 = Interpolation - extrémités asymétriques (continues) 2 = Observation la plus proche (discrètes) 3 = Fonction dist. empirique (discrètes) 4 = Interpolation - extrémités symétriques (continues) 5 = Fonction dist. empirique avec étalement (discrètes)

MIN

Description	MIN (Données1;Données2;...DonnéesN) calcule le minimum des données spécifiées dans Données1;Données2;...DonnéesN.
Exemples	StatMin (A1:A100;{1;2;3;2,4}) calcule la valeur minimum de l'ensemble de données situé dans A1:A100 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

MODE

Description	MODE (<i>Données</i> ; <i>Discrètes</i>) calcule le mode d'un ensemble de données.
Exemples	StatMode (A1:A100; <i>FAUX</i>) calcule le mode de l'ensemble de données situé dans A1:A100. Les données sont continues.
Directives	<i>Données</i> représente la matrice ou plage de données pour laquelle le mode doit être calculé. <i>Discrètes</i> est un argument facultatif spécifiant si les données doivent être traitées comme discrètes (vrai) ou continues (faux). En son absence, sa valeur est déterminée automatiquement en fonction des données.

NORMAL

Description	NORMAL (<i>moyenne</i> ; <i>éc_type</i> ; <i>statistique</i> ; <i>valeur</i>) calcule la <i>statistique</i> pour la distribution normale spécifiée par <i>moyenne</i> et <i>éc_type</i> .
Exemples	StatNormal (2; 1; "x to p"; 3) calcule la valeur p pour une valeur x de 3 dans la distribution normale à moyenne=2 et écart type=1.
Directives	<i>moyenne</i> représente la moyenne arithmétique de la distribution. <i>éc_type</i> représente l'écart type de la distribution. Sa valeur doit être > 0. <i>statistique</i> est une valeur entière (1 à 12) ou une chaîne indiquant la statistique à renvoyer. Pour plus de détails, voir la section intitulée Fonctions de distribution en début de chapitre. <i>valeur</i> doit être >=0 et <=1 quand une valeur p est entrée.

PAIRCOUNT

Description	PAIRCOUNT (<i>Données1</i> ; <i>Données2</i>) compte le nombre de paires de cellules dans <i>Données1</i> et <i>Données2</i> . Seules les paires de cellules numériques sont comptées. Les valeurs manquantes, dans l'une ou l'autre plage, ne sont pas comptées.
Exemples	StatPairCount (A1:A100; B1:B100) compte le nombre de paires de cellules numériques dans les ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>Données1</i> et <i>Données2</i> doivent être des plages Excel de taille égale. Les paires de cellules de <i>Données1</i> et <i>Données2</i> se sélectionnent de ligne en ligne, à partir du coin supérieur gauche.

PAIRMEAN

Description	PAIRMEAN (Données1;Données2) calcule la moyenne des différences entre les paires de cellules dans <i>Données1</i> et <i>Données2</i> . Seules les paires de cellules numériques sont soumises aux calculs. Les valeurs manquantes, dans l'une ou l'autre plage, ne sont pas comptées.
Exemples	StatPairMean (A1:A100;B1:B100) calcule la moyenne des différences entre les paires de cellules numériques dans les ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>Données1</i> et <i>Données2</i> doivent être des plages Excel de taille égale. Les paires de cellules de <i>Données1</i> et <i>Données2</i> se sélectionnent de ligne en ligne, à partir du coin supérieur gauche.

PAIRMEDIAN

Description	PAIRMEDIAN (Données1;Données2;drapMéth) calcule la médiane des différences entre les paires de cellules dans <i>Données1</i> et <i>Données2</i> . Seules les paires de cellules numériques sont soumises aux calculs. Les valeurs manquantes, dans l'une ou l'autre plage, ne sont pas comptées. La médiane peut être calculée selon l'une de cinq méthodes, telle que facultativement spécifiée par <i>drapMéth</i> .
Exemples	StatPairMedian (A1:A100;B1:B100) calcule la médiane des différences entre les paires de cellules numériques dans les ensembles de données situés dans A1:A100 et B1:B100.
Directives	<i>Données1</i> et <i>Données2</i> doivent être des plages Excel de taille égale. Les paires de cellules de <i>Données1</i> et <i>Données2</i> se sélectionnent de ligne en ligne, à partir du coin supérieur gauche. <i>drapMéth</i> est un argument facultatif représenté par une valeur entière comprise entre -1 et 5. Cette valeur représente la méthode de calcul désirée pour la médiane. -1 ou non spécifié = Automatique (selon données en entrée) 0 = Identique à la fonction de centiles Excel (continues) 1 = Interpolation - extrémités asymétriques (continues) 2 = Observation la plus proche (discrètes) 3 = Fonction dist. empirique (discrètes) 4 = Interpolation - extrémités symétriques (continues) 5 = Fonction dist. empirique avec étalement (discrètes)

PAIRSTDDEV

Description	PAIRSTDDEV (Données1;Données2) calcule l'écart type d'échantillon des différences entre les paires de cellules dans <i>Données1</i> et <i>Données2</i> . Seules les paires de cellules numériques sont soumises aux calculs. Les valeurs manquantes, dans l'une ou l'autre plage, ne sont pas comptées.
Exemples	StatPairStdDev (A1:A100;B1:B100) calcule l'écart type d'échantillon des différences entre les paires de cellules numériques dans les ensembles de données situés dans <i>A1:A100</i> et <i>B1:B100</i> .
Directives	<i>Données1</i> et <i>Données2</i> doivent être des plages Excel de taille égale. Les paires de cellules de <i>Données1</i> et <i>Données2</i> se sélectionnent de ligne en ligne, à partir du coin supérieur gauche.

PERCENTILE

Description	PERCENTILE (<i>données;p;drapMéth</i>) calcule le p^{e} centile de <i>données</i> . Les centiles peuvent se calculer selon l'une de cinq méthodes, telle que facultativement spécifiée par <i>drapMéth</i> .
Exemples	StatPercentile (<i>A1:A100;0,15;0</i>) calcule le 15 ^e centile des données situées dans <i>A1:A100</i> . Les données sont continues et la méthode de calcul de moyenne pondérée est utilisée.
Directives	<i>p</i> doit être compris entre 0 et 1 (inclusif). <i>drapMéth</i> est un argument facultatif représenté par une valeur entière comprise entre -1 et 5. Cette valeur représente la méthode de calcul désirée pour le centile. -1 ou non spécifié = Automatique (selon données en entrée) 0 = Identique à la fonction de centiles Excel (continues) 1 = Interpolation - extrémités asymétriques (continues) 2 = Observation la plus proche (discrètes) 3 = Fonction dist. empirique (discrètes) 4 = Interpolation - extrémités symétriques (continues) 5 = Fonction dist. empirique avec étalement (discrètes)

PRODUCT

Description	PRODUCT (<i>Données1;Données2;...DonnéesN</i>) calcule le produit des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatProduct (<i>A1:A10;{1;2;3;2,4}</i>) calcule le produit de toutes les valeurs de l'ensemble de données situé dans <i>A1:A10</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

QUARTILE

Description	QUARTILE (<i>données;q;drapMéth</i>) calcule le quartile spécifié de <i>données</i> . Les quartiles peuvent être calculés selon l'une de cinq méthodes, telle que facultativement spécifiée par <i>drapMéth</i> .
Exemples	StatQuartile (<i>A1:A100;1;FAUX</i>) calcule le 1 ^{er} quartile des données situées dans <i>A1:A100</i> . Les données sont continues.
Directives	<i>données</i> doit être une plage Excel. <i>Q</i> représente le quartile. 0=minimum, 1=1er quartile, 2=2e quartile (médiane), 3=3e quartile, 4=maximum. <i>drapMéth</i> est un argument facultatif représenté par une valeur entière comprise entre -1 et 5. Cette valeur représente la méthode de calcul désirée pour le centile. -1 ou non spécifié = Automatique (selon données en entrée) 0 = Identique à la fonction de centiles Excel (continues) 1 = Interpolation - extrémités asymétriques (continues) 2 = Observation la plus proche (discrètes) 3 = Fonction dist. empirique (discrètes) 4 = Interpolation - extrémités symétriques (continues) 5 = Fonction dist. empirique avec étalement (discrètes)

RAND

Description	RAND () renvoie un nombre aléatoire compris entre 0 et 1. Cette fonction utilise le générateur de nombres aléatoires du logiciel Palisade @RISK et pas celui intégré d'Excel.
Exemples	StatRand () renvoie un nombre aléatoire compris entre 0 et 1.

RANGE

Description	RANGE (<i>Données1;Données2;...DonnéesN</i>) calcule l'étendue (maximum-minimum) des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatRange (<i>A1:A100;</i> {1;2;3;2,4}) calcule l'étendue (maximum-minimum) des données situées dans <i>A1:A100</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

RUNSTEST

Description	RUNSTEST (<i>données;limite</i>) calcule la statistique des runs pour les valeurs de <i>données</i> selon une valeur <i>limite</i> .
Exemples	StatRunsTest (<i>A1:A100;StatMean(A1:A100)</i>) calcule la statistique du test des runs sur les données de <i>A1:A100</i> avec la moyenne des données comme limite.
Directives	<i>données</i> doit être une plage Excel.

SKEWNESS

Description	SKEWNESS (<i>Données1;Données2;...DonnéesN</i>) calcule l'asymétrie d'échantillon des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatSkewness (<i>A1:A10;</i> {1;2;3;2,4}) calcule l'asymétrie d'échantillon de toutes les valeurs de l'ensemble de données situé dans <i>A1:A10</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

SKEWNESSP

Description	SKEWNESSP (<i>Données1;Données2;...DonnéesN</i>) calcule l'asymétrie de population des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatSkewnessP (<i>A1:A10;</i> {1;2;3;2,4}) calcule l'asymétrie de population de toutes les valeurs de l'ensemble de données situé dans <i>A1:A10</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

STANDARDIZE

Description	STANDARDIZE (<i>x</i> ; <i>moyenne</i> ; <i>éc_type</i>) calcule une valeur normalisée d'une distribution à <i>moyenne</i> et <i>écart type</i> spécifiés. <i>x</i> représente la valeur à normaliser.
Exemples	StatStandardize (2;1;3) calcule une valeur normalisée à la valeur 2 d'une distribution à moyenne 1 et écart type 3.
Directives	<i>x</i> représente la valeur à normaliser. <i>moyenne</i> représente la moyenne arithmétique de la distribution. <i>éc_type</i> représente l'écart type de la distribution. Sa valeur doit être > 0.

STDDEV

Description	STDDEV (<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i>) calcule l'écart type d'échantillon des données spécifiées dans <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> .
Exemples	StatStdDev (A1:A10;{1;2;3;2,4}) calcule l'écart type d'échantillon de toutes les valeurs de l'ensemble de données situé dans A1:A10 et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

STDDEVP

Description	STDDEVP (<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i>) calcule l'écart type de population des données spécifiées dans <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> .
Exemples	StatStdDevP (A1:A10;{1;2;3;2,4}) calcule l'écart type de population de toutes les valeurs de l'ensemble de données situé dans A1:A10 et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

SUM

Description	SUM (<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i>) calcule la somme des données spécifiées dans <i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> .
Exemples	StatSum (A1:A10;{1;2;3;2,4}) calcule la somme de toutes les valeurs de l'ensemble de données situé dans A1:A10 et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1</i> ; <i>Données2</i> ; ... <i>DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

SUMDEVSQ

Description	SUMDEVSQ (Données1;Données2;...DonnéesN) calcule la somme du carré de l'écart par rapport à la moyenne de ses arguments, qui peuvent être des nombres, des matrices ou des plages. Les nombres manquants renvoient un blanc.
Exemples	StatSumDevSq (A1:A10;{1;2;3;2,4}) calcule la somme du carré de l'écart par rapport à la moyenne de toutes les valeurs de l'ensemble de données situé dans A1:A10 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

SUMSQ

Description	SUMSQ (Données1;Données2;...DonnéesN) calcule la somme du carré de ses arguments, qui peuvent être des nombres, des matrices ou des plages. Les nombres manquants renvoient un blanc.
Exemples	StatSumSq (A1:A10;{1;2;3;2,4}) calcule la somme du carré de toutes les valeurs de l'ensemble de données situé dans A1:A10 et des valeurs 1, 2, 3 et 2,4.
Directives	Données1;Données2;...DonnéesN peuvent être des nombres, des matrices ou des plages.

STUDENT

Description	STUDENT (deg_liberté;statistique;valeur) calcule la <i>statistique</i> pour la distribution de Student entrée.
Exemples	StatStudent (5;1;"x to p";2) calcule la valeur p de la distribution de Student avec 5 degrés de liberté à la valeur x 2.
Directives	<i>deg_liberté</i> est un entier indiquant le nombre de degrés de liberté. Sa valeur doit être comprise entre 1 et 32767. <i>statistique</i> est une valeur entière (1 à 12) ou une chaîne indiquant la statistique à renvoyer. Pour plus de détails, voir la section intitulée Fonctions de distribution en début de chapitre. <i>x</i> représente la valeur numérique à laquelle la distribution doit être évaluée. Sa valeur doit être ≥ 0 .

VARIANCE

Description	VARIANCE (<i>Données1;Données2;...DonnéesN</i>) calcule la variance d'échantillon des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatVariance (<i>A1:A10;</i> {1;2;3;2,4}) calcule la variance d'échantillon de toutes les valeurs de l'ensemble de données situé dans <i>A1:A10</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

VARIANCEP

Description	VARIANCEP (<i>Données1;Données2;...DonnéesN</i>) calcule la variance de population des données spécifiées dans <i>Données1;Données2;...DonnéesN</i> .
Exemples	StatVarianceP (<i>A1:A10;</i> {1;2;3;2,4}) calcule la variance de population de toutes les valeurs de l'ensemble de données situé dans <i>A1:A10</i> et des valeurs 1, 2, 3 et 2,4.
Directives	<i>Données1;Données2;...DonnéesN</i> peuvent être des nombres, des matrices ou des plages.

Index

- Autorisation, 179
- Barres d'outils
 - StatTools, 29
- Commandes
 - À propos, 179
 - Analyse discriminante, 133
 - ANOVA double, 98
 - ANOVA simple, 95
 - Autocorrélation, 113
 - Autorisation, 179
 - Boîte à moustaches, 78
 - Chronogramme, 111
 - Combinaison, 61
 - Corrélation et covariance, 70
 - de graphique U, 154
 - Décharger StatTools, 177
 - Désempiler les variables, 52
 - Diagramme de dispersion, 76
 - Diagramme de Pareto, 140
 - Différence, 57
 - Échantillons aléatoires, 65
 - Effacer la mémoire de dialogue, 177
 - Empiler les variables, 49
 - Gestionnaire des ensembles de données, 43
 - Graphique C, 151
 - Graphique P, 147
 - Graphique Q-Q normal, 109
 - Graphiques X/R, 143
 - Histogramme, 73
 - Interaction, 59
 - Intervalle de confiance - Moyenne/Écart type, 81
 - Intervalle de confiance - Proportion, 84
 - Paramètres d'application, 171
 - Prévision, 117
 - Régression, 124
 - Régression logistique, 129
 - Retard, 55
 - Sélection de taille d'échantillon, 93
 - Supprimer les ensembles de données, 177
 - Synthèse de variable, 67
 - Test d'hypothèse – Moyenne / Écart type, 87
 - Test d'hypothèse - Proportion, 90
 - Test d'indépendance chi carré, 101
 - Test de Lilliefors, 107
 - Test de Mann-Whitney, 167
 - Test de normalité chi carré, 103
 - Test de rang de Wilcoxon, 164
 - Test des runs, 115
 - Test des signes, 161
 - Transformer les variables, 53
 - Variable nominale, 63
- Configuration requise, 6
- Désinstallation de StatTools, 7
- Fonctions StatTools, 181
 - Liste alphabétique, 185
 - StatAutocorrelation, 188
 - StatAveDev, 188
 - StatBinomial, 188
 - StatCategoryIndices, 189
 - StatCategoryNames, 189
 - StatCategoryOccurrenceCount, 189
 - StatChiDist, 190
 - StatCorrelationCoeff, 190
 - StatCount, 190
 - StatCountCategories, 191
 - StatCountCellByType, 191
 - StatCountRange, 191
 - StatCovariance, 192

StatCovarianceP, 192
StatDestack, 193
StatDurbinWatson, 193
StatF, 194
StatGetCellValues, 194
StatKurtosis, 194
StatKurtosisP, 195
StatLN, 195
StatMax, 195
StatMean, 195
StatMeanAbs, 196
StatMedian, 196
StatMin, 196
StatNormal, 197
StatPairCount, 197
StatPairMean, 198
StatPairMedian, 198
StatPairStdDev, 199
StatPercentile, 199
StatProduct, 200
StatQuartile, 200
StatRand, 200
StatRange, 201
StatRunsTest, 201
StatSkewness, 201
StatSkewnessP, 201
StatStandardize, 202

StatStdDev, 202
StatStdDevP, 202
StatStudent, 203
StatSum, 202
StatSumDevSq, 203
StatSumSq, 203
StatVariance, 204
StatVarianceP, 204

Icônes

Bureau, 8
StatTools, 29

Installation, 6

Menus

Aide, 179
Contrôle de qualité, 139
Graphiques de synthèse, 73
Inférence statistique, 81
Régression et classification,
123
Série temporelle et prévision,
111
Statistiques de synthèse, 67
Tests de normalité, 103
Tests non paramétriques, 159
Utilitaires, 171
Utilitaires de données, 49

Palisade Corporation, 5
