



Table des matières :

- 1 Introduction
 - 1 Pourquoi les tableurs sont si populaires
 - 7 Une alternative aux tableurs
 - 9 La courbe d'apprentissage avec IBM SPSS Statistics
 - 10 Conclusion
 - 11 À propos d'IBM Business Analytics
 - 11 Notes
 - 12 Autres sources
-

Les risques de l'utilisation de tableurs pour l'analyse statistique

Introduction

Les tableurs sont largement employés dans le cadre de l'analyse statistique et même si leur utilité n'est plus à démontrer, elle a cependant ses limites. Lorsqu'ils sont utilisés pour effectuer des tâches pour lesquelles ils ne sont pas conçus ou qui vont au-delà de leurs possibilités, les tableurs peuvent s'avérer dangereux.

Ce document présente quelques points qu'il peut être judicieux de prendre en compte si vous utilisez ou envisagez d'utiliser un outil de tableur pour effectuer des analyses statistiques. Il propose également une alternative qui peut s'avérer plus pertinente dans de nombreux cas.

Pourquoi les tableurs sont si populaires

Un tableur est un choix attirant lorsqu'il s'agit d'effectuer des calculs, car il est très simple d'utilisation. Nous savons tous les utiliser (ou nous pensons le savoir). De plus, ceux-ci sont fournis en tant que ressource standard avec la plupart des ordinateurs, ils sont donc directement disponibles.

Le tableur est une superbe invention et un outil excellent – pour certains travaux. Cependant, les tableurs sont trop souvent utilisés pour effectuer des tâches qui dépassent leurs capacités. Comme le dit le vieux proverbe « Tout ressemble à un clou pour qui ne possède qu'un marteau ». Certains problèmes peuvent pourtant être bien mieux réglés à l'aide d'un tournevis, d'un tube de colle ou d'une boucle de ceinture.

De plus, le sentiment qu'un tableur est un outil facile à utiliser est, dans une certaine mesure, une illusion. Il est toujours simple d'obtenir une réponse avec un tableur – mais il n'est pas forcément facile d'obtenir une réponse correcte.

La décision de se tourner vers autre chose (une technologie ou un outil inconnu) reste cependant difficile. Lorsque l'on envisage une alternative, deux questions viennent à l'esprit : Dans quelle mesure cet outil est-il utile ? Est-il difficile d'apprendre à le maîtriser ?



Les tableurs peuvent être des outils utiles pour l'analyse statistique ; mais lorsqu'ils sont utilisés lors de tâches pour lesquelles ils ne sont pas conçus, ils peuvent en fait s'avérer dangereux.

La réponse à la première question dépend de l'ampleur et de la complexité de votre analyse de données. Un tableur traditionnel pourra contenir un nombre limité d'enregistrements, donc si la tâche est de taille importante, un autre outil peut être utile.

En ce qui concerne la complexité, si vous n'avez besoin que d'un contrôle superficiel de vos données, un tableur peut être un outil convenable. Si toutefois vous pensez que vos données contiennent des informations qui ne sont pas détectables de prime abord, si vous devez effectuer une analyse détaillée ou rechercher des schémas cachés, un tableur s'avérera insuffisant.

Un autre facteur à prendre en compte est le degré de précision requis. Les résultats des tableurs peuvent manquer de fiabilité, particulièrement avec de grands ensembles de données et/ou dans le cas de calculs complexes. Si une précision absolue est requise, un tableur n'est donc pas suffisant. À la place, il est préférable d'envisager l'utilisation d'un outil plus fiable en matière de précision.

Enfin, si la tâche consiste simplement à analyser une quantité limitée de données historiques, un tableur suffit, mais si vous souhaitez effectuer des prévisions ou trouver des tendances, il existe des outils bien meilleurs, particulièrement pour travailler avec de grands ensembles de données.

Ce document va maintenant tenter de répondre à la deuxième question : est-il difficile à maîtriser les logiciels spécialisés pour les calculs statistiques ?

La création d'un tableur est une tâche aussi complexe et susceptible de générer des erreurs que la programmation informatique.

Cependant, avant de continuer, il ne sert à rien d'utiliser des tableurs pour effectuer des tâches autres que des calculs numériques. Par exemple, les tableurs sont souvent utilisés comme des bases de données, pour créer et gérer des listes. Une fois de plus, on applique les facteurs d'échelle et de complexité. Au delà de certaines limites, une base de données propre, avec des règles de structuration des données intégrées, de maintien de l'intégrité des données et de développement des audits de qualité, convient bien mieux.

Deux éléments à retenir à propos des tableurs **Les tableurs sont vraiment des programmes informatiques**

Lorsque vous concevez une disposition de tableur, vous écrivez un programme informatique. Les programmes de tableur tels que Microsoft® Excel utilisent ce que l'on appelle un « langage de programmation non procédural ». Bien qu'il soit également possible d'écrire des programmes procéduraux pour Excel dans Visual Basic, le fait de saisir tous les jours des formules dans des cellules est un exercice de programmation non procédurale.

En principe, lorsque l'on pense aux langages de programmation, on pense à BASIC, C, Java™, FORTRAN etc. ces langages sont tous des « langages procéduraux » et chacun d'entre eux possède une méthodologie cohérente développée pour des programmes dans ces langages. Il est devenu évident depuis quelques années qu'il est nécessaire d'appliquer ces règles pour que les programmes fonctionnent correctement.

Cependant, pour qu'un programme compliqué produise les bons chiffres, de grandes quantités de tests et de déboguage peuvent être nécessaires.

La programmation non procédurale est aussi remplie de décisions, de complexité et de risques de se tromper comme le plus simple des programmes procéduraux.

Avec une méthodologie de développement logiciel standard, les programmes informatiques procéduraux sont vérifiés et revérifiés. En contraste, un tableur, bien que d'une importance vitale pour l'activité d'une entreprise, est généralement le travail d'une seule personne. Il n'est presque jamais vérifié ou testé en détail et part bien souvent en production avec peu ou pas de vérification. Pourtant, certaines décisions administratives importantes, comme par exemple la prévision des revenus et les plans d'investissements futurs, se fondent sur les chiffres qu'il produit

Les études révèlent que 90 pour cent des tableurs contiennent au moins une erreur.

Les tableurs sont source d'erreur

Un certain nombre d'études ont été menées sur la fréquence des erreurs dans les tableurs. Selon ces études, il semblerait que 90% de toutes les feuilles contiennent au moins une erreur. Ces études consistaient en des inspections visuelles de tableurs stratégiques, il est donc possible que de nombreuses autres erreurs n'aient pas été trouvées. De même, on a constaté que les tentatives de correction des erreurs en amenaient souvent de nouvelles.

Brown et Gould¹ ont détecté 17 erreurs dans un tableur, dont 15 qui ont généré des chiffres erronés, 11 mauvaises formules, 2 fautes de frappe et 1 erreur d'arrondis. Dans une analyse de 88 tableurs, 94% d'entre eux contenaient des erreurs et un examen approfondi de 43 autres a révélé que 5,2% des cellules contenaient des erreurs.

Teo et Tan² ont réalisé une expérience afin de déterminer la fréquence des erreurs faites par des personnes utilisant un tableur. Ils ont montré que 30% des aberrations étaient corrigées par ceux qui apportaient une modification à un tableur existant. Cependant, cette amélioration était compensée par une augmentation du nombre total d'erreurs. Sur l'ensemble des tableurs, 49% recelaient de nouvelles erreurs mécaniques, 30% des erreurs de logique et 65% de nouvelles omissions;

D'autres études ont été réalisées et les résultats restent cohérents. Toutes ces études indiquent que plus de 90% des tableurs étudiés contenaient des erreurs. Panko³ a rassemblé les éléments de plusieurs études et le taux d'erreur le plus bas qu'il ait trouvé était de 86% et le plus haut, d'après une autre source, était de 100%.

Types d'erreurs de tableur

Les erreurs de tableur peuvent être divisées en trois types principaux.

Le type « le plus amical » est celui des erreurs fonctionnelles. Ces erreurs sont les plus faciles à trouver car elles empêchent purement et simplement le tableur de fonctionner. Plutôt que de donner des chiffres erronés, ils affichent des messages d'erreur ou des valeurs manquantes.

Les types d'erreurs présentes dans les tableurs incluent des erreurs fonctionnelles, des aberrations et des erreurs furtives, sur une échelle d'importance plus ou moins élevée.

Ensuite, il y a les aberrations. Le tableur semble fonctionner, mais les chiffres sont incorrects. Souvent, ces erreurs sont généralement repérées par un utilisateur expert du domaine et donc capable de remarquer les résultats incohérents.

Les pires erreurs sont celles qu'on appelle furtives. Elles produisent des résultats incorrects mais personne ne réalise qu'ils le sont. Ils passent les contrôles et sont validés. Les erreurs furtives apparaissent soit parce que le résultat correct n'est pas connu. (ce qui est souvent le cas avec les calculs statistiques), soit parce que les chiffres ne sont que légèrement différents du résultat attendu et semblent raisonnables. Ces erreurs peuvent perdurer plusieurs années avant d'être détectées.

Il existe quelques histoires sur les conséquences embarrassantes dues à des erreurs de tableur. Par exemple, la ville de Nevada City, en Californie, a découvert en janvier 2006 que son budget était déficitaire de cinq millions de dollars. Le tableur du budget était le même que celui employé les années précédentes, mais en entrant les données de l'année en cours, une formule avait été écrasée par inadvertance. Heureusement, il s'agissait d'une erreur aberrante qui fut rapidement remarquée par les membres du conseil municipal. Cependant, la correction du tableur a pris une journée entière au directeur financier (et c'est à ce moment qu'il a trouvé plusieurs autres erreurs).

Dans une autre anecdote datant de 2003, une université a découvert des erreurs dans les moyennes de l'année de certains étudiants. Les chiffres n'avaient tout simplement aucun sens. Après avoir refait les calculs à la main, les examinateurs ont corrigé les notes et ont fini par découvrir que l'erreur dans les équations du tableur était causée par des opérations de copier-coller qui ne parvenaient pas à prendre en compte la différence entre adressage de cellule absolu et relatif. Bien que le tableur ait été vérifié par un supérieur hiérarchique, ce dernier n'avait porté attention qu'à la première ligne, la seule qui était effectivement correcte.

Causes des erreurs de tableur

Les utilisateurs de tableurs doivent connaître les facteurs qui causent ces erreurs. Malheureusement, il y a trop de causes différentes pour en dresser une liste exhaustive ici, mais les plus courantes sont les suivantes:

- **Fautes de logique** : elles peuvent prendre une forme simple, comme l'appel d'une mauvaise fonction, une soustraction à la place d'une addition ou l'oubli d'une parenthèse lors de la création d'une formule. Ce genre d'erreur peut également être causé par des relations implicites de cellules dans le tableur.
- **Formules mal copiées** : entrer une équation tout en lisant autre part amène souvent à faire des erreurs, tout comme le copier-coller. La copie d'équations existantes vers de nouveaux emplacements modifie généralement la cellule de référence, c'est pourquoi il est important de vérifier que les bons changements sont opérés de la bonne façon.

Des erreurs se produisent dans les tableurs pour différentes raisons : erreurs de logique, formules incorrectes, formules écrasées ou mauvaise utilisation des fonctionnalités intégrées.

- **Formules accidentellement écrasées** : une cellule contenant une équation a l'apparence d'un chiffre ; ce que voit l'utilisateur à première vue n'est autre que le résultat. Donc, insérer accidentellement un chiffre dans une cellule contenant déjà une formule aura pour effet d'écraser l'équation et de transformer le contenu de la cellule en une constante. Si d'autres formules reposent sur les résultats de cette cellule, l'erreur peut s'en trouver considérablement aggravée.
- **Mauvaise utilisation de fonctions intégrées** : une fonction incorrecte peut être utilisée par erreur ; par exemple, utiliser AVERAGEA, qui évalue le texte et remplace les entrées erronées par zéro, au lieu de AVERAGE, qui les ignore. Ce genre d'erreur est malheureusement très vite arrivé.
- **Omission de facteurs** : il est très facile de tout simplement oublier quelque chose. Il peut s'agir d'une équation, de données ou des deux. Les erreurs de ce type ont lieu relativement souvent lorsque de nouvelles données sont ajoutées à une feuille de tableur existante. Il se peut que toutes les données ne soient pas entrées ou que certaines des nouvelles cellules ne soient pas incluses dans toutes les équations qui leur sont relatives.
- **Erreurs dans la saisie des données** : celles-ci peuvent induire des erreurs aberrantes mais ce n'est pas toujours le cas. Par exemple, si on entre 3,5 plutôt que 3.5, le tableur suppose qu'il s'agit d'une chaîne et non d'un chiffre. La valeur zéro est alors utilisée dans toute formule se rapportant à cette cellule. Une autre erreur serait d'entrer 3/5, reconnu comme une date, c'est-à-dire un très grand nombre dans un calcul.

Il existe de nombreuses autres possibilités. Par exemple, si vous triez une colonne contenant à la fois des nombres et des équations, vous trierez les équations en plus des nombres, ce qui peut se traduire par des erreurs de calcul.

Enfin, il y a le problème de la fiabilité du logiciel de tableur. Les développeurs de tableurs publient continuellement des correctifs pour leurs logiciels. Cependant, Gregg Keizer⁴ a en 2008 signalé qu'un correctif résolvant une erreur dans Excel causait de nouvelles erreurs lors des calculs : en fait, les erreurs étaient encore plus nombreuses qu'auparavant. (Le logiciel avait été publié cinq ans plus tôt, mais soit le problème n'avait pas été détecté, soit il l'avait été et aucune tentative de réparation n'avait été faite.) Ce n'est pas qu'une question de correctif. De nombreuses études ont montré⁵ que les tableurs ne sont tout simplement pas très précis pour les calculs mathématiques complexes, même s'ils sont programmés correctement.

De nombreux articles détaillent les erreurs des procédures statistiques d'Excel couramment utilisé même dans les enseignements statistiques. Et de nombreux sites Web mettent pourtant en évidence ses limites dans les calculs analytiques avancés (voir citations à la fin de cet article.) Pour en résumer les conclusions, une analyse sérieuse d'Excel montre qu'il peut être tout simplement dangereux de l'utiliser en analyse statistique. McCullough⁵ déclare : « les statisticiens professionnels continuent d'écrire des livres avec des titres comme « Les Statistiques avec Excel » mais ils mettent en garde leurs étudiants de ne pas parier leur avenir sur la précision d'Excel ».

Lorsque l'on utilise des tableurs, il est aisé de faire une erreur de saisie des données, mais il est souvent très difficile de localiser les problèmes et de les résoudre.

Lorsque l'on enregistre les résultats d'une enquête dans des tableurs, la représentation précise des données manquantes ou catégorielles peut s'avérer être particulièrement ardue.

Autres problèmes associés à l'utilisation des tableurs

Saisie et vérification des données

Lorsque vous saisissez des données dans un tableur, il est important de connaître le lien qui existe entre les informations d'une cellule et celles qui l'entourent. Cela signifie vérifier que les chiffres se trouvent dans une plage valide, que les formules ne sont pas écrasées par des chiffres et que les nouvelles cellules sont incluses dans toutes les formules. Il s'agit de vérifier, revérifier et revérifier encore et d'être toujours prêt à cliquer sur le bouton Annuler.

Il est facile de faire une erreur lorsqu'on saisit des chiffres, par exemple d'entrer 95,3 au lieu de 9,53. Si cela se produit, le tableau utilisera logiquement le chiffre incorrect et avec un peu de chance, le résultat sera assez étrange pour apparaître comme une erreur aberrante. Il faudra ensuite étudier le tableur, élément par élément, pour trouver la cause (ou les causes) du problème.

Prise en charge de certains types de données particuliers

Il existe plusieurs types de données, communs à de nombreux types de recherche, qui nécessitent une prise en charge particulière.

Un problème qui surgit fréquemment est de savoir comment gérer les valeurs de données manquantes. Lorsque vous utilisez des tableurs, vous devez prêter attention aux valeurs manquantes. Si vous attribuez la valeur zéro à ces données, cela risque de modifier la moyenne d'une plage de valeurs. Si vous saisissez une chaîne dans une cellule pour indiquer une valeur manquante, certaines équations ignoreront cette chaîne alors que d'autres la considéreront comme zéro. Parce que zéro est une valeur valide dans certains cas, il vous faut un autre moyen d'indiquer une valeur manquante. Cependant dans ce cas, vous devez faire attention à utiliser la désignation de la valeur manquante de manière cohérente pour toutes les données et vous devez également documenter avec soin votre approche afin que les modifications ultérieures du tableur n'invalident pas les conventions de données utilisées. Notez par ailleurs qu'aucune des approches ci-dessus n'est considérée comme une méthode approuvée de saisie fiable des valeurs manquantes.

Une autre situation particulière surgit lorsque vous devez traiter des données catégorielles (généralement rencontrées dans les résultats d'enquêtes). Par exemple, imaginons que les quatre valeurs 1, 2, 3 et 4 sont attribuées pour représenter les réponses « Oui », « Non », « Ne sait pas » et « Refuse de répondre » à la question d'une enquête. Si vous utilisez un tableur pour stocker ce type de données, vous devez avant tout documenter les valeurs et ce qu'elles signifient, vous assurer que les données sont correctement saisies (la bonne valeur attribuée à chaque réponse) et que ces données sont traitées comme il le faut. Dans le cas contraire, la pertinence des données disparaîtra dès que la personne ayant développé le tableur le fermera.

Même si certaines applications de tableurs proposent des fonctions de prévisions des tendances et des résultats, ces méthodes sont rarement fiables et manquent souvent de précision.

Les projections futures

Généralement, les tableurs sont utilisés pour extraire des informations et des relations associées à des événements passés. Cependant, de plus en plus souvent, les organisations souhaitent savoir ce qu'il peut se passer dans le futur. Les dernières versions d'Excel ont ce genre de fonctions, par exemple, PRÉVISION, TENDANCE et CROISSANCE qui permettent de prévoir de nouvelles valeurs en fonction de données existantes et toute une gamme de programmes d'extension est également disponible. Mais la question qui se pose est la fiabilité et la précision de ces fonctionnalités, qui, de toutes façons, ne proposent aucun des tests que des mathématiciens sérieux utiliseraient pour la validation des résultats.

Gestion des données

Parce qu'ils se concentrent sur les cellules, les tableurs présentent un certain nombre de problèmes quant à la gestion de données. Une modification simple, comme celle de l'heure de début, l'ajout de nouveaux membres ou un changement de formule, peut nécessiter des douzaines, voire des centaines, d'autres modifications.

Même une modification toute simple peut nécessiter des insertions ou des suppressions de cellules/lignes/colonnes, la modification ou la copie de formules dans un grand nombre de cellules ou la reconfiguration de tout le tableur. Ces opérations prennent non seulement beaucoup de temps mais elles peuvent également être source de nouvelles erreurs.

La plupart du temps, de nouvelles données doivent être ajoutées à une feuille de tableur terminée. Mais comment intégrer ces nouveaux chiffres ? Une solution est de configurer le tableur pour que les équations soient étendues et puissent intégrer toutes les nouvelles données : le problème qui se pose alors est que certaines équations pourraient malencontreusement être étendues pour intégrer des données qui ne devraient pas l'être. D'un autre côté, si le tableur est configuré pour ne pas s'étendre automatiquement, certaines données qui devraient être incluses risquent d'être oubliées. Dans les deux cas, il est peu probable que le tableur produise des résultats corrects à moins que les modifications soient vérifiées avec attention.

Une alternative aux tableurs

Jusqu'ici cet article a proposé une présentation générale des situations dans lesquelles les tableurs peuvent s'avérer inefficaces, ou peu appropriées, pour effectuer des analyses statistiques. Cela ne signifie pas qu'ils ne présentent aucun intérêt. S'il s'agit d'effectuer des tests simples sur un petit nombre de variables alors un tableur s'avère être un outil aussi efficace que bien d'autres.

Cela étant dit, un programme de tableur, comme nous l'avons mentionné auparavant, est un logiciel généraliste. Avec ou sans programmes d'extension, la gamme d'outils d'analyse est limitée et les algorithmes d'un programme de tableur ne sont pas conçus ou testés de manière aussi rigoureuse que ceux des programmes logiciels conçus spécifiquement pour les analyses statistiques.

IBM SPSS Statistics Base était auparavant nommé PASW[®] Statistics

IBM SPSS Statistics offre aux organisations la possibilité d'effectuer des analyses statistiques approfondies et robustes sans avoir recours à la programmation.

Tout comme un charpentier pourrait utiliser une scie à main pour couper une douzaine de planches de bois mais se tournerait vers des outils spécialisés pour construire des meubles et vers des outils plus puissants pour gérer le stock de bois nécessaire à la construction d'un bâtiment, toute personne souhaitant effectuer une analyse solide et détaillée devra utiliser un outil conçu spécifiquement pour cette tâche. L'un de ces outils est IBM® SPSS® Statistics* de IBM Business Analytics..

Les mécanismes intégrés de validation des données et de détection des erreurs d'IBM SPSS Statistics permettent de garantir que les données saisies sont valides et correctes.

IBM SPSS Statistics a été conçu en 1968 et n'a pas cessé d'être amélioré depuis. Durant tout ce temps, de nombreuses formes d'analyses statistiques ont été intégrées au logiciel et les algorithmes qui exécutent les équations ont été testés à la fois par les développeurs et par les utilisateurs dans les milieux universitaires, dans les laboratoires et dans tous types d'entreprises. C'est la raison pour laquelle les utilisateurs peuvent être assurés que le logiciel a été minutieusement testé et que ses résultats sont jugés comme fiables.

Sans avoir à effectuer aucune programmation, les utilisateurs peuvent utiliser une large gamme d'analyses statistiques. De plus, les besoins des utilisateurs pouvant évoluer avec le temps, ils pourront être amenés à utiliser des méthodes statistiques parmi les plus avancées, présentes dans le logiciel.

Naturellement, IBM SPSS Statistics est optimisé pour gérer les calculs statistiques comme aucun tableur ne pourrait le faire. En fait, ce logiciel est optimisé pour les travaux statistiques à tous les niveaux, de la saisie de données à la création de rapports de prise de décision.

La saisie de données selon IBM SPSS Statistics

Avec IBM SPSS Statistics, le processus de saisie de données démarre avec les définitions des types de données qui vont être utilisés. Celles-ci sont particulièrement détaillées. Par exemple, chaque type de données comporte à la fois un nom abrégé et un nom long. (Le nom le plus adapté est celui qui est utilisé pour annoter les tableaux et les graphiques). De plus, le type de données qui peut être saisi (des chiffres ou du texte, pour donner un exemple simple) peut être spécifié. À ce moment, la vérification de premier niveau a lieu. Les données doivent correspondre aux caractéristiques du type défini ou elles seront refusées. Aucun type de données ni aucune autre caractéristique de la disposition ne peut être modifié accidentellement. Vous ne pouvez pas non plus modifier les relations entre les données. La saisie de données n'est ni plus ni moins que de la saisie de données : elle est totalement indépendante de la programmation.**

Les mécanismes de validation de données et de vérification des erreurs que contient IBM SPSS Statistics sont simples à comprendre. Des procédures automatisées recherchent les valeurs qui semblent dissonantes, ce qui

*** Remarque : IBM Business Analytics, recommande l'utilisation d'un produit spécialisé tel qu'IBM® SPSS® Data Collection Data Entry ou tout autre produit de la gamme de logiciels d'enquêtes d'opinion de l'entreprise destiné à la saisie de données. En effet, ces produits sont conçus pour détecter les erreurs lors de la saisie des données. IBM SPSS Data Collection Data Entry était anciennement nommé PASW® Data Collection Data Entry.*

Les résultats des données analysées à l'aide d'IBM SPSS Statistics peuvent être extraits dans divers formats de sortie, y compris dans une large variété de diagrammes et de graphiques.

Même si les tableurs peuvent être utilisés pour effectuer certaines projections, vous avez besoin d'un outil tel qu'IBM SPSS Statistics pour prendre en compte les variables complexes.

permet de localiser la majorité des erreurs typographiques. Cependant, si la valeur semble acceptable mais paraît néanmoins présenter quelque chose d'anormal lorsqu'elle est comparée aux autres chiffres saisis, IBM SPSS Statistics la retient et vous pose la question.

Préparation des données pour l'analyse : l'approche d'IBM SPSS Statistics

Comme mentionné auparavant, il arrive souvent que les données disponibles pour l'analyse soient incomplètes. Lors d'une enquête, par exemple, certaines personnes peuvent manquer une question ou choisir de ne pas y répondre. Comme nous l'avons souligné, la gestion de données incomplètes dans un tableur pose de nombreux problèmes. Grâce à IBM SPSS Statistics, les chercheurs peuvent examiner les données disponibles et calculer des valeurs pour les éléments manquants (un processus appelé « imputation »). Ils peuvent examiner les données en utilisant un des six rapports de diagnostic pour découvrir les schémas de données manquantes. Ils peuvent également créer des résumés de statistiques et ajouter les valeurs manquantes à l'aide d'une procédure automatisée qui choisit la méthode d'imputation la plus adaptée en fonction des caractéristiques des données. Puis l'analyse peut être effectuée comme si toutes les données étaient présentes, ce qui, au sens mathématique, est le cas.

Les autres étapes de préparation des données à l'analyse comprennent l'étude de la distribution des données, la vérification des données aberrantes et l'organisation ou la répartition en classes de données afin que les algorithmes que vous souhaitez utiliser, comme les modèles Naïve Bayes ou logistiques, fonctionnent correctement. IBM SPSS Statistics effectue ces étapes de préparation des données, ce qu'aucun programme de tableur n'est conçu à faire.

Analyses statistiques avec IBM SPSS Statistics

Lorsqu'IBM SPSS Statistics passe en mode d'analyse et effectue les actions nécessaires pour produire des résultats, les données ne sont pas modifiées : elles sont utilisées comme entrées du processus et les résultats (disponibles sous différents formats, notamment une très grande quantité de diagrammes et de graphiques) sont affichés dans une autre fenêtre.

De plus, lorsqu'un type d'analyse est effectué, le logiciel génère automatiquement un programme sous la forme de syntaxe, qui peut être enregistré et exécuté à l'infini sur différents ensembles de données sans avoir besoin de le modifier. (Il peut néanmoins être modifié si nécessaire).

IBM SPSS Statistics a également l'avantage de permettre aux utilisateurs avancés de mettre en place de nouvelles procédures et fonctionnalités grâce à l'extension de sa programmabilité. Cette fonctionnalité avancée permet aux utilisateurs qui maîtrisent le langage de programmation statistique R ou Python® d'intégrer de nouveaux algorithmes ou fonctions directement dans le produit. Il est ensuite possible de concevoir une interface graphique spécifique pour cette nouvelle fonctionnalité et ainsi la rendre accessible à d'autres utilisateurs.

Prédire le futur avec IBM SPSS Statistics

Les tableurs permettent généralement d'effectuer des prédictions et, d'estimer les événements à venir en fonction de données historiques. Par exemple, un usage professionnel commun serait de prévoir les bénéfices des deux semestres à venir en fonction des résultats de l'année précédente. Bien qu'il soit possible d'effectuer ce genre de calcul à l'aide d'un tableur, tenir compte de facteurs tels que la saisonnalité des ventes ou développer des scénarios en fonction de nombreuses autres variables n'est possible qu'à l'aide d'un logiciel mathématique solide comme IBM SPSS Statistics.

IBM SPSS Statistics est un outil facile à maîtriser qui permet à votre organisation de se passer des tableurs et d'opter pour une analyse mathématiquement robuste des données complexes.

La courbe d'apprentissage avec IBM SPSS Statistics

Au début de cet article, nous avons posé la question qui surgit souvent dans la tête des gens lorsqu'ils pensent à utiliser un nouveau logiciel : Est-il difficile d'apprendre à le maîtriser ?

Dans le cas de IBM SPSS Statistics, la réponse est la suivante : « Pas difficile du tout ». Tout comme pour un tableur, ce logiciel dispose d'une interface graphique avec une présentation claire et dont les fonctionnalités sont accessibles à l'aide des menus et barres d'outils habituels. Les fonctions statistiques du programme sont regroupées de manière logique : lorsqu'une d'elles est sélectionnée, les options associées apparaissent dans une fenêtre contextuelle et le calcul est effectué à l'aide des options choisies et en cliquant le bouton « OK » ou « Exécuter ».

De plus, IBM SPSS Statistics est fourni avec un didacticiel complet, des fichiers d'aide extrêmement détaillés et des études de cas qui expliquent des exemples d'utilisation de l'analyse statistique dans des situations professionnelles et de recherche. Ces avantages peuvent transformer rapidement un statisticien débutant en analyste compétent. Bien sûr, l'entreprise propose plusieurs options de formation, notamment une formation Internet à la demande. De plus, en raison de son utilisation de longue date par des analystes dans de nombreux environnements différents, des ressources d'enseignement supplémentaires provenant de parties tierces sont disponibles, notamment des forums de discussion en ligne avec les conseils d'autres utilisateurs, des livres et des vidéos d'enseignement et des manuels d'instruction.

Conclusion

Au fur et à mesure de la lecture de cet article, vous avez découvert plusieurs choses importantes : d'abord, que les tableurs sont bien plus largement utilisés que ce que l'on croit, souvent sans prendre la peine de chercher d'autres solutions. Ensuite que le taux d'erreur dû à l'utilisation des tableurs est incroyablement élevé, notamment en comparaison du taux d'erreur accepté par les autres formes de calcul. Enfin que les tableurs sont utilisés pour traiter d'un grand nombre de problèmes différents, dont certains ne sont pas du tout adaptés aux fonctionnalités des programmes.

Votre ensemble de données est unique de même que la façon dont vous utilisez un tableur.

Pour que vous puissiez savoir si un tableur suffira à vos besoins ou si vous pourriez profiter d'un outil spécialisé tel qu'IBM SPSS Statistics, il est nécessaire que vous compreniez par vous-même comment chaque programme fonctionne avec vos données et comment il effectue les tâches analytiques dont vous avez habituellement besoin.

Tester IBM SPSS Statistics est simple : vous pouvez contacter l'entreprise ou télécharger une copie d'évaluation gratuite du logiciel sur www.spss.com/statistics. Si vos données sont déjà sous forme de tableur, IBM SPSS Statistics peut facilement les importer. Et une fois que vos données sont importées, vous pouvez évaluer les différents types d'analyses disponibles et voir s'il vous sera profitable d'utiliser un outil conçu pour les analyses statistiques au lieu d'un tableur généraliste.

À propos d'IBM Business Analytics

Les logiciels IBM Business Analytics fournissent des informations complètes, cohérentes et précises permettant aux décideurs d'améliorer les performances de leur entreprise. Un portefeuille complet de solutions de Business Intelligence, d'Analyse Prédictive, de Performance financière & gestion de la stratégie et d'Applications Analytiques permettent d'avoir une vision claire et précise de la situation actuelle et de prédire les événements futurs. Combinées à de puissantes solutions métiers, les organisations de toute taille peuvent améliorer leur productivité, optimiser leur prise de décisions et délivrer de meilleurs résultats.

Dans ce cadre, les logiciels IBM SPSS Predictive Analytics aident les organisations à prévoir les événements futurs pour une meilleure prise de décisions et donc une amélioration de leurs résultats. Nos clients des secteurs privé, public et universitaire se fient à la technologie IBM SPSS et considèrent cette technologie comme un avantage concurrentiel pour : fidéliser sa clientèle, attirer de nouveaux clients, maximiser ses ventes, réduire la fraude, minimiser les risques. En intégrant le logiciel IBM SPSS à leurs opérations quotidiennes, les organisations deviennent des entreprises prédictives capables d'optimiser leur prise de décision afin d'atteindre leurs objectifs et d'obtenir un avantage concurrentiel important.

Pour plus de renseignements, visitez le site www.ibm.com/spss/fr.

Notes

1. P. Brown, J. Gould, « An experimental study of people creating spreadsheets, » *ACM Transactions on Office Information Systems*, (1987) Vol. 5, 258-272.
2. T. Teo, M. Tan. « Quantitative and qualitative errors in spreadsheet development, » *Proceedings of the 30th Hawaii International Conference on Systems Sciences* (1997) 149-155.
3. Ray Panko, Professor, University of Hawaii. www.panko.cba.hawaii.edu/ssr.
4. Gregg Keizer, « Microsoft fixes Excel math mistake, » *Computerworld* (March 2008).
5. Bruce D. McCullough, « The Unreliability of Excel's Statistical Procedures, » *Foresight*, (February 2006) Vol. 3, 44-45.

Autres sources

www.daheiser.info/excel/frontpage.html

www.practicalstats.com/xlsstats/excelstats.html

McCullough, B.D. and Wilson, B. « On the accuracy of statistical procedures in Microsoft Excel 2003. » *Computational Statistics and Data Analysis*. (2005) Vol. 49, 1244-1252.

– « Teaching statistics with Excel 2007 and other spreadsheets. »
Computational Statistics and Data Analysis, (June 2008) Vol. 52, issue 10



© Copyright IBM Corporation 2010

IBM Corporation
Route 100
Somers, NY 10589

Droits restreints pour les utilisateurs du gouvernement américain - l'utilisation, la duplication ou la divulgation sont soumises aux restrictions visées dans le contrat GSA ADP Schedule conclu avec IBM Corp.

Produit aux États-Unis d'Amérique

Mai 2010.
Tous droits réservés

IBM, le logo d'IBM, ibm.com, WebSphere, InfoSphere et Cognos sont des marques commerciales ou des marques déposées d'International Business Machines Corporation aux États-Unis, dans d'autres pays ou les deux. Si ces termes ou si d'autres termes déposés d'IBM sont représentés pour la première fois dans ce document suivis du symbole de marque déposée (® ou ™), ces symboles indiquent des marques déposées ou de droit commun appartenant à IBM lors de la publication de ce document. Ces marques commerciales doivent également faire l'objet d'un enregistrement ou doivent être des marques de droit commun dans les autres pays. Une liste des marques commerciales actuelles d'IBM est disponible sur Internet sous « Droits d'auteur et marques » à l'adresse www.ibm.com/legal/copytrade.shtml.

SPSS est une marque commerciale de SPSS, Inc., an IBM Company, déposée dans de nombreuses juridictions dans le monde.

Les autres noms d'entreprises, de produits ou de services peuvent être des marques commerciales ou des marques de service d'autres organisations.



Veillez recycler