



Les points clés

- Examinez facilement vos données sous différents angles.
 - Diagnostiquez rapidement les problèmes de données manquantes.
 - Remplacez les valeurs manquantes par des estimations.
 - Affichez les types des données manquantes et les valeurs extrêmes.
 - Supprimez les biais cachés.
-

IBM SPSS Missing Values

Pour créer de meilleurs modèles quand vous devez remplacer les valeurs manquantes

Les créateurs d'enquêtes, les analystes de marché, les sociologues, les spécialistes de l'exploration des données (data mining) ainsi que de nombreux autres professionnels font confiance à IBM SPSS Missing Values pour valider leurs données d'étude.

Lorsque vous ignorez ou excluez les données manquantes, vous risquez d'obtenir des résultats biaisés ou non significatifs. SPSS Missing Values vous permet d'imputer ces valeurs afin de pouvoir tirer des conclusions plus fiables. SPSS Missing Values est un outil indispensable pour toutes les personnes soucieuses de la validité de leurs données. Il vous permet d'examiner facilement vos données afin de découvrir des schémas dans les données manquantes, puis d'estimer les statistiques récapitulatives et d'imputer les valeurs manquantes grâce à des algorithmes statistiques.

Cela vous permet, par exemple, d'améliorer les questions d'une enquête que vous avez identifiées comme présentant un risque de confusion d'après les schémas repérés dans les données manquantes. Grâce au tableau des pourcentages de non-concordance des schémas, vous pouvez même déterminer si les valeurs manquantes pour une variable sont liées aux valeurs manquantes pour une autre variable.

Il se peut que vous découvriez ainsi que les personnes qui ignorent une question relative à leurs revenus ignorent également une question concernant leur niveau d'éducation. Ce type d'information vous permettra d'améliorer la qualité de vos enquêtes.



Diagnostiquez facilement et rapidement les problèmes de données manquantes

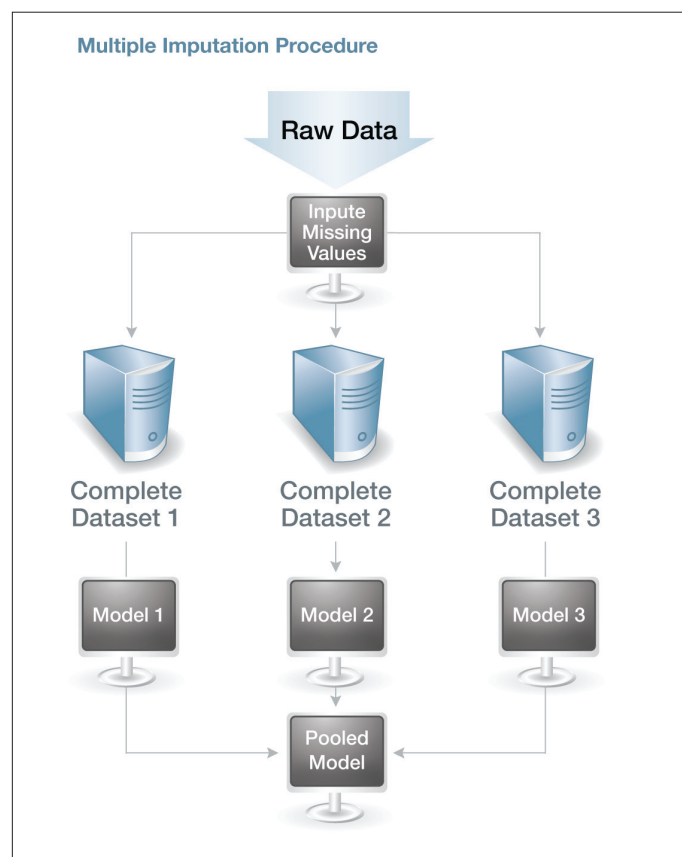
Le rapport récapitulatif global sur les valeurs manquantes vous permet de diagnostiquer rapidement l'existence d'un problème de données manquantes sérieux. Le rapport sur les schémas de valeurs manquantes vous offre ensuite une présentation cas par cas de vos données. Il affiche un instantané de chaque type de valeur manquante ainsi que, le cas échéant, les valeurs extrêmes pour chaque cas.

Utilisez ensuite les outils comme le test t de variance séparée et les tableaux croisés de variables catégorielles pour découvrir s'il existe des différences significatives entre les personnes qui ont répondu et celles qui n'ont pas répondu. Ces rapports vous aident à déterminer si les données manquantes risquent de poser des problèmes lors de l'analyse.

Quant aux rapports tabulés sur les schémas, ils présentent un récapitulatif des schémas de données manquantes et les points clés des ensembles de variables correspondants.

Utilisez l'imputation multiple pour remplacer les valeurs de données manquantes

Dans SPSS Missing Values, la procédure d'imputation multiple vous aide à comprendre les schémas d'absence de données dans votre ensemble de données et vous permet de remplacer les valeurs manquantes par des estimations plausibles. Il propose un mode d'imputation entièrement automatique, qui choisit la méthode d'imputation la plus appropriée d'après les caractéristiques de vos données tout en vous permettant de personnaliser votre modèle d'imputation.



Tout d'abord, plusieurs ensembles de données complets sont générés (3 à 5 la plupart du temps), comportant chacun un ensemble différent de valeurs de remplacement. Ensuite, vous modélisez ces ensembles à l'aide des techniques habituelles, telles que la régression linéaire, pour produire des estimations paramétriques pour chacun d'eux. Enfin, vous générez les estimations paramétriques définitives. Cette opération consiste à regrouper les différents ensembles d'estimations paramétriques obtenus à l'étape 2 et à calculer les statistiques déductives qui tiennent compte des variations au sein des différentes imputations et entre celles-ci.

L'analyse des différents ensembles de données et le regroupement des résultats sont effectués à l'aide de certaines procédures SPSS Statistics existantes, telles que la régression. Utilisées sur des ensembles de données comportant des données imputées, ces procédures existantes produiront automatiquement des estimations paramétriques regroupées.

De meilleures conclusions

Remplacez les valeurs manquantes par des estimations et augmentez vos chances d'obtenir des résultats significatifs d'un point de vue statistique. Supprimez les biais cachés de vos données en remplaçant les valeurs manquantes par des estimations afin d'inclure tous les groupes dans votre analyse – même ceux présentant un faible taux de réponse.

SPSS Missing Values fait partie de la suite logicielle IBM SPSS Statistics. Les produits de cette suite sont disponibles individuellement ainsi que sous forme de trois éditions : IBM SPSS Statistics Standard, IBM SPSS Statistics Professional et IBM SPSS Statistics Premium. En regroupant les fonctionnalités essentielles, ces éditions offrent un moyen efficace de garantir que toute votre équipe ou service dispose des outils nécessaires pour effectuer les analyses qui contribuent au succès de votre entreprise.

SPSS Missing Values est disponible sous forme d'application cliente autonome, mais aussi en version serveur pour de meilleures performances et une plus grande évolutivité.

La collaboration, pour générer encore plus de valeur

Pour partager et réutiliser efficacement les ressources, pour les protéger selon les exigences de conformité internes et externes et publier les résultats afin qu'un plus grand nombre d'utilisateurs puisse les consulter et interagir avec eux, envisagez d'enrichir votre logiciel IBM SPSS Statistics avec IBM SPSS Collaboration and Deployment Services. Vous trouverez des informations supplémentaires sur ces fonctionnalités sur le site suivant : ibm.com/spss/cds

Fonctionnalités

Analyse des schémas

- Grâce au tableau des schémas de données, visualisez les données manquantes et les valeurs extrêmes pour tous les cas et toutes les variables :
 - Visualisez les données manquantes définies par le système ainsi que trois types de valeurs manquantes définies par l'utilisateur.
 - Triez par ordre croissant ou décroissant.
 - Visualisez les valeurs réelles des variables spécifiées.
- Grâce au tableau des schémas d'absence, visualisez les schémas de valeurs manquantes pour tous les cas qui présentent au moins une valeur manquante :
 - Regroupez les schémas de valeurs manquantes similaires.
 - Triez par variables et schémas d'absence.
 - Visualisez les valeurs réelles des variables spécifiées.
- Grâce au tableau du test t de variance séparée, déterminez pour une variable liée les différences entre les groupes avec et sans valeurs manquantes :
 - test t, degrés de liberté, moyenne, valeur et nombre.
- Grâce au tableau de distribution des variables catégorielles, visualisez les différences entre les données présentes et manquantes pour les variables catégorielles :
 - Générez des tableaux croisés montrant les données produits et les données manquantes pour chaque catégorie d'une variable pour chacune des autres variables.

- Grâce au tableau des pourcentages de non-concordance des schémas, évaluez quelle quantité de données manquantes d'une variable correspond à la quantité de données manquantes d'une autre variable :
 - Triez les matrices par variables ou schémas de valeurs manquantes.
- Grâce au tableau des schémas, qui récapitule chaque schéma de données manquantes et affiche le nombre pour chacun d'eux ainsi que les moyennes et les fréquences pour chaque variable, identifiez tous les schémas uniques :
 - Grâce au tableau récapitulatif des schémas de valeurs manquantes, visualisez le nombre et les moyennes pour chacun de ces schémas.

Statistiques

- Univariée : Calculez le nombre, la moyenne, l'écart type et l'erreur standard de la moyenne pour tous les cas, sauf ceux contenant des valeurs manquantes, les nombres, le pourcentage de valeurs manquantes et les valeurs extrêmes pour toutes les variables.
- Listwise (par liste) : Calculez la moyenne, la matrice de covariance et la matrice de corrélation pour toutes les variables quantitatives pour tous les cas, en excluant les valeurs manquantes.
- Pairwise (par paires) : Calculez la fréquence, la moyenne, la variance, la matrice de covariance et la matrice de corrélation.

Imputation multiple

- Spécifiez quelles variables doivent être imputées ainsi que les contraintes qui régissent les valeurs imputées, telles que les valeurs minimum et maximum, et spécifiez quelles variables doivent être utilisées comme variables prédictives lors de l'imputation des valeurs manquantes des autres variables.
- Imputez les valeurs des variables catégorielles et continues : la régression logistique est utilisée pour les variables catégorielles, et la régression linéaire, pour les variables continues, et la mise en correspondance prédictive avec la moyenne est une option pour les résultats continus, ce qui garantit que des valeurs raisonnables (c'est-à-dire dans la plage des valeurs d'origine) sont imputées.
- La détection des schémas de données manquantes aide à déterminer quelle méthode d'imputation utiliser.
- Trois méthodes d'imputation sont proposées :
 - Monotone : une méthode efficace pour les données qui présentent un schéma d'absence monotone.
 - FCS (spécification entièrement conditionnelle) : une méthode MCMC (Markov Chain Monte Carlo) itérative appropriée lorsque les données présentent un schéma d'absence arbitraire (monotone ou non).
 - Automatique : analyse les données pour déterminer la meilleure méthode d'imputation (monotone ou FCS).
- Spécifiez :
 - le nombre d'imputations ;
 - la plage des valeurs imputées ;
 - si les effets d'interaction doivent être utilisés lors de l'imputation ;
 - facultatif : la désactivation de l'imputation pour les variables qui présentent un pourcentage plus élevé de valeurs manquantes ;
 - les niveaux de tolérance, pour rechercher les singularités.
- Spécifiez une variable contenant les pondérations d'analyse (régression) : cette procédure intègre les pondérations d'analyse dans les modèles de régression et de discrimination utilisés pour imputer les valeurs manquantes. Ces pondérations sont également utilisées dans les récapitulatifs des valeurs imputées (c'est-à-dire la moyenne, l'écart type et l'erreur standard).
- Vous pouvez visualiser un récapitulatif global des absences dans vos données, un récapitulatif d'imputation et le modèle d'imputation pour chaque variable dont les valeurs sont imputées, ainsi qu'une analyse des valeurs manquantes par variable et un tableau des schémas des valeurs manquantes, ou bien des statistiques descriptives des valeurs imputées.
- Résumez visuellement les absences pour les cas, les variables et les valeurs de données (cellules) individuelles.
- Obtenez sur demande un fichier de données SPSS Statistics contenant les valeurs imputées et/ou un historique d'itération FCS.
- Les ensembles de données générés par l'imputation multiple peuvent être analysés à l'aide des procédures d'analyse prises en charge, afin d'obtenir des estimations paramétriques définitives (combinées) qui tiennent compte de l'incertitude inhérente à ces différents ensembles.

Analyse

- Procédures d'analyse prises en charge pour l'imputation multiple (remarque : vous devez posséder le module qui contient cette procédure.)
- Procédures descriptives : fréquences, descriptives, tableaux croisés, corrélations*, corrélation non paramétrique, corrélation partielle*.
- Comparaison des moyennes : moyennes, test t, tests non paramétriques, ANOVA à 1 facteur, ANOVA univariée.
- Modèles : modèles linéaires généraux, modèles linéaires généralisés, régression linéaire*, régression logistique multinomiale*, régression logistique binaire, analyse discriminante*, régression ordinale, modèles linéaires mixtes.
- Techniques d'analyse de survie : régression de Cox.

Regroupement

- Regroupement des sorties : les sorties sont regroupées selon l'un des deux niveaux de regroupement pour produire un paramètre regroupé.
- Diagnostics de regroupement :
 - Augmentation relative de la variance : mesure de la variabilité relative de l'estimation des paramètres sur les différentes imputations.
 - Fraction des informations manquantes : augmentation relative de la variance mise à l'échelle sous forme de proportion. Une mesure de l'incertitude due à l'absence de réponse.
 - Efficacité relative : efficacité de l'estimation pour M imputations relative à celle pour un nombre infini d'imputations.
- Générez automatiquement les modèles PMML pour les estimations paramétriques regroupées : régression linéaire, modèles linéaires généralisés, régression logistique multinomiale*, régression logistique binaire, analyse discriminante, régression de Cox.

Configuration système requise

La configuration dépend de la plateforme. Pour plus d'informations, consultez le site [ibm.com/spss/requirements](https://www.ibm.com/spss/requirements)

À propos d'IBM Business Analytics

Les logiciels IBM Business Analytics permettent aux décideurs de disposer des précieux éclairages dont ils ont besoin pour améliorer les performances métier. IBM propose à cet effet une gamme complète et unifiée d'applications d'aide à la décision, d'analyse prédictive avancée, de pilotage de la stratégie et des performances financières, de gouvernance et de gestion des risques et de la conformité.

Avec les logiciels IBM, les entreprises peuvent non seulement détecter les tendances, les schémas récurrents et les anomalies, comparer des scénarios de simulation et prédire les menaces et opportunités potentielles, mais aussi planifier, élaborer les budgets et prévoir les ressources nécessaires. Grâce aux puissantes fonctions analytiques dont ils disposent, nos clients à travers le monde sont à même de mieux comprendre, anticiper et maîtriser leurs résultats métier.

Pour plus d'informations

Pour plus d'informations, visitez le site

ibm.com/business-analytics/fr

Je souhaite être contacté(e)

Pour demander à être appelé ou pour poser une question, accédez au site ibm.com/business-analytics/fr

Un représentant IBM vous répondra sous deux jours ouvrés.



Compagnie IBM France

17 Avenue de l'Europe
92 275 Bois-Colombes Cedex

La page d'accueil d'IBM est accessible à l'adresse suivante :

ibm.com

IBM, le logo IBM, ibm.com et SPSS sont des marques d'International Business Machines aux États-Unis et/ou dans certains autres pays. Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés.

Une liste actualisée de toutes les marques d'IBM est disponible sur la page Web « Copyright and trademark information » à l'adresse suivante :

ibm.com/legal/copytrade.shtml

Le présent document contient des informations qui étaient en vigueur et valides à la date de la première publication et qui peuvent être modifiées par IBM à tout moment. Toutes les offres mentionnées ne sont pas distribuées dans tous les pays où IBM exerce son activité.

LES INFORMATIONS DU PRÉSENT DOCUMENT SONT FOURNIES « EN L'ÉTAT » ET SANS GARANTIE EXPLICITE OU IMPLICITE D'AUCUNE SORTE. IBM DÉCLINE NOTAMMENT TOUTE RESPONSABILITÉ RELATIVE À CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DÉFAUT D'APTITUDE À L'EXÉCUTION D'UN TRAVAIL DONNÉ. Les produits IBM sont garantis conformément aux dispositions des contrats avec lesquels ils sont fournis.

© Copyright IBM Corporation 2013



Pensez à recycler ce document