

Guide de l'utilisateur de IBM SPSS
Modeler 14.2



Remarque : avant d'utiliser ces informations et le produit dont elles traitent, consultez les informations générales dans la rubrique Avis sur p. 269.

Ce document contient des informations sur la propriété de SPSS Inc., an IBM Company. Il est fourni sous accord de licence et est protégé par la loi sur le copyright. Les informations contenues dans cette publication n'incluent pas les garanties de produit et aucune des provisions de ce manuel ne devra être interprétée comme tel.

Lorsque vous envoyez des informations à IBM ou SPSS, vous octroyez à IBM et SPSS le droit non exclusif d'utiliser ou de distribuer ces informations de la façon dont IBM ou SPSS le juge approprié sans aucune obligation envers vous.

© **Copyright IBM Corporation 1994, 2011..**

Préface

IBM® SPSS® Modeler est le puissant utilitaire de Data mining de IBM Corp.. SPSS Modeler aide les entreprises et les organismes à améliorer leurs relations avec les clients et les citoyens grâce à une compréhension approfondie des données. À l'aide des connaissances plus précises obtenues par le biais de SPSS Modeler, les entreprises et les organismes peuvent conserver les clients rentables, identifier les opportunités de vente croisée, attirer de nouveaux clients, détecter les éventuelles fraudes, réduire les risques et améliorer les services gouvernementaux.

L'interface visuelle de SPSS Modeler met à contribution les compétences professionnelles de l'utilisateur, ce qui permet d'obtenir des modèles prédictifs plus efficaces et de trouver des solutions plus rapidement. SPSS Modeler dispose de nombreuses techniques de modélisation, telles que les algorithmes de prévision, de classification, de segmentation et de détection d'association. Une fois les modèles créés, l'utilisateur peut utiliser IBM® SPSS® Modeler Solution Publisher pour les remettre aux responsables, où qu'ils se trouvent dans l'entreprise, ou pour les transférer vers une base de données.

A propos de IBM Business Analytics

Le logiciel IBM Business Analytics fournit des informations complètes, cohérentes et précises que les preneurs de décision utilisent avec confiance pour améliorer la performance du marché. Un portefeuille étendu d'outils de [business intelligence](#), d'[analyses prédictives](#), de [performance financière et de gestion de stratégie](#), et des [applications analytiques](#) offre des connaissances claires, immédiates et applicables pour améliorer l'efficacité actuelle ainsi que la capacité de prévoir les résultats futurs. Combinées avec de riches solutions industrielles, des pratiques éprouvées et des services professionnels, les organisations de toutes tailles peuvent atteindre la productivité la plus élevée, automatiser des décisions en toute tranquillité et fournir de meilleurs résultats.

Dans le cadre de ce portefeuille, le logiciel IBM SPSS Predictive Analytics aide les organisations à prévoir des événements futurs et à agir en conséquence pour mener à de meilleurs résultats. Des clients dans le domaine commercial, gouvernemental et académique à travers le monde font confiance à la technologie IBM SPSS et considèrent qu'elle représente un avantage compétitif pour attirer, retenir et ajouter des clients, tout en réduisant la fraude et en atténuant les risques. En incorporant le logiciel IBM SPSS dans leur opérations quotidiennes, les organisations deviennent des entreprises prédictives – capables de diriger et d'automatiser les décisions pour atteindre les buts qu'ils se sont fixés et obtenir des avantages compétitifs sensibles. Pour informations supplémentaires ou pour joindre un revendeur, visitez le site <http://www.ibm.com/spss>.

Assistance technique

L'assistance technique est à la disposition des clients pour la maintenance des produits. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour joindre l'assistance technique, consultez le site Web de IBM Corp. à l'adresse <http://www.ibm.com/support>. Lorsque vous contactez l'assistance technique, n'oubliez pas de préparer vos identifiants, le nom de votre société et votre contrat d'assistance.

Contenu

1	<i>A propos de IBM SPSS Modeler</i>	1
	IBM SPSS Modeler Server	1
	IBM SPSS Modeler Options	2
	IBM SPSS Text Analytics	2
	Documentation de IBM SPSS Modeler	2
	Exemples d'application	4
	Dossier Demos	4
2	<i>Nouvelles fonctionnalités</i>	6
	Nouvelles fonctions et fonctions modifiées dans IBM SPSS Modeler 14.2.	6
	Nouveaux nœuds dans IBM SPSS Modeler 14.2	7
3	<i>Présentation de IBM SPSS Modeler</i>	8
	Démarrage	8
	Démarrage de IBM SPSS Modeler	8
	Lancement de l'application à partir de la ligne de commande.	9
	Connexion au IBM SPSS Modeler Server.	9
	Changement de répertoire temporaire	13
	Démarrage de plusieurs sessions IBM SPSS Modeler	14
	Interface IBM SPSS Modeler en un clin d'oeil	14
	Espace de travail de flux IBM SPSS Modeler	15
	Palette de noeuds	15
	Gestionnaires IBM SPSS Modeler	16
	Projets IBM SPSS Modeler	18
	Barre d'outils IBM SPSS Modeler	19
	Personnalisation de la barre d'outils	20
	Personnalisation de la fenêtre IBM SPSS Modeler	21
	Utilisation de la souris dans IBM SPSS Modeler	22
	Utilisation de touches de raccourci	22
	Impression	23
	Automatisation de IBM SPSS Modeler	24

4 Compréhension du Data mining 25

Présentation du Data mining	25
Evaluation des données	27
Stratégie pour le Data mining	28
Méthodologie CRISP-DM	29
Types de modèle.	30
Exemples de Data mining	37

5 Création de flux 38

Présentation de la création de flux	38
Création de flux de données	38
Utilisation des noeuds	39
Utilisation des flux	51
Descriptions de flux	65
Exécution des flux	68
Utilisation des modèles	69
Ajout de commentaires et d'annotations à des noeuds et à des flux	70
Enregistrement des flux de données	80
Chargement de fichiers	83
Mappage de flux de données.	84
Astuces et raccourcis.	89

6 Traitement des valeurs manquantes 93

Description des valeurs manquantes	93
Traitement des valeurs manquantes	94
Traitement des enregistrements contenant des valeurs manquantes	95
Traitement des champs contenant des valeurs manquantes.	95
Attribution ou remplacement de valeurs manquantes	96
Fonctions CLEM pour les valeurs manquantes	97

7 Création d'expressions CLEM 99

A propos de CLEM	99
CLEM Exemples	102

Valeurs et types de données	104
Expressions et conditions	105
Paramètres de flux, de session et de super noeud	106
Utilisation de chaînes	107
Traitement des valeurs non renseignées et des valeurs manquantes	108
Utilisation de nombres	109
Utilisation de l'heure et de la date	109
Récapitulation de plusieurs champs	110
Utilisation des données à réponses multiples	112
Générateur de formules	113
Accès au Générateur de formules	114
Création d'expressions	114
Sélection de fonctions	115
Sélection de Champs, Paramètres et Variables globales	116
Affichage ou sélection de valeurs	116
Vérification des expressions CLEM	118
Rechercher et remplacer	118

8 Référence du langage CLEM 122

Présentation de la référence CLEM.	122
Types de données CLEM	122
Entiers	123
Réels	123
Caractères	124
Chaînes	124
Listes	124
Champs	124
Dates	125
Heure	126
Opérateurs CLEM	126
Informations de référence sur les fonctions	128
Conventions pour la description des fonctions	130
Fonctions d'informations	130
Fonctions de conversion	131
Fonctions de comparaison	132
Fonctions logiques	134
Fonctions numériques	135
Fonctions trigonométriques	136
Fonctions de probabilité	137
Opérations d'entiers sur les bits	137

Fonctions aléatoires	139
Fonctions sur chaînes	139
Fonctions SoundEx	144
Fonctions date et heure	145
Fonctions séquentielles	149
Fonctions globales	155
Traitement des valeurs nulles et non renseignées	156
Champs spéciaux	157

9 Utilisation de IBM SPSS Modeler avec un référentiel 159

A propos du IBM SPSS Collaboration and Deployment Services Repository	159
Stockage et de déploiement des objets IBM SPSS Collaboration and Deployment Services Repository	161
Connexion au IBM SPSS Collaboration and Deployment Services Repository	162
Saisie des informations d'identification pour IBM SPSS Collaboration and Deployment Services Repository	163
Accès au contenu du IBM SPSS Collaboration and Deployment Services Repository	164
Stockage des objets dans le IBM SPSS Collaboration and Deployment Services Repository	166
Définition des propriétés d'un objet	166
Stockage des flux	172
Stockage de projets	172
Stockage des noeuds	173
Stockage des objets de sortie	173
Stockage des modèles et des palettes de modèles	174
Extraction d'objets du IBM SPSS Collaboration and Deployment Services Repository	175
Choix d'un objet à extraire	176
Sélection de la version d'objet	177
Recherche d'objets dans le IBM SPSS Collaboration and Deployment Services Repository	177
Modification des objets IBM SPSS Collaboration and Deployment Services Repository	180
Création, changement de nom et suppression de dossiers	180
Verrouillage et déverrouillage d'objets du IBM SPSS Collaboration and Deployment Services Repository	181
Suppression d'objets du IBM SPSS Collaboration and Deployment Services Repository	181
Gestion des propriétés des objets du IBM SPSS Collaboration and Deployment Services Repository	182
Affichage des propriétés d'un dossier	182
Affichage et modification des propriétés des objets	184
Gestion des étiquettes de version d'objet	186
Déploiement de flux	187
Options de déploiement des flux	188
La branche de scoring	192

10 Exportation vers des applications externes 200

À propos de l'exportation vers des applications externes	200
Ouverture d'un flux dans IBM SPSS Modeler Advantage	201
Assistant des applications prédictives 4.x.	201
Avant d'utiliser l'assistant des applications prédictives	202
Exportation de prévisions binaires en tant que scores de propension.	203
Etape 1 : Présentation de l'assistant des applications prédictives	204
Etape 2 : Sélection d'un noeud terminal	204
Etape 3 : Sélection d'un noeud UCV	205
Etape 4 : Spécification d'un logiciel	207
Etape 5 : Création du logiciel	207
Etape 6 : Récapitulatif	209
Importation et exportation de modèles au format PMML.	209
Types de modèle prenant en charge le format PMML.	212

11 Projets et rapports 214

Introduction aux projets	214
Mode CRISP-DM	215
Mode Classes	216
Création d'un projet	216
Création d'un projet	217
Ajout à un projet	217
Transfert de projets vers le IBM SPSS Collaboration and Deployment Services Repository	219
Définition des propriétés d'un projet	220
Annotation d'un projet	221
Propriétés de l'objet.	223
Fermeture d'un projet	224
Génération d'un rapport	224
Enregistrement et exportation de rapports générés	228

12 Personnaliser IBM SPSS Modeler 230

Options de personnalisation de IBM SPSS Modeler	230
Définition des options de IBM SPSS Modeler	230
Options du système	230
Définition des répertoires par défaut	231

Définition des options utilisateur	232
Définition des informations utilisateur	241
Personnalisation de la palette Noeuds	242
Personnalisation du gestionnaire de palette.	242
Modification de l’affichage d’un onglet Palette	248
Gestion du noeud CEMI	248

13 Remarques relatives aux performances des flux et des noeuds **249**

Ordre des noeuds	249
Caches de noeud	250
Performances : noeuds d’exécution	252
Performances : Noeuds de modélisation.	254
Performances : Expressions CLEM	254

Annexes

A Accessibilité dans IBM SPSS Modeler **255**

Présentation de l’accessibilité dans IBM SPSS Modeler.	255
Types de prise en charge pour l’accessibilité	255
Accessibilité pour les utilisateurs ayant une déficience visuelle.	255
Accessibilité pour les utilisateurs aveugles	256
Accessibilité via le clavier	257
Utilisation d’un lecteur d’écran	265
Conseils d’utilisation.	266
Interférence avec d’autres logiciels.	267
JAWS et Java	267
Utilisation des graphiques dans IBM SPSS Modeler	267

B Prise en charge d’Unicode **268**

Prise en charge d’Unicode dans IBM SPSS Modeler	268
---	-----

C Avis

269

Index

272

A propos de IBM SPSS Modeler

IBM® SPSS® Modeler est un ensemble d'outils de data mining qui vous permet de développer rapidement, grâce à vos compétences professionnelles, des modèles prédictifs et de les déployer dans des applications professionnelles afin de faciliter la prise de décision. Conçu autour d'un modèle confirmé, le modèle CRISP-DM, SPSS Modeler prend en charge l'intégralité du processus de Data mining, des données à l'obtention de meilleurs résultats commerciaux.

SPSS Modeler propose différentes méthodes de modélisation issues des domaines de l'apprentissage automatique, de l'intelligence artificielle et des statistiques. Les méthodes disponibles dans la palette Modélisation vous permettent d'extraire de nouvelles informations de vos données et de développer des modèles prédictifs. Chaque méthode possède ses propres avantages et est donc plus adaptée à certains types de problème spécifiques.

Il est possible d'acquérir SPSS Modeler comme produit autonome ou de l'utiliser en combinaison avec SPSS Modeler Server. Plusieurs autres options sont également disponibles, telles que décrites dans les sections suivantes. Pour plus d'informations, consultez <http://www.ibm.com/software/analytics/spss/products/modeler/>.

IBM SPSS Modeler Server

Grâce à une architecture client/serveur, SPSS Modeler adresse les demandes d'opérations très consommatrices de ressources à un logiciel serveur puissant. Il offre ainsi des performances accrues sur des ensembles de données plus volumineux. D'autres mises à jour ou produits que ceux mentionnés ici peuvent également être disponibles. Pour plus d'informations, consultez <http://www.ibm.com/software/analytics/spss/products/modeler/>.

SPSS Modeler. La version SPSS Modeler intègre toutes les fonctions du produit. Elle est installée et exécutée sur l'ordinateur de bureau de l'utilisateur. Pour obtenir de meilleures performances lors du traitement d'ensembles de données volumineux, vous pouvez l'exécuter en mode local, comme produit autonome, ou en mode réparti, en association avec IBM® SPSS® Modeler Server.

SPSS Modeler Server. SPSS Modeler Server s'exécute en continu en mode d'analyse réparti, conjointement avec une ou plusieurs installations IBM® SPSS® Modeler, obtenant ainsi de meilleures performances lors du traitement des ensembles de données volumineux. En effet, les opérations consommatrices de mémoire s'effectuent sur le serveur : inutile de télécharger les données sur l'ordinateur client. SPSS Modeler Server offre également une prise en charge de l'optimisation SQL et des capacités de modélisation au sein de la base de données, ce qui présente de nouveaux avantages en matière de performances et d'automatisation. Pour lancer les analyses, vous devez disposer d'au moins une installation SPSS Modeler.

IBM SPSS Modeler Options

Vous pouvez acheter et obtenir une licence pour les fonctions et composants suivants à utiliser avec SPSS Modeler. Notez que des produits ou des mises à jour supplémentaires deviennent parfois disponibles. Pour plus d'informations, consultez <http://www.ibm.com/software/analytics/spss/products/modeler/>.

- Accès à SPSS Modeler Server, assurant une meilleure extensibilité et des performances accrues pour les ensembles de données volumineux, ainsi qu'une prise en charge de l'optimisation SQL et des capacités de modélisation au sein de la base de données.
- SPSS Modeler Solution Publisher, pour le scoring en temps réel ou automatique en dehors de l'environnement SPSS Modeler. [Pour plus d'informations, reportez-vous à la section IBM SPSS Modeler Solution Publisher dans le chapitre 2 dans IBM SPSS Modeler 14.2 Solution Publisher.](#)
- Adaptateurs permettant le déploiement dans IBM SPSS Collaboration and Deployment Services ou dans l'application client léger IBM SPSS Modeler Advantage. [Pour plus d'informations, reportez-vous à la section Stockage et et déploiement des objets IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 161.](#)

IBM SPSS Text Analytics

IBM® SPSS® Text Analytics est un complément totalement intégré pour SPSS Modeler, qui utilise des technologies linguistiques avancées et le traitement du langage naturel pour traiter rapidement un grand nombre de données texte non structurées, extraire et organiser les concepts-clés et grouper ces concepts en catégories. Les concepts extraits et les catégories peuvent ensuite être combinés aux données structurées existantes, telles que les données démographiques, et appliqués à la modélisation grâce à la gamme complète d'outils de Data mining de IBM® SPSS® Modeler, afin de favoriser une prise de décision précise et efficace.

- Le Text Mining offre une modélisation des concepts et des catégories, ainsi qu'un utilitaire interactif où vous pouvez exécuter une exploration avancée des liens textuels et des classes, créer vos propres catégories et affiner les modèles de ressources linguistiques.
- De nombreux formats d'importation sont pris en charge, y compris les blogs et sources Web.
- Enfin, des modèles personnalisés, des bibliothèques et des dictionnaires spécialisés dans des domaines précis, tels que la gestion de la relation client et la génomique, sont fournis.

Remarque : Une licence distincte est requise pour accéder à ce composant. Pour plus d'informations, consultez <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Documentation de IBM SPSS Modeler

Une documentation complète au format d'aide en ligne est disponible dans le menu Aide de SPSS Modeler. Vous y trouverez la documentation de SPSS Modeler, SPSS Modeler Server et de SPSS Modeler Solution Publisher, ainsi que le Guide des applications et d'autres documentations utiles.

La documentation complète de chaque produit au format PDF est disponible dans le dossier *Documentation* de chaque DVD de produit.

- **Guide de l'utilisateur IBM SPSS Modeler.** Introduction générale à SPSS Modeler : création de flux de données, traitement des valeurs manquantes, création d'expressions CLEM, utilisation des projets et des rapports et regroupement des flux pour le déploiement dans IBM SPSS Collaboration and Deployment Services, des applications prédictives ou IBM SPSS Modeler Advantage.
- **Noeuds de Source, d'exécution et de sortie IBM SPSS Modeler.** Descriptions de tous les noeuds utilisés pour lire, traiter et renvoyer les données de sortie dans différents formats. En pratique, cela signifie tous les noeuds autres que les noeuds de modélisation.
- **IBM SPSS Modeler Noeuds de modélisation.** Description de tous les noeuds utilisés pour créer des modèles de Data mining. IBM® SPSS® Modeler propose différentes méthodes de modélisation issues des domaines de l'apprentissage automatique, de l'intelligence artificielle et des statistiques. [Pour plus d'informations, reportez-vous à la section Description des noeuds de modélisation dans le chapitre 3 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)
- **Guide des Algorithmes IBM SPSS Modeler.** Descriptions des fondements mathématiques des méthodes de modélisation utilisées dans SPSS Modeler.
- **Guide des applications IBM SPSS Modeler.** Les exemples de ce guide fournissent des introductions brèves et ciblées aux méthodes et techniques de modélisation. Un version en ligne de ce guide est également disponible dans le menu Aide. [Pour plus d'informations, reportez-vous à la section Exemples d'application sur p. 4.](#)
- **Génération de scripts et automatisation IBM SPSS Modeler.** Informations sur l'automatisation du système via la génération de scripts, y compris les propriétés permettant de manipuler les noeuds et les flux.
- **IBM SPSS Modeler Guide de déploiement.** Informations sur l'exécution des scénarios et des flux SPSS Modeler comme étapes des tâches d'exécution sous IBM® SPSS® Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler CLEF Guide du développeur.** CLEF permet d'intégrer des programmes tiers tels que des programmes de traitement de données ou des algorithmes de modélisation en tant que noeuds dans SPSS Modeler.
- **Guide d'exploration de base de données IBM SPSS Modeler.** Informations sur la manière de tirer parti de la puissance de votre base de données pour améliorer les performances et étendre la gamme des fonctions analytiques via des algorithmes tiers.
- **IBM SPSS Modeler Server et Guide des performances.** Informations sur le mode de configuration et d'administration de IBM® SPSS® Modeler Server.
- **Guide de l'utilisateur de IBM SPSS Modeler Administration Console.** Informations concernant l'installation et l'utilisation de l'interface utilisateur de la console permettant de surveiller et de configurer SPSS Modeler Server. La console est implémentée en tant que plug-in à l'application Deployment Manager.
- **Guide IBM SPSS Modeler Solution Publisher.** SPSS Modeler Solution Publisher est un module complémentaire qui permet aux entreprises de publier des flux destinés à être utilisés en dehors de l'environnement SPSS Modeler.
- **Guide CRISP-DM IBM SPSS Modeler** Guide détaillé sur l'utilisation de la méthodologie CRISP-DM pour le Data mining avec SPSS Modeler

Exemples d'application

Tandis que les outils de Data mining de SPSS Modeler peuvent vous aider à résoudre une grande variété de problèmes commerciaux et organisationnels, les exemples d'application fournissent des introductions brèves et ciblées aux méthodes et aux techniques de modélisation. Les ensembles de données utilisés ici sont beaucoup plus petits que les énormes entrepôts de données gérés par certains Data miners, mais les concepts et les méthodes impliqués doivent pouvoir être adaptés à des applications réelles.

Vous pouvez accéder aux exemples en cliquant Exemples d'application dans le menu Aide de SPSS Modeler. Les fichiers de données et les flux d'échantillons sont installés dans le dossier *Demos*, sous le répertoire d'installation du produit. [Pour plus d'informations, reportez-vous à la section Dossier Demos sur p. 4.](#)

Exemples de modélisation de bases de données. Consultez les exemples dans le *IBM SPSS ModelerGuide d'exploration de base de données*.

Exemples de génération de scripts. Consultez les exemples dans le *IBM SPSS ModelerGuide de génération de scripts et d'automatisation*.

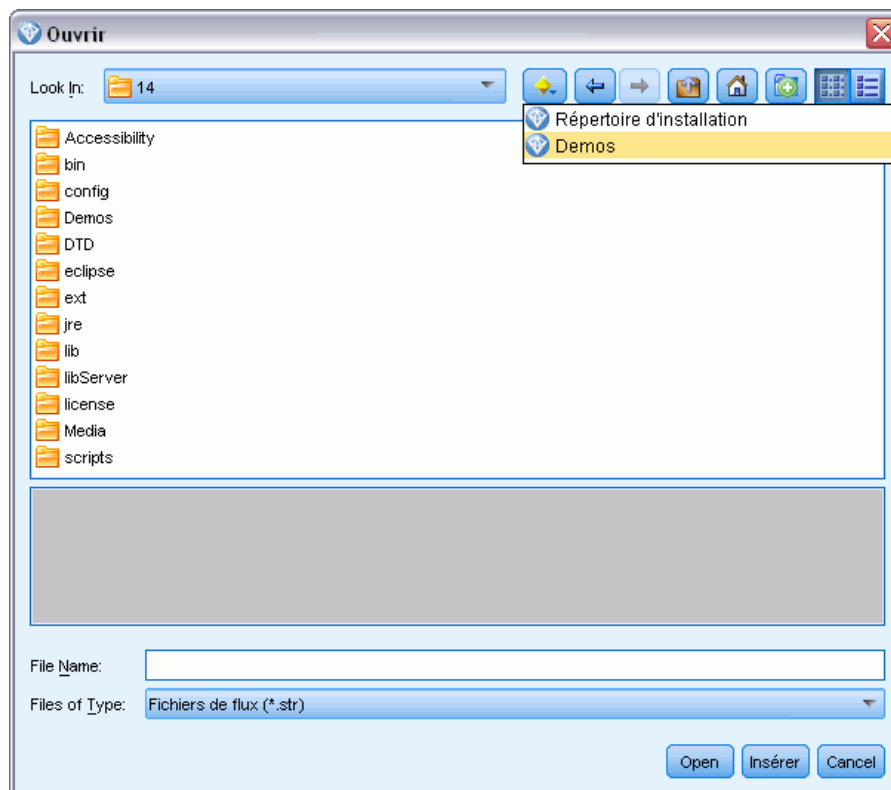
Dossier Demos

Les fichiers de données et les flux d'échantillons utilisés avec les exemples d'application sont installés dans le dossier *Demos*, sous le répertoire d'installation du produit. Ce dossier est également accessible à partir du groupe de programmes sous IBM SPSS Modeler 14.2 dans le

menu Démarrer de Windows, ou en cliquant sur *Demos* dans la liste des répertoires récents de la boîte de dialogue Ouverture de fichier.

Figure 1-1

Sélection du dossier *Demos* dans la liste des répertoires récemment consultés



Nouvelles fonctionnalités

Nouvelles fonctions et fonctions modifiées dans IBM SPSS Modeler 14.2

La version IBM® SPSS® Modeler 14.2 ajoute les fonctionnalités ci-dessous.

Prise en charge de l'exploration de base de données pour IBM Netezza Analytics. Cette version prend en charge l'exploration de base de données pour la technologie IBM Netezza Analytics. [Pour plus d'informations, reportez-vous à la section IBM SPSS Modeler et IBM Netezza Analytics dans le chapitre 6 dans *Guide d'exploration de base de données IBM SPSS Modeler 14.2*.](#)

Cette prise en charge est fournie à travers deux nouveaux noeuds de modélisation de base de données : Arbres décision Netezza et Netezza K-Means. [Pour plus d'informations, reportez-vous à la section Nouveaux nœuds dans IBM SPSS Modeler 14.2 sur p. 7.](#)

Niveau 1 de prise en charge pour les bases de données IBM Netezza. Les bases de données IBM Netezza sont désormais prises en charge au niveau 1, ce qui signifie que toutes les répercussions SQL possibles sont maintenant disponibles, avec une optimisation SQL spécifique à la base de données. [Pour plus d'informations, reportez-vous à la section Accès aux données dans le chapitre 2 dans *Administration et guide des performances de IBM SPSS Modeler Server 14.2*.](#)

Prise en charge améliorée pour la génération SQL dans le noeud Echantillonner. La prise en charge pour la génération SQL dans le noeud Echantillonner, lors de l'utilisation d'échantillonnage simple, a été améliorée pour des bases de données DB2 exécutées sur une plateforme Windows ou UNIX, et pour des bases de données IBM Netezza et Teradata. [Pour plus d'informations, reportez-vous à la section Noeuds prenant en charge la génération SQL dans le chapitre 6 dans *Administration et guide des performances de IBM SPSS Modeler Server 14.2*.](#)

Prise en charge de Cognos 10. Les noeuds source et les noeuds d'exportation IBM Cognos BI prennent désormais en charge Cognos BI version 10.1 en plus de la version 8.4. Pour basculer de la version 8.4 (la version par défaut) vers la version 10.1, exécutez l'utilitaire Cognos 10.1 approprié (fichier de commandes Windows ou script de l'interface UNIX) disponible sur le site Web de support technique de l'entreprise à l'adresse <http://www.ibm.com/support>. Si vous voulez ensuite utiliser de nouveau la version 8.4, vous aurez besoin d'exécuter l'utilitaire Cognos 8.4 approprié, disponible au même emplacement.

Support côté serveur pour IBM System z. SPSS Modeler a ajouté un support côté serveur pour les systèmes IBM System z qui exécutent Red Hat Enterprise Linux, Red Hat Enterprise Linux Advanced Platform, ou SuSE Linux Enterprise Server. Pour plus d'informations, consultez les *Instructions d'installation de IBM SPSS Modeler Server 14.2 pour UNIX* fournies sur le DVD IBM® SPSS® Modeler Server.

Nouveaux nœuds dans IBM SPSS Modeler 14.2



Le nœud Arbres de décision Netezza crée un arbre de décision que vous pouvez utiliser pour prévoir ou classer les observations futures. La méthode utilise la technique de partition récursive afin de diviser les données d'apprentissage en segments en réduisant l'index d'impureté à chaque étape, un nœud de l'arbre étant considéré comme "pur" si 100 % de ses observations appartiennent à une catégorie spécifique (« étiquette de classe ») du champ cible. [Pour plus d'informations, reportez-vous à la section Arbres décision Netezza dans le chapitre 6 dans *Guide d'exploration de base de données IBM SPSS Modeler 14.2*.](#)



Le nœud Netezza K-Means effectue une analyse de classes, et vous permet de diviser les membres d'un ensemble de données en groupes de mêmes enregistrements. L'algorithme effectue plusieurs itérations du même processus de base dans lesquelles chaque instance d'apprentissage est attribuée à la classe la plus proche. Tous les centres de classe sont alors recalculés en tant que vecteurs de valeur d'attribut moyenne des instances attribuées à des classes spécifiques. [Pour plus d'informations, reportez-vous à la section Netezza K-means dans le chapitre 6 dans *Guide d'exploration de base de données IBM SPSS Modeler 14.2*.](#)

Présentation de IBM SPSS Modeler

Démarrage

En tant qu'application de Data mining, IBM® SPSS® Modeler constitue une méthode stratégique de recherche de relations utiles dans les grands ensembles de données. Contrairement aux méthodes statistiques plus traditionnelles, il n'est pas indispensable de savoir ce que vous recherchez avant de commencer. Vous pouvez explorer vos données, créer divers modèles et explorer diverses relations, jusqu'à ce que vous trouviez des informations utiles.

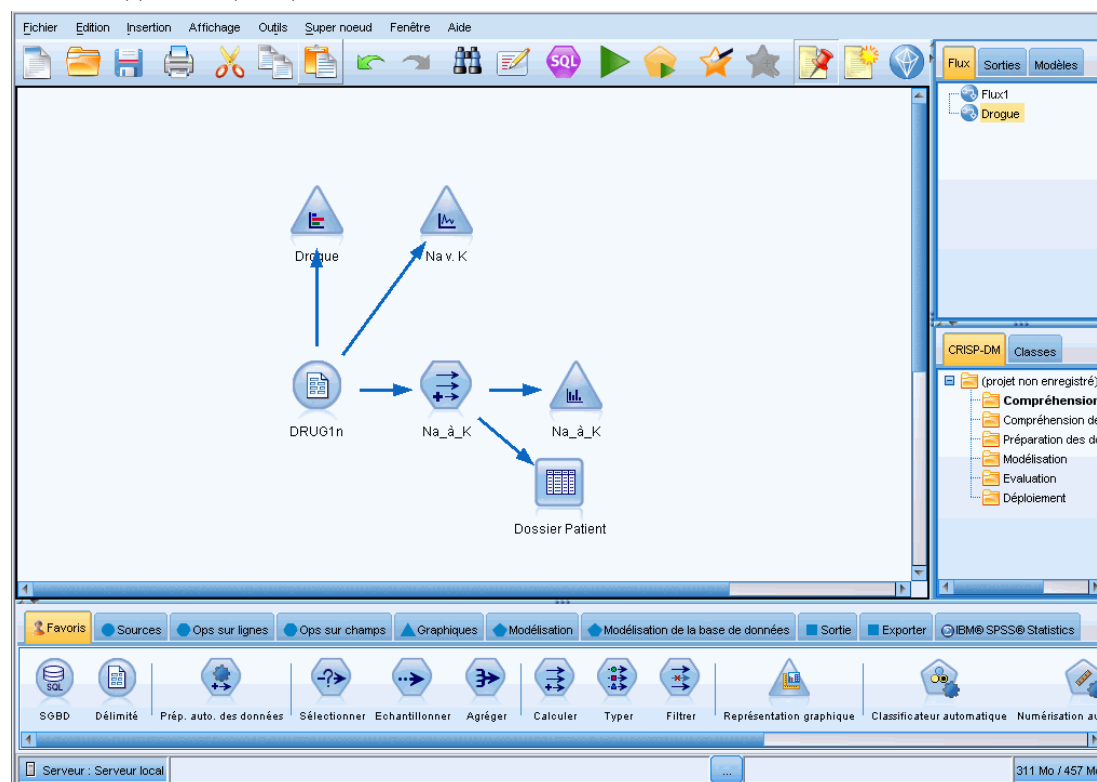
Démarrage de IBM SPSS Modeler

Pour démarrer l'application, cliquez :

Début > Programmes [tout] > IBM SPSS Modeler14.2 > IBM SPSS Modeler14.2

La fenêtre principale apparaît après quelques secondes.

Figure 3-1
fenêtre d'application principale de IBM SPSS Modeler



Lancement de l'application à partir de la ligne de commande

Vous pouvez utiliser la ligne de commande de votre système d'exploitation pour lancer IBM® SPSS® Modeler comme suit :

- ▶ Dans le cas d'un ordinateur sur lequel est installé IBM® SPSS® Modeler, ouvrez une fenêtre DOS ou une invite de commande.
- ▶ Pour lancer l'interface SPSS Modeler en mode interactif, tapez la commande `modelerclient` suivie des arguments souhaités, par exemple :

```
modelerclient -stream report.str -execute
```

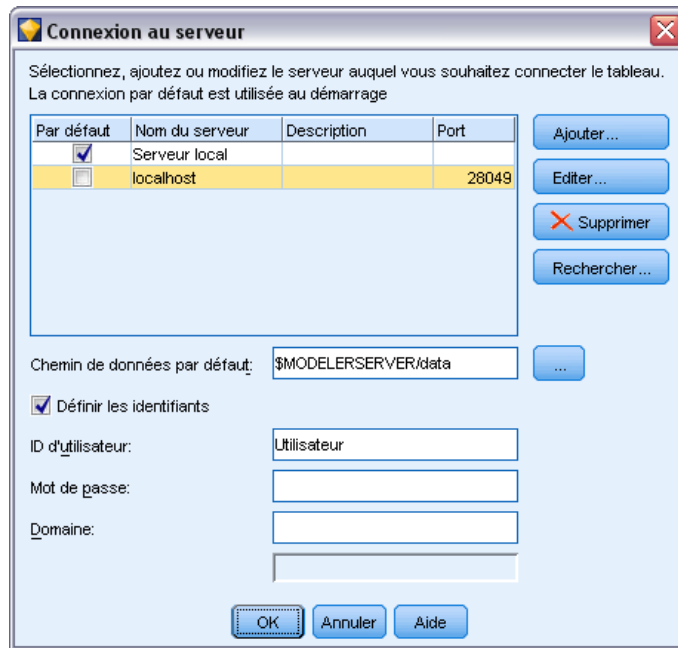
Les arguments disponibles (drapeaux) vous permettent de vous connecter à un serveur, de charger des flux, d'exécuter des scripts, ou d'indiquer les autres paramètres nécessaires.

Connexion au IBM SPSS Modeler Server

Il est possible d'exécuter IBM® SPSS® Modeler comme une application autonome ou un comme un client connecté directement à IBM® SPSS® Modeler Server ou à SPSS Modeler Server or à un groupe de serveurs par le biais du Coordinateur des processus connecté à partir de IBM® SPSS® Collaboration and Deployment Services. L'état de la connexion apparaît en bas à gauche de la fenêtre SPSS Modeler.

Lorsque vous souhaitez vous connecter à un serveur, vous pouvez saisir manuellement son nom ou sélectionner un nom que vous aurez préalablement défini. En revanche, si vous avez IBM SPSS Collaboration and Deployment Services, vous avez la possibilité de chercher dans une liste de serveurs ou de groupes de serveurs à partir de la boîte de dialogue Connexion au serveur. Vous pouvez naviguer via les services Statistics s'exécutant sur un réseau grâce au Coordinateur des processus. [Pour plus d'informations, reportez-vous à la section Equilibrage de charge avec classes de serveur dans l'annexe D dans Administration et guide des performances de IBM SPSS Modeler Server 14.2.](#)

Figure 3-2
Boîte de dialogue Connexion au serveur



Pour vous connecter à un serveur

- ▶ Dans le menu Outils, cliquez sur Connexion au serveur. La boîte de dialogue Connexion au serveur s'affiche. Vous pouvez également double-cliquer sur la zone d'état de la connexion dans la fenêtre SPSS Modeler.
- ▶ Dans la boîte de dialogue, indiquez les options de connexion à l'ordinateur du serveur local ou sélectionnez une connexion dans le tableau.
 - Cliquez sur Ajouter ou Modifier pour ajouter ou modifier une connexion. [Pour plus d'informations, reportez-vous à la section Ajout et modification d'une connexion à IBM SPSS Modeler Server sur p. 11.](#)
 - Cliquez sur Rechercher pour accéder au serveur ou à un groupe de serveurs dans le Coordinateur de processus. [Pour plus d'informations, reportez-vous à la section Recherche de serveurs dans IBM SPSS Collaboration and Deployment Services sur p. 12.](#)

Tableau de serveur. Ce tableau comprend un ensemble de connexions au serveur définies. Il affiche la connexion par défaut, le nom du serveur, sa description et le numéro du port. Vous pouvez ajouter manuellement une nouvelle connexion ainsi que sélectionner ou rechercher une connexion existante. Pour définir un serveur particulier comme connexion par défaut, cochez la case dans la colonne Par défaut du tableau de la connexion.

Chemin de données par défaut. Indiquez le chemin d'accès aux données situées sur l'ordinateur serveur. Cliquez sur le bouton ... pour accéder à l'emplacement requis.

Définir les informations d'identification. Laissez cette case décochée pour permettre à la fonction de **connexion unique** de se connecter au serveur à l'aide de vos informations de nom d'utilisateur et de mot de passe locaux. Si la connexion unique n'est pas disponible, ou si vous cochez la case pour désactiver la connexion unique (par exemple pour vous connecter à un compte administrateur), les champs suivants sont activés et vous permettent d'entrer vos informations d'identification.

Nom d'utilisateur. Entrez le nom d'utilisateur avec lequel effectuer la connexion au serveur.

Mot de passe : Entrez le mot de passe associé au nom d'utilisateur défini.

Domaine. Indiquez le domaine utilisé pour la connexion au serveur. Le nom de domaine n'est requis que si l'ordinateur serveur se trouve dans un autre domaine Windows que l'ordinateur client.

- ▶ Cliquez sur OK pour terminer la connexion.

Pour se déconnecter d'un serveur

- ▶ Dans le menu Outils, cliquez sur Connexion au serveur. La boîte de dialogue Connexion au serveur s'affiche. Vous pouvez également double-cliquez sur la zone d'état de la connexion dans la fenêtre SPSS Modeler.
- ▶ Dans la boîte de dialogue, sélectionnez le serveur local puis cliquez sur OK.

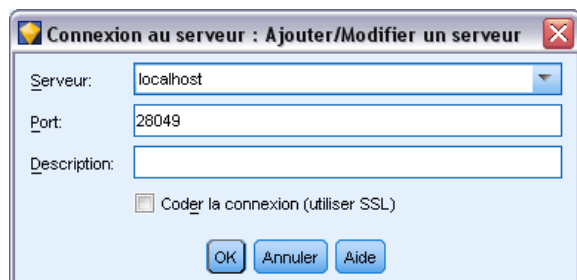
Ajout et modification d'une connexion à IBM SPSS Modeler Server

Dans la boîte de dialogue Connexion au serveur, vous pouvez modifier ou ajouter une connexion au serveur. Cliquez sur Ajouter pour accéder à une boîte de dialogue Ajouter/Modifier un serveur non renseignée, dans laquelle vous pourrez entrer les données de la connexion au serveur. Si vous sélectionnez une connexion existante et cliquez sur Modifier, dans la boîte de dialogue Connexion au serveur, la boîte de dialogue Ajouter/Modifier un serveur s'ouvre, affichant les données de cette connexion, vous permettant ainsi d'apporter toutes les modifications que vous souhaitez.

Remarque : Vous ne pouvez pas modifier une connexion au serveur qui a été ajoutée à partir de IBM® SPSS® Collaboration and Deployment Services, car le nom, le port et d'autres détails sont définis dans IBM SPSS Collaboration and Deployment Services.

Figure 3-3

Boîte de dialogue Ajouter/Modifier un serveur pour la connexion au serveur



Pour ajouter des connexions au serveur

- ▶ Dans le menu Outils, cliquez sur Connexion au serveur. La boîte de dialogue Connexion au serveur s'affiche.

- ▶ Dans la boîte de dialogue, cliquez sur Ajouter. La boîte de dialogue Ajouter/Modifier un serveur pour la connexion au serveur s'ouvre.
- ▶ Saisissez les données de connexion au serveur puis cliquez sur OK pour enregistrer la connexion et retourner à la boîte de dialogue Connexion au serveur.
 - **Serveur.** Indiquez un serveur disponible ou sélectionnez-en un dans la liste. L'ordinateur serveur peut être identifié par un nom alphanumérique (par exemple, *monserveur*) ou une adresse IP qui lui est affectée (par exemple, 202.123.456.78).
 - **Port.** Indiquez le numéro de port d'écoute du serveur. Si ce port par défaut ne fonctionne pas, demandez à l'administrateur système le numéro de port correct.
 - **Description.** Saisissez une description optionnelle pour la connexion à ce serveur.
 - **Coder la connexion (utiliser SSL).** Indique si une connexion SSL (**Secure Sockets Layer**) doit être utilisée. Le protocole SSL est fréquemment utilisé pour la sécurisation des données sur un réseau. Pour pouvoir utiliser cette fonction, vous devez activer le protocole SSL sur le serveur hébergeant le IBM® SPSS® Modeler Server. Si nécessaire, contactez votre administrateur local pour plus d'informations.

Pour modifier des connexions au serveur

- ▶ Dans le menu Outils, cliquez sur Connexion au serveur. La boîte de dialogue Connexion au serveur s'affiche.
- ▶ Dans la boîte de dialogue, sélectionnez la connexion que vous souhaitez modifier puis cliquez sur Modifier. La boîte de dialogue Ajouter/Modifier un serveur pour la connexion au serveur s'ouvre.
- ▶ Modifiez les données de connexion au serveur puis cliquez sur OK pour enregistrer les changements et retourner à la boîte de dialogue Connexion au serveur.

Recherche de serveurs dans IBM SPSS Collaboration and Deployment Services

Au lieu d'entrer manuellement une connexion au serveur, vous pouvez sélectionner un serveur ou un groupe de serveurs disponible sur le réseau par le biais du Coordinateur de processus, disponible dans IBM® SPSS® Collaboration and Deployment Services. Un groupe de serveurs contient plusieurs serveurs, et permet au Coordinateur de processus de déterminer le serveur qui répond le mieux à la demande de traitement. [Pour plus d'informations, reportez-vous à la section Equilibrage de charge avec classes de serveur dans l'annexe D dans Administration et guide des performances de IBM SPSS Modeler Server 14.2.](#)

Bien que vous ne vous puissiez pas ajouter manuellement de serveurs dans la boîte de dialogue Connexion au serveur, la recherche de serveurs disponibles vous permet de vous connecter aux serveurs sans que vous ayez besoin de connaître le nom du serveur et le numéro du port. Ces informations sont fournies automatiquement. Il vous faut néanmoins corriger les données de connexion telles que le nom de l'utilisateur, le domaine et le mot de passe.

Remarque : Si vous n'avez pas accès au Coordinateur de processus, vous pouvez tout de même saisir manuellement le nom du serveur auquel vous souhaitez vous connecter ou sélectionner un nom que vous aurez défini au préalable. [Pour plus d'informations, reportez-vous à la section Ajout et modification d'une connexion à IBM SPSS Modeler Server sur p. 11.](#)

Figure 3-4
Boîte de dialogue Recherche de serveurs



Pour rechercher des serveurs et des groupes de serveurs

- ▶ Dans le menu Outils, cliquez sur Connexion au serveur. La boîte de dialogue Connexion au serveur s'affiche.
- ▶ Cliquez sur Rechercher pour ouvrir la boîte de dialogue Recherche de serveurs. Si vous n'êtes plus connecté à IBM SPSS Collaboration and Deployment Services, lors de votre tentative d'accès au Coordinateur de processus, il vous sera demandé de vous reconnecter. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 162.](#)
- ▶ Sélectionnez le serveur ou le groupe de serveurs dans la liste.
- ▶ Cliquez sur OK pour fermer la boîte de dialogue et ajouter cette connexion au tableau de la boîte de dialogue Connexion au serveur.

Changement de répertoire temporaire

Certaines opérations effectuées par IBM® SPSS® Modeler Server peuvent nécessiter la création de fichiers temporaires. Par défaut, IBM® SPSS® Modeler crée les fichiers temporaires dans le répertoire temporaire du système. Vous pouvez modifier l'emplacement du répertoire temporaire en effectuant les opérations suivantes.

- ▶ Créez un répertoire intitulé *spss* et un sous-répertoire intitulé *servertemp*.
- ▶ Editez le fichier *options.cfg*, situé dans le répertoire */config* du dossier d'installation de SPSS Modeler. Editez le paramètre *temp_directory* de ce fichier en saisissant : *temp_directory, "C:/spss/servertemp"*.
- ▶ Redémarrez ensuite le service SPSS Modeler Server. Pour ce faire, cliquez sur Services dans les outils d'administration du Panneau de configuration Windows. Il vous suffit d'arrêter le service et de le redémarrer pour appliquer les modifications apportées. Vous pouvez redémarrer l'ordinateur pour redémarrer le service.

Tous les fichiers temporaires sont désormais écrits dans ce nouveau répertoire.

Remarque : L'erreur la plus courante est de ne pas utiliser le type correct de barre oblique. En raison de l'historique UNIX de SPSS Modeler, les barres obliques sont utilisées.

Démarrage de plusieurs sessions IBM SPSS Modeler

Si vous devez lancer plus d'une session IBM® SPSS® Modeler à la fois, vous devez effectuer certaines modifications de votre IBM® SPSS® Modeler et des paramètres Windows. Par exemple, il vous faudra effectuer ces modifications si vous avez deux licences de serveur distinctes et que vous souhaitez exécuter deux flux pour deux serveurs distincts du même ordinateur client.

Pour activer plusieurs sessions SPSS Modeler :

- ▶ Cliquez :
Début > Programmes [tout] > IBM SPSS Modeler14.2
- ▶ Dans le raccourci de IBM SPSS Modeler14.2 (celui avec l'icône), cliquez avec le bouton droit de la souris et sélectionnez Propriétés.
- ▶ Dans la zone de texte Cible, ajoutez -noshare à la fin de la chaîne.
- ▶ Dans Windows Explorer, sélectionnez :
Outils > Options des dossiers...
- ▶ Dans l'onglet Types de fichier, sélectionnez l'option Flux de SPSS Modeler et cliquez sur Avancé.
- ▶ Dans la boîte de dialogue Modifier le type de fichier, sélectionnez Ouvrir avec SPSS Modeler et cliquez sur Modifier.
- ▶ Dans la zone de texte Application utilisée pour effectuer l'action, ajoutez -noshare avant l'argument -flux.

Interface IBM SPSS Modeler en un clin d'oeil

A chaque étape du processus de Data mining, l'interface simplifiée de IBM® SPSS® Modeler met à contribution vos compétences professionnelles. Les algorithmes de modélisation, comme la prévision, la classification, la segmentation et la détection d'association, permettent l'obtention de modèles performants et précis. Les résultats du modèle peuvent facilement être déployés et lus dans des bases de données, dans IBM® SPSS® Statistics et dans de nombreuses autres applications.

Travailler avec SPSS Modeler est un processus en trois étapes de travail avec les données.

- Pour commencer, lisez les données de SPSS Modeler.
- Ensuite, exécutez les données par une série de manipulations.
- Et pour finir, envoyez les données vers une destination choisie.

Cette séquence d'opérations est appelée **flux de données** car les données circulent, enregistrement par enregistrement, de la source à la destination (modèle ou type de sortie de données), en passant par chaque manipulation.

Figure 3-5
Un flux simple



Espace de travail de flux IBM SPSS Modeler

L'espace de travail de flux est la plus grande zone de la fenêtre IBM® SPSS® Modeler. C'est dans cette zone que vous créez et manipulez les flux de données.

Les flux sont créés en dessinant des diagrammes des opérations de données nécessaires à votre entreprise sur l'espace de travail principal de l'interface. Chaque opération est représentée par une icône ou un **noeud**, et les noeuds sont reliés entre eux dans un **flux** représentant le flux de données passant par chaque opération.

Vous pouvez utiliser plusieurs flux à la fois dans SPSS Modeler, que ce soit dans le même espace de travail de flux ou par l'ouverture d'un nouveau flux. Au cours d'une session, les flux sont stockés dans le gestionnaire de flux, en haut à droite de la fenêtre SPSS Modeler.

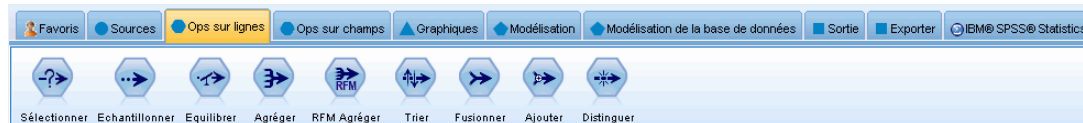
Palette de noeuds

La plupart des données et des outils de modélisation de IBM® SPSS® Modeler se trouvent dans la **Palette de noeuds**, au bas de la fenêtre sous l'espace de travail de flux.

Par exemple, l'onglet de la palette Ops sur lignes contient des noeuds permettant d'effectuer des opérations sur les **lignes**, telles que la sélection, la fusion et l'ajout.

Pour ajouter des noeuds à l'espace de travail, double-cliquez sur les icônes de la palette de noeuds ou faites glisser les icônes vers l'espace de travail. Vous pouvez ensuite les relier afin de créer un **flux** représentant le flux des données.

Figure 3-6
Onglet Ops sur lignes de la palette de noeuds



Chaque onglet de palette contient un ensemble de noeuds connexes employés pour différentes étapes des opérations de flux, comme :

- **Sources.** Les noeuds amènent les données dans SPSS Modeler.

- **Ops sur lignes.** Noeuds utilisés pour les opérations sur les **lignes** de données, comme la sélection, la fusion et l'ajout.
- **Ops sur champs.** Noeuds utilisés pour les opérations sur les **champs** de données, comme le filtrage, le calcul de nouveaux champs et la détermination du niveau de mesure de champs particuliers.
- **Graphiques.** Noeuds utilisés pour visualiser les données avant et après la modélisation. Les graphiques peuvent être des nuages, des histogrammes, des noeuds Relations, ainsi que des graphiques d'évaluation.
- **Modélisation.** Noeuds utilisant les algorithmes de modélisation disponibles dans SPSS Modeler, comme les réseaux de neurones, les arbres décision, les algorithmes de classification et la création de séquences de données.
- **Modélisation de bases de données.** Les noeuds utilisent les algorithmes de modélisation disponibles dans les bases de données Microsoft SQL Server, IBM DB2 et Oracle.
- **Résultats.** Les noeuds produisent diverses sorties pour les données, les diagrammes et les résultats de modèles qui peuvent être affichés dans SPSS Modeler.
- **Exporter.** Les noeuds produisent diverses sorties qui peuvent être affichées dans des applications externes telles que IBM® SPSS® Data Collection ou Excel.
- **SPSS Statistics.** Les noeuds importent des données à partir de, ou exportent des données vers IBM® SPSS® Statistics, et exécutent aussi des procédures SPSS Statistics.

Au fur et à mesure que vous maîtrisez mieux l'application SPSS Modeler, vous pouvez personnaliser le contenu de la palette en fonction de vos besoins. [Pour plus d'informations, reportez-vous à la section Personnalisation de la palette Noeuds dans le chapitre 12 sur p. 242.](#)

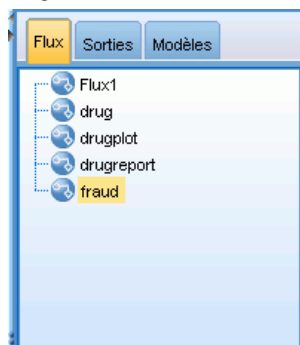
Situé sous la palette de noeuds, un panneau de rapports fournit des informations sur la progression des diverses opérations, telles que la lecture des données dans le flux de données. Egalement situé sous la palette de noeuds, un panneau d'état fournit des informations sur l'activité actuelle de l'application, ainsi que des indications lorsqu'une saisie par l'utilisateur est requise.

Gestionnaires IBM SPSS Modeler

En haut à droite de la fenêtre se trouve le panneau des gestionnaires. Il contient trois onglets qui permettent de gérer les flux, les sorties et les modèles.

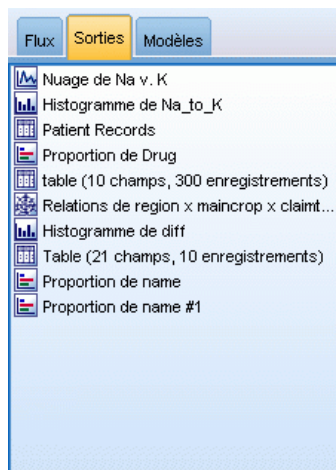
Vous pouvez utiliser l'onglet Flux pour ouvrir, renommer, enregistrer et supprimer les flux créés dans une session.

Figure 3-7
Onglet Flux



L'onglet Sortie contient différents fichiers, tels que des graphiques et des tableaux, produits par des opérations de flux dans IBM® SPSS® Modeler. Vous pouvez afficher, sauvegarder, renommer et fermer les tableaux, les graphiques et les rapports qui figurent dans cet onglet.

Figure 3-8
Onglet Sorties



L'onglet Modèle est le plus puissant des onglets du gestionnaire. Cet onglet contient tous les **nuggets** de modèle qui contiennent les modèles générés dans SPSS Modeler, pour la session en cours. Vous pouvez accéder à ces modèles directement à partir de l'onglet Modèles ou les ajouter au flux dans l'espace de travail.

Figure 3-9
Onglet Modèles qui contient des nugget de modèles

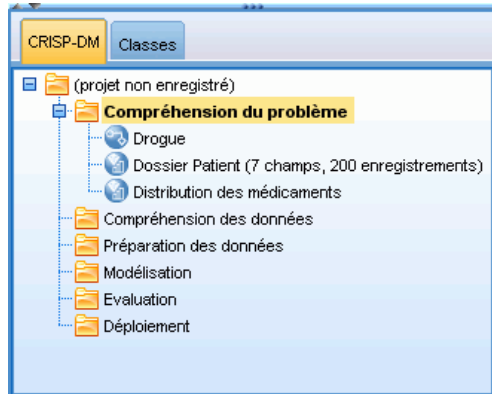


Projets IBM SPSS Modeler

Dans la partie inférieure droite de la fenêtre se trouve le panneau de projet qui permet de créer et de gérer les **projets** de Data mining (groupes de fichiers en rapport avec une tâche de data mining). Vous pouvez afficher les projets créés de deux façons dans IBM® SPSS® Modeler— en mode Classes et en mode CRISP-DM.

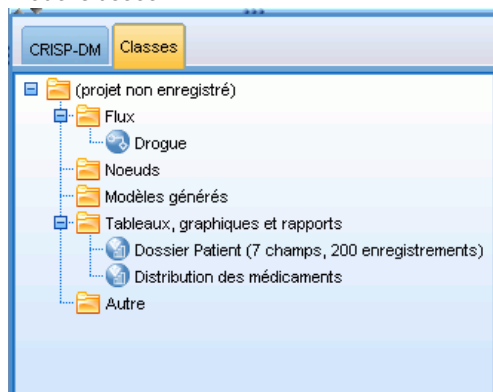
L'onglet CRISP-DM permet d'organiser les projets en fonction de la méthodologie Cross-Industry Standard Process for Data mining commune utilisée dans le domaine. Que vous soyez un utilisateur chevronné ou novice, l'outil CRISP vous aidera à mieux organiser et communiquer vos efforts.

Figure 3-10
Mode CRISP-DM









L'onglet Classes permet d'organiser votre travail dans SPSS Modeler en catégories, selon les types d'objet que vous créez. Ce mode d'affichage est utile lorsque vous effectuez l'inventaire des données, des flux et des modèles.










Figure 3-11
mode Classes



Barre d'outils IBM SPSS Modeler

Une barre d'outils, composée d'icônes fournissant des options très utiles, se trouve en haut de la fenêtre IBM® SPSS® Modeler. Voici les boutons de la barre d'outils et leurs fonctions.

	Créer un flux		Permet d'ouvrir un flux
	Enregistrer le flux		Imprimer le flux actuel
	Déplacer la sélection vers le Presse-papiers		Copier dans le Presse-papiers
	Coller la sélection		Annuler l'action précédente
	Rétablir		Recherche de noeuds
	Editer les propriétés du flux		Aperçu de génération SQL
	Exécuter le flux actuel		Exécuter la sélection de flux

	Arrêter le flux (actif uniquement pendant l'exécution du flux)		Ajouter un super noeud
	Zoom avant (super noeuds uniquement)		Zoom arrière (super noeuds uniquement)
	Aucun balisage dans le flux		Insérer un commentaire
	Masquer le balisage de flux (le cas échéant)		Afficher le balisage de flux masqué
	Permet d'ouvrir un flux dans IBM® SPSS® Modeler Advantage		

Le balisage de flux se compose des commentaires de flux, des liens de modèle et des indications de branche de scoring.

Pour plus d'informations sur les commentaires de flux, reportez-vous à la rubrique [Ajout de commentaires et d'annotations à des noeuds et à des flux sur p. 70](#).

Pour de plus amples informations sur les indications de branche de scoring, reportez-vous à [La branche de scoring sur p. 192](#).

Les liens de modèle sont décrits dans le guide *Noeuds de modélisation IBM SPSS*.

Personnalisation de la barre d'outils

Vous pouvez modifier plusieurs aspects de la barre d'outils, tels que :

- choisir si elle sera affichée ou non
- Choisir si les icônes comporteront ou non des info-bulles
- Choisir si elle utilisera des petites ou des grandes icônes

activer ou désactiver l'affichage de la barre d'outils :

- ▶ Dans le menu principal, cliquez sur :
Affichage > Barre d'outils > Afficher

Pour modifier les paramètres des info-bulles ou de la taille des icônes :

- ▶ Dans le menu principal, cliquez sur :
Affichage > Barre d'outils > Personnaliser

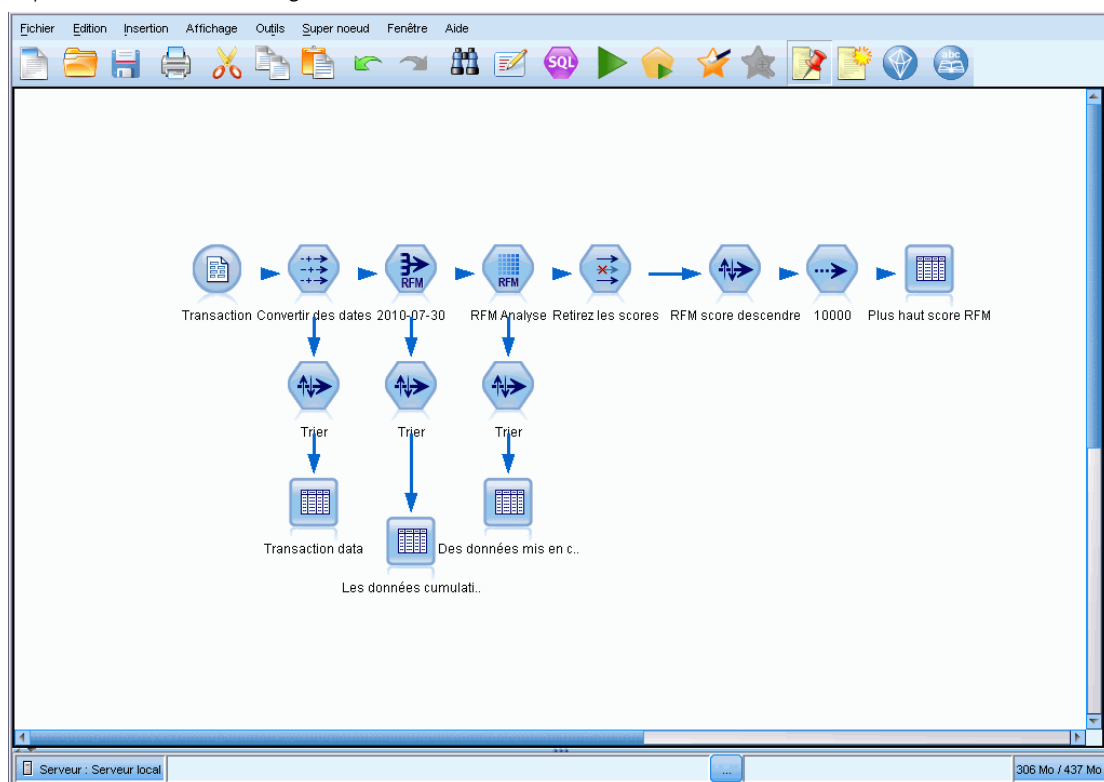
Cliquez sur **Afficher les info-bulles** ou **Gros boutons** le cas échéant.

Personnalisation de la fenêtre IBM SPSS Modeler

Vous pouvez utiliser les séparateurs situés entre les différentes zones de l'interface IBM® SPSS® Modeler pour redimensionner ou fermer des outils en fonction de vos besoins. Par exemple, si vous travaillez avec un flux volumineux, vous pouvez utiliser les petites flèches situées sur chaque séparateur pour fermer la palette de noeuds, le panneau des gestionnaires et le panneau des projets. Ainsi, vous agrandissez l'espace de travail de flux et libérez suffisamment d'espace pour les flux volumineux ou multiples.

À partir du menu Affichage, vous pouvez aussi cliquer sur Palette de noeuds, Gestionnaires ou Projet pour activer ou désactiver l'affichage de ces éléments.

Figure 3-12
Espace de travail de flux agrandi



Vous pouvez également garder ouvertes la palette des noeuds, et les panneaux des gestionnaires et des projets, et utiliser les barres de défilement de l'espace de travail de flux pour vous déplacer dans cet espace ; ces barres sont situées sur le côté et en bas de la fenêtre SPSS Modeler.

Vous pouvez aussi commander l'affichage du balisage de l'écran, lequel se compose des commentaires de flux, des liens de modèles et des indications de branche de scoring. Pour activer ou désactiver cet affichage, cliquez sur :

Affichage > Balisage de flux

Utilisation de la souris dans IBM SPSS Modeler

Dans IBM® SPSS® Modeler, les utilisations les plus courantes de la souris sont les suivantes :

- **Clic simple.** Utilisez le bouton droit ou le bouton gauche de la souris pour sélectionner des options dans les menus, ouvrir des menus contextuels, ou accéder à diverses autres commandes et options standard. Cliquez avec la souris et maintenez le bouton de la souris enfoncé pour faire glisser des noeuds.
- **Double-clic.** Double-cliquez avec le bouton gauche de la souris pour placer des noeuds dans l'espace de travail de flux et éditer des noeuds existants.
- **Clic à l'aide du bouton central.** Cliquez avec le bouton central de la souris et faites glisser le curseur pour connecter des noeuds dans l'espace de travail de flux. Double-cliquez avec le bouton central de la souris pour déconnecter un noeud. Si vous ne possédez pas de souris à trois boutons, vous pouvez simuler cette fonction en appuyant sur la touche Alt tout en cliquant avec la souris et en la faisant glisser.

Utilisation de touches de raccourci

Dans IBM® SPSS® Modeler, de nombreuses opérations de programmation visuelle sont associées à des touches de raccourci. Par exemple, vous pouvez supprimer un noeud en cliquant dessus et en appuyant sur la touche Suppr de votre clavier. De la même façon, vous pouvez enregistrer rapidement un flux en appuyant sur la touche S tout en maintenant la touche Ctrl enfoncée. Les commandes de ce type sont indiquées par Ctrl et une autre touche (par exemple, Ctrl+S).

De nombreuses touches de raccourci sont utilisées dans les opérations Windows standard, telles que Ctrl+X pour couper un élément. Ces raccourcis sont pris en charge dans SPSS Modeler, parallèlement à ceux présentés ci-après, propres à l'application.

Remarque : dans certains cas, les anciennes touches de raccourci utilisées dans SPSS Modeler sont en conflit avec les touches de raccourci Windows standard. Pour que ces anciennes touches de raccourci fonctionnent, il faut utiliser la touche Alt. Par exemple, Ctrl+Alt+C peut activer ou désactiver la mise en cache.

Table 3-1

Touches de raccourci prises en charge

Touche de raccourci	Fonction
Ctrl+A	Tout sélectionner
Ctrl+X	Couper
Ctrl+N	Permet de créer un flux
Ctrl+O	Permet d'ouvrir un flux
Ctrl+P	Imprimer
Ctrl+C	Copier
Ctrl+V	Coller
Ctrl+Z	Annuler
Ctrl+Q	Permet de sélectionner tous les noeuds situés en aval du noeud sélectionné.
Ctrl+W	Permet de désélectionner tous les noeuds en aval (bascule du raccourci Ctrl+Q)

Touche de raccourci	Fonction
Ctrl+E	Exécuter à partir d'un noeud sélectionné
Ctrl+S	Permet d'enregistrer le flux en cours
Alt+flèches	Permettent de déplacer les noeuds sélectionnés dans l'espace de travail de flux dans le sens indiqué par la flèche utilisée
Maj+F10	Permet d'ouvrir le menu contextuel du noeud sélectionné

Table 3-2

Touches de raccourci prises en charge pour les anciennes touches d'accès rapide

Touche de raccourci	Fonction
Ctrl+Alt+D	Permet de dupliquer un noeud
Ctrl+Alt+L	Permet de charger un noeud
Ctrl+Alt+R	Permet de renommer un noeud
Ctrl+Alt+U	Permet de créer un noeud Utilisateur
Ctrl+Alt+C	Permet d'activer et de désactiver le cache
Ctrl+Alt+F	Permet de vider le cache
Ctrl+Alt+X	Développer le super noeud
Ctrl+Alt+Z	Permet d'effectuer un zoom avant/arrière
Suppr	Permet de supprimer un noeud ou une connexion

Impression

Les objets suivants peuvent être imprimés dans IBM® SPSS® Modeler :

- Diagrammes de flux
- Graphiques
- Tableaux
- Rapports (à partir du noeud Rapport et des rapports de projet)
- Scripts (à partir des boîtes de dialogue Propriétés du flux, Script autonome ou Script Super noeud)
- Modèles (navigateurs de modèle, onglets de boîte de dialogue avec élément en cours, afficheurs d'arbres)
- Annotations (à partir de l'onglet Annotations de la sortie)

Pour imprimer un objet :

- Pour imprimer sans afficher d'aperçu, cliquez sur le bouton Imprimer de la barre d'outils.
- Pour définir la mise en page avant d'imprimer, sélectionnez Mise en page dans le menu Fichier.
- Pour afficher un aperçu avant d'imprimer, sélectionnez Aperçu avant impression dans le menu Fichier.
- Pour afficher la boîte de dialogue d'impression standard vous permettant de sélectionner les imprimantes et de définir des options d'aspect, sélectionnez Imprimer dans le menu Fichier.

Automatisation de IBM SPSS Modeler

Etant donné que le Data mining avancé peut être complexe et parfois long, IBM® SPSS® Modeler comprend plusieurs types d'assistance au codage et à l'automatisation.

- **Control Language for Expression Manipulation (CLEM)** est un langage permettant d'analyser et de manipuler les données circulant au sein des flux de SPSS Modeler. Les data miners utilisent beaucoup le langage CLEM dans les opérations de flux pour exécuter des tâches aussi simples que le calcul du profit à partir des données de coûts et de revenus, ou aussi complexes que la transformation de données du log Web en un ensemble de champs et d'enregistrements contenant des informations utilisables. [Pour plus d'informations, reportez-vous à la section A propos de CLEM dans le chapitre 7 sur p. 99.](#)
- **La génération de scripts** est un outil performant pour automatiser les processus dans l'interface utilisateur. Les scripts effectuent des opérations semblables à celles qui peuvent être exécutées à la souris ou au clavier. Vous pouvez définir des options pour les noeuds et effectuer des dérivations en utilisant un sous-ensemble du CLEM. Vous pouvez également définir une sortie et manipuler des modèles générés. [Pour plus d'informations, reportez-vous à la section Génération de scripts - Présentation dans le chapitre 2 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.*](#)

Compréhension du Data mining

Présentation du Data mining

A l'aide de plusieurs techniques, le **Data mining** identifie des portions d'informations dans des bases de données. Le Data mining permet d'extraire des informations de manière à pouvoir les utiliser dans des domaines tels que l'aide à la décision, les prévisions et les estimations. Les données sont souvent volumineuses, mais peu significatives et ne sont pas directement utilisables si elles restent à « l'état brut ». Seules les informations masquées des données ont de la valeur.

En Data mining, vous atteindrez votre objectif en associant vos connaissances des données ou celles de votre expert à des techniques d'analyse élaborées et actives où l'ordinateur identifie les relations et options sous-jacentes des données. Le processus de Data mining génère des modèles à partir des données historiques utilisées ensuite pour les prévisions, la détection de caractéristiques, etc. La technique de création de ces modèles est appelée **apprentissage automatique** ou **modélisation**.

Méthodes de modélisation

IBM® SPSS® Modeler offre plusieurs technologies de modélisation et d'apprentissage automatique, qui peuvent être approximativement regroupées selon les types de problème qu'elles sont censées résoudre.

- Les méthodes de modélisation prédictive incluent les arbres décision, les réseaux de neurones et les modèles statistiques.
- Les modèles de classification se chargent essentiellement d'identifier des groupes d'enregistrements similaires et de répertorier les enregistrements en fonction du groupe auquel ils appartiennent. Les méthodes de classification non supervisée sont les suivantes : Kohonen, *k*-means et TwoStep.
- Les règles d'association associent une conclusion particulière (l'achat d'un produit particulier) à un ensemble de conditions (l'achat de plusieurs autres produits).
- Vous pouvez utiliser les modèles de filtrage pour filtrer les données afin de repérer les champs et enregistrements présentant le plus d'intérêt pour une modélisation et d'identifier les valeurs éloignées qui ne correspondent pas aux patrons connus. Vous disposez pour cela de deux méthodes : la sélection de fonction et la détection d'anomalies.

Découverte et manipulation des données

SPSS Modeler inclut aussi des fonctions vous permettant d'appliquer vos connaissances aux données :

- **Manipulation des données.** Construit des éléments de données dérivés d'éléments existants et scinde les données en sous-ensembles plus compréhensibles. Vous pouvez fusionner et filtrer des données provenant de différentes sources.

- **Navigation et visualisation.** Affiche certains aspects des données à l'aide du nœud Audit données pour exécuter un audit initial, y compris les graphiques et les statistiques. La visualisation avancée comprend des graphiques interactifs, qui peuvent être exportés pour être inclus dans les rapports de projet.
- **Statistics.** Confirme les liens soupçonnés entre des variables dans les données. Vous pouvez également utiliser les statistiques de IBM® SPSS® Statistics dans SPSS Modeler.
- **Test des hypothèses.** Crée des modèles sur le comportement des données et vérifie ces modèles.

En règle générale, vous utiliserez ces outils pour identifier une série d'attributs de données pouvant se révéler très utile. Vous pouvez ensuite intégrer ces attributs aux techniques de modélisation en vue de l'identification des règles et des relations sous-jacentes.

Applications standard

Les applications standard utilisées pour les méthodes de Data mining sont les suivantes :

Publipostage direct. Déterminer les groupes démographiques qui ont la plus grande fréquence de réponses. Utilisez ces informations pour optimiser le nombre de réponses aux publipostages ultérieurs.

Cote de solvabilité. Utiliser l'historique des crédits d'un individu pour les prises de décisions concernant des acceptations ou des refus de crédit.

Ressources humaines. Comprendre les anciennes pratiques et créer des règles de décision visant à simplifier le processus d'embauche.

Recherche médicale. Créer des règles de décision qui suggèrent des procédures sur la base de preuves médicales.

Analyse de marché. Déterminer les variables (caractéristiques géographiques, tarifaires et liées aux clients) qui sont associées aux ventes.

Contrôle qualité. Analyser les données de la fabrication d'un produit et identifier les variables qui déterminent les défauts sur les produits.

Etudes stratégiques. Utiliser des données d'enquête pour établir une stratégie en appliquant des règles de décision qui permettent de sélectionner les variables les plus importantes.

Soins de santé. Associer des études et des données cliniques afin de déterminer les variables à prendre en compte dans le cadre de soins de santé.

Terminologie

Les termes **attribut**, **champ** et **variable** se rapportent à un élément de données unique commun à toutes les observations en cours. Un ensemble de valeurs d'attribut se rapportant à une observation spécifique est appelé **enregistrement**, **exemple** ou **observation**.

Evaluation des données

Pour qu'un travail de Data mining soit fructueux, les données à utiliser doivent répondre à certains critères. Les sections suivantes présentent certains aspects des données et de leur application à prendre en considération.

Assurez-vous que les données sont disponibles

Ce point peut sembler évident, mais il faut savoir que même si les données sont disponibles, elles peuvent ne pas être exploitables immédiatement. IBM® SPSS® Modeler peut importer des données à partir de bases de données (via ODBC) ou de fichiers. Les données peuvent se présenter sous une forme différente sur une machine à laquelle on ne peut pas accéder directement. Elles devront pour cela être téléchargées ou transférées dans un format convenable avant de pouvoir être utilisées. Au besoin, elles devront être dispersées parmi plusieurs bases de données et sources avant d'être de nouveau regroupées. Ce processus n'a pas forcément lieu en ligne. Si les données sont inscrites sur papier, une phase de saisie des données est nécessaire avant de procéder au Data mining.

Vérifiez que les données couvrent tous les attributs nécessaires

L'objectif du Data mining est d'identifier les attributs qui conviennent, par conséquent, ce point peut sembler étrange. Cependant, il s'avère très utile de consulter les données disponibles et d'essayer d'identifier les facteurs pertinents non enregistrés. Par exemple, si vous souhaitez prévoir la vente de glaces, vous disposerez d'une quantité importante d'informations via les magasins de vente au détail et l'historique des ventes, ce qui n'est pas le cas des conditions météorologiques dont l'impact peut être significatif. Le fait que des attributs manquent ne signifie pas nécessairement que le Data mining ne produira pas de résultats fiables. Par contre, l'exactitude des prévisions peut être limitée.

L'exécution d'un audit complet de vos données permet d'évaluer rapidement la situation. Avant de continuer, associez un nœud Audit données à votre source de données, puis exécutez-le pour générer un rapport complet. [Pour plus d'informations, reportez-vous à la section Nœud Audit données dans le chapitre 6 dans IBM SPSS Modeler 14.2 Nœuds source, exécution et de sortie.](#)

Faites attention aux données parasites

Les données contiennent souvent des erreurs ou peuvent contenir des points de vue subjectifs qui peuvent changer dans le temps. Ces phénomènes sont généralement désignés sous le nom de **bruits parasites**. Parfois, ces parasites sont normaux. Même s'il existait une règle générale, elle ne pourrait pas s'appliquer dans cent pour cent des cas.

En résumé, plus les données contiennent de parasites, plus il est difficile d'obtenir des résultats précis. Cependant, les méthodes d'apprentissage automatique de SPSS Modeler sont capables de gérer les données parasites, et ont été utilisées avec succès sur les ensembles de données contenant environ 50 % de données parasites.

Assurez-vous d'avoir assez de données

Dans le Data mining, ce n'est pas nécessairement la taille d'un ensemble de données qui importe. La *représentativité* de l'ensemble de données est bien plus significative, ainsi que la couverture des résultats possibles et des combinaisons de variables.

En général, plus le nombre d'attributs considérés est important, plus le nombre d'enregistrements nécessaires pour obtenir une couverture représentative est élevé.

Si les données sont représentatives et s'il existe des règles sous-jacentes générales, un échantillon de données contenant plusieurs milliers d'enregistrements (ou plusieurs centaines), peut donner des résultats aussi performants qu'un échantillon contenant un million d'enregistrements. De plus, vous obtiendrez les résultats plus rapidement.

Adressez-vous aux spécialistes de ces données

Dans la plupart des cas, vous travaillez sur vos propres données et vous maîtrisez donc leur sens et contenu. Cependant, si vous travaillez sur des données pour un autre service de votre entreprise ou pour un client, il est fortement conseillé d'entrer en contact avec les experts qui maîtrisent ces données. Ils peuvent aider à identifier les attributs pertinents et à interpréter les résultats de Data mining, en faisant la différence entre des informations pertinentes et des informations erronées (informations sans aucune valeur) dues à des anomalies dans les ensembles de données.

Stratégie pour le Data mining

Comme dans le cas de la plupart des efforts commerciaux, le Data mining est beaucoup plus profitable si vous procédez de manière méthodique. Même si vous utilisez des outils de Data mining de pointe comme IBM® SPSS® Modeler, vous avez besoin de la présence d'un analyste en informatique de gestion expérimenté pour assurer le suivi du processus. Pour vous guider dans votre planification, répondez aux questions suivantes :

- Quel réel problème souhaitez-vous résoudre ?
- De quelles sources de données disposez-vous et quelles sont les données nécessaires au problème en cours ?
- Quel genre de prétraitement et de nettoyage de données devez-vous effectuer avant d'utiliser le Data mining ?
- Quelles techniques de Data mining pensez-vous utiliser ?
- Comment pensez-vous évaluer les résultats de l'analyse de Data mining ?
- Comment tirer profit au maximum des informations obtenues suite au Data mining ?

Le Data mining peut devenir très rapidement compliqué. Vous devez surveiller beaucoup d'éléments à la fois, tels que les difficultés commerciales complexes, les sources de données multiples, les différents niveaux de qualité des données selon la source, la multiplicité des techniques de Data mining et des méthodes d'évaluation des résultats, etc.

Il est également important d'avoir à disposition un modèle de processus clairement défini pendant le Data mining. Le modèle de processus vous aide à répondre aux questions répertoriées précédemment dans cette section et permet de s'assurer que les points importants sont traités. Il

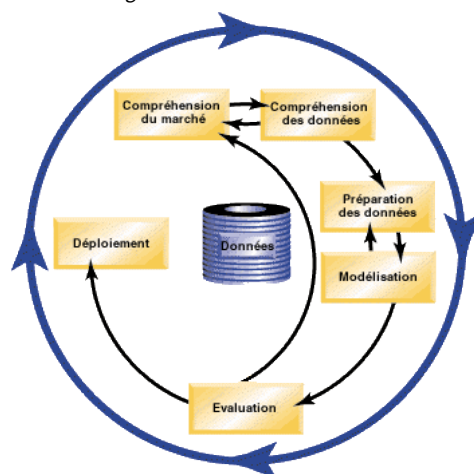
vous indique la marche à suivre pour que vous restiez concentré sur le fond du problème tout en faisant face à la complexité des données à explorer.

Le modèle de processus de Data mining recommandé avec SPSS Modeler est le processus CRISP-DM (Cross-Industry Standard Process for Data Mining). En d'autres termes, ce modèle est conçu pour résoudre une grande diversité de problèmes propres aux entreprises ou au commerce.

Méthodologie CRISP-DM

La méthodologie CRISP-DM comprend six étapes au cours desquelles vous aborderez les principaux problèmes de Data mining. Les six étapes s'ajustent au sein d'un processus cyclique conçu pour incorporer le Data mining dans le cadre de vos activités commerciales en augmentation.

Figure 4-1
méthodologie CRISP-DM



Les six étapes comprennent :

- **Compréhension du problème.** Il s'agit peut-être de l'étape de Data mining la plus importante. La compréhension du problème vise à déterminer des objectifs commerciaux, à évaluer la situation, à déterminer les objectifs de Data mining et à établir une planification de projet.
- **Compréhension des données.** Les données sont la « matière première » du Data mining. Cette étape vise à vous sensibiliser à l'importance de bien maîtriser les ressources de données et les caractéristiques de ces ressources. Elle aborde la collecte de données initiale, la description et l'exploration des données, et la vérification de la qualité de ces données. Le nœud Audit données disponible dans la palette de noeuds Sortie est un outil indispensable pour comprendre les données.
- **Préparation des données.** Après avoir catalogué vos ressources de données, vous devez préparer les données à explorer. La préparation comprend la sélection, le nettoyage, la construction, l'intégration et le formatage des données.
- **Modélisation.** Cette étape est la plus intéressante. Des méthodes d'analyse élaborées sont utilisées pour extraire des informations à partir des données. Lors de cette étape, il s'agit de sélectionner des techniques de modélisation, de générer des conceptions de test, et de construire et d'évaluer des modèles.

- **Evaluation.** Une fois que vous avez choisi vos modèles, vous pouvez évaluer l'aide que peuvent apporter les résultats de Data mining à la concrétisation de vos objectifs commerciaux. Cette étape aborde l'évaluation des résultats, la vérification du processus de Data mining et la prise de décision des étapes à suivre.
- **Déploiement.** Cette étape permet de rentabiliser pleinement tous vos efforts. L'objectif de cette phase est d'intégrer vos nouvelles connaissances aux processus commerciaux quotidiens pour résoudre votre problème commercial initial. Elle comprend le déploiement du plan, la surveillance et la maintenance, la production d'un rapport final et la révision du projet.

Il ya des éléments clés dans ce modèle de processus. Tout d'abord, il faut savoir que, même si le processus respecte l'ordre des étapes présentées dans les paragraphes précédents, il existe des situations où les étapes ont une influence les unes sur les autres dans un ordre aléatoire. Par exemple, la préparation des données précède généralement la modélisation. Cependant, les décisions prises et les informations collectées lors de la modélisation remettent en question certaines parties de la préparation des données, ce qui entraîne de nouveaux problèmes de modélisation. Les deux étapes sont alors interdépendantes jusqu'à ce qu'elles soient satisfaisantes. De même, l'étape d'évaluation peut vous amener à réévaluer votre première analyse du marché et vous aider à prendre conscience du fait que vous êtes en train de répondre à la mauvaise question. A ce stade, vous pouvez revoir votre compréhension du problème, puis passer au reste du processus dans le but d'atteindre un objectif clairement défini.

Le second élément-clé est la nature itérative du Data mining. Ce serait un tort de croire que le Data mining se limite à planifier un projet, à le réaliser et à rassembler vos données. L'utilisation du Data mining pour satisfaire aux exigences de votre clientèle est un effort continu. Les connaissances acquises, suite à un cycle de Data mining, aboutissent presque systématiquement à de nouvelles questions, de nouveaux problèmes et de nouvelles opportunités, permettant d'identifier les besoins de votre clientèle et d'y répondre. Ces nouvelles questions, opportunités et nouveaux problèmes peuvent être abordés en renouvelant l'extraction de données. Le Data mining et l'identification de nouvelles opportunités doivent désormais faire partie de vos activités professionnelles et être un élément-clé de votre stratégie commerciale globale.

Cette introduction ne donne qu'un aperçu général de la méthodologie CRISP-DM. Pour plus d'informations sur le modèle, consultez les ressources suivantes :

- Le *Guide CRISP-DM*, accessible de même que d'autres documents à partir du dossier *Documentation* du disque d'installation.
- Le système d'aide de CRISP-DM, disponible à partir du menu Démarrer ou en cliquant sur Aide CRISP-DM à partir du menu Aide de IBM® SPSS® Modeler.

Types de modèle

IBM® SPSS® Modeler propose différentes méthodes de modélisation issues des domaines de l'apprentissage automatique, de l'intelligence artificielle et des statistiques. Les méthodes disponibles dans la palette Modélisation vous permettent d'extraire de nouvelles informations de vos données et de développer des modèles prédictifs. Chaque méthode possède ses propres avantages et est donc plus adaptée à certains types de problème spécifiques.

Le *Guide des applications SPSS Modeler* fournit des exemples pour de nombreuses méthodes, ainsi qu'une présentation générale du processus de modélisation. Ce guide est disponible en tant que tutoriel en ligne ainsi qu'au format PDF. [Pour plus d'informations, reportez-vous à la section Exemples d'application dans le chapitre 1 sur p. 4.](#)

Les méthodes de modélisation sont divisées en trois catégories :

- Classification
- Association
- Segmentation

Modèles de classification

Les *modèles de classification* utilisent les valeurs d'un ou de plusieurs champs d'**entrée** afin de prédire la valeur d'un ou de plusieurs champs de résultat ou **cible**. Certains exemples de ces techniques sont : Les arbres de décision (algorithmes C&R Tree, QUEST, CHAID et C5.0), la régression (algorithmes linéaires, logistiques, linéaires généralisés et régression de Cox), les réseaux de neurones, les machines à vecteurs de prise en charge et les réseaux Bayésiens.

Les modèles de classification permettent aux entreprises de prévoir un résultat connu, par exemple si un client va acheter ou ne pas acheter un produit, ou si une transaction entre dans le cadre d'une fraude connue. Les techniques de modélisation comprennent le processus d'apprentissage automatique, l'induction de règles, l'identification de sous-groupes, les méthodes statistiques et la génération de modèles multiples.

noeuds Classification



Le noeud Classificateur automatique crée et compare les résultats binaires de plusieurs modèles différents (oui ou non, avec ou sans attrition, etc.), ce qui vous permet de choisir la meilleure approche pour une analyse donnée. Plusieurs algorithmes de modélisation sont pris en charge. Vous pouvez alors sélectionner les méthodes que vous souhaitez utiliser, les options spécifiques pour chacune d'elles et le critère de comparaison des résultats. Le noeud génère un ensemble de modèles basé sur les options spécifiées et classe les meilleurs candidats en fonction des critères indiqués. [Pour plus d'informations, reportez-vous à la section Noeud Classificateur automatique dans le chapitre 5 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Numérisation automatique évalue et compare des modèles pour des résultats d'intervalle numérique continus par le biais de différentes méthodes. Le noeud fonctionne de la même manière que le noeud Classificateur automatique, vous permettant ainsi de choisir les algorithmes à utiliser et à tester avec différentes combinaisons d'options en un seul passage de modélisation. Les algorithmes pris en charge comprennent les réseaux neuraux, l'arborescence C&R Tree, CHAID, la régression linéaire, la régression linéaire généralisée et support vector machines (SVM). Les modèles peuvent être comparés selon la corrélation, l'erreur relative ou le nombre de variables utilisées. [Pour plus d'informations, reportez-vous à la section Noeud Numérisation automatique dans le chapitre 5 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Arbre Classification et Regression, (C&R) Tree, génère un arbre décision qui vous permet de prévoir ou de classifier les observations futures. La méthode utilise la technique de partition récursive afin de diviser les données d'apprentissage en segments en réduisant l'index d'impureté à chaque étape, un noeud de l'arbre étant considéré comme "pur" si 100 % de ses observations appartiennent à une catégorie spécifique du champ cible. Les champs cible et les champs d'entrée peuvent être des champs d'intervalle numériques ou des champs catégoriels numériques (nominal, ordinal ou booléen). Toutes les divisions sont binaires (deux sous-groupes uniquement). [Pour plus d'informations, reportez-vous à la section Noeud Arbre C&RT dans le chapitre 6 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud QUEST est une méthode de classification supervisée binaire permettant de créer des arbres décision, développée pour réduire le temps de traitement nécessaire aux analyses C&R Tree importantes, tout en limitant la tendance, observée parmi les méthodes d'arbre de classification, à favoriser les entrées autorisant un nombre supérieur de divisions. Les champs d'entrée peuvent être des intervalles numériques (continues) mais les champs cible doivent être catégoriels. Toutes les divisions sont binaires. [Pour plus d'informations, reportez-vous à la section Noeud QUEST dans le chapitre 6 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud CHAID génère des arbres décision à l'aide des statistiques du Chi-deux pour identifier les séparations optimales. Contrairement aux noeuds C&R Tree et QUEST, CHAID peut générer des arbres non binaires, ce qui implique que certaines divisions possèdent plusieurs branches. Les champs cibles et les champs d'entrée peuvent être d'intervalle numérique (continu) ou catégoriels. La méthode Exhaustive CHAID correspond à une modification du CHAID qui examine plus en détail toutes les divisions possibles, mais dont les calculs sont plus longs. [Pour plus d'informations, reportez-vous à la section Noeud CHAID dans le chapitre 6 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud C5.0 crée un arbre décision ou un ensemble de règles. Le fonctionnement de ce modèle repose sur un découpage de l'échantillon basé sur le champ qui fournit le gain d'informations le plus important à chaque niveau. Le champ cible doit être catégoriel. Les divisions multiples en plus de deux sous-groupes sont autorisées. [Pour plus d'informations, reportez-vous à la section Noeud C5.0 dans le chapitre 6 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud Liste de décision identifie les sous-groupes, ou les segments, qui présentent une probabilité plus élevée ou plus faible d'un résultat binaire donné par rapport à la population globale. Vous pouvez, par exemple, rechercher les clients qui ont une faible probabilité d'attrition ou ceux qui ont une plus forte probabilité de répondre favorablement à une campagne. Vous pouvez incorporer votre connaissance du marché dans le modèle en ajoutant vos propres segments personnalisés et en prévisualisant des modèles alternatifs côte à côte de façon à comparer les résultats. Les modèles Liste de décision se composent d'une liste de règles dans laquelle chaque règle présente une condition et un résultat. Les règles sont appliquées dans l'ordre et la première règle correspondante détermine le résultat. [Pour plus d'informations, reportez-vous à la section Liste de décisions dans le chapitre 9 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Les modèles de régression linéaire prédisent une cible continue en fonction de relations linéaires entre la cible et une ou plusieurs variables prédites. [Pour plus d'informations, reportez-vous à la section Modèles linéaires dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



La régression linéaire est une technique statistique couramment utilisée dans le domaine de la synthèse de données et de la prévision. Cette technique établit une ligne droite ou une surface afin de réduire les écarts entre les valeurs de sortie prévues et observées.



Le noeud ACP/Analyse factorielle propose des techniques de factorisation puissantes qui vous permettent de réduire la complexité de vos données. L'analyse en composantes principales (ACP) recherche les combinaisons linéaires des champs d'entrée qui permettent de capturer au mieux la variance dans l'ensemble de champs, où les composantes sont orthogonales (perpendiculaires) les unes par rapport aux autres. L'analyse factorielle a pour but d'identifier les facteurs sous-jacents qui expliquent la tendance des corrélations dans un ensemble de champs observés. Quelle que soit l'approche choisie, le but consiste à trouver un nombre limité de champs dérivés récapitulant les informations contenues dans l'ensemble de champs d'origine. [Pour plus d'informations, reportez-vous à la section Noeud ACP/Analyse factorielle dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud Sélection de fonction filtre les champs d'entrée en vue de leur suppression, en fonction d'un ensemble de critères donné (tel que le pourcentage de valeurs manquantes) ; il classe ensuite les entrées restantes selon leur importance par rapport à la cible indiquée. Si l'on prend, par exemple, un ensemble de données comportant des centaines d'entrées potentielles, quelles sont celles susceptibles d'être les plus utiles dans la modélisation des résultats de patients ? [Pour plus d'informations, reportez-vous à la section Noeud Sélection de fonction dans le chapitre 4 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



L'analyse discriminante crée des hypothèses plus strictes que la régression logistique mais peut constituer une alternative ou un complément précieux à une analyse de régression logistique lorsque ces hypothèses sont réunies. [Pour plus d'informations, reportez-vous à la section Noeud discriminant dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



La régression logistique est une technique statistique de classification des enregistrements sur la base des valeurs des champs d'entrée. Excepté le fait qu'elle utilise un champ cible catégoriel et non pas numérique, cette régression est similaire à la régression linéaire. [Pour plus d'informations, reportez-vous à la section Noeud Logistique dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



La procédure Modèles linéaires généralisés développe le modèle linéaire général de sorte que la variable dépendante soit linéairement reliée aux facteurs et covariables via une fonction de lien précise. En outre, le modèle permet à la variable dépendante de suivre une distribution non normale. Il couvre les fonctionnalités d'un grand nombre de modèles statistiques, notamment le modèle de régression linéaire, le modèle de régression logistique, le modèle log-linéaire pour les données d'effectif et le modèle de survie avec censure par intervalle. [Pour plus d'informations, reportez-vous à la section Noeud Modèles linéaires généralisés dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud de régression de Cox vous permet de créer un modèle de survie pour les données de durée jusqu'à l'événement en présence d'enregistrements censurés. Ce modèle produit une fonction de survie qui prédit la probabilité que l'événement en question se soit produit à un moment (t) pour des valeurs données des variables d'entrée. [Pour plus d'informations, reportez-vous à la section Noeud de Cox dans le chapitre 10 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud Support Vector Machine (SVM) vous permet de classer les données dans l'un de deux groupes sans sur-ajustement. SVM fonctionne bien avec les grands ensembles de données, comme ceux qui disposent d'un très grand nombre de champs d'entrée. [Pour plus d'informations, reportez-vous à la section Noeud SVM dans le chapitre 15 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Réseau Bayésien permet de créer un modèle de probabilité en combinant les preuves observées et enregistrées avec les connaissances réelles pour établir la probabilité des occurrences. Le noeud est axé sur le Tree Augmented Naïve Bayes (TAN) et sur les réseaux Couverture de Markov qui servent principalement à la classification. [Pour plus d'informations, reportez-vous à la section Noeud Réseau Bayésien dans le chapitre 7 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Modèle de réponse en auto-apprentissage (SLRM) vous permet de créer un modèle dans lequel une nouvelle observation unique, ou un petit nombre de nouvelles observations, peuvent être utilisés pour réestimer un modèle sans qu'un recyclage de toutes les données soit nécessaire. [Pour plus d'informations, reportez-vous à la section Noeud MRAA dans le chapitre 14 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Séries temporelles estime les modèles de lissage exponentiel, d'ARIMA (Autoregressive Integrated Moving Average) univariable et d'ARIMA multivariable (ou fonction de transfert) pour les données de séries temporelles et génère des prévisions d'une performance future. Un noeud Séries temporelles doit toujours être précédé d'un noeud Intervalles de temps. [Pour plus d'informations, reportez-vous à la section Noeud de modélisation Séries temporelles dans le chapitre 13 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud k -Voisin le plus proche (KNN) associe une nouvelle observation à la catégorie ou à la valeur des objets k les plus proches dans l'espace du prédicteur, où k est un entier. Les observations semblables sont proches l'une de l'autre et les observations dissemblables sont éloignées l'une de l'autre. [Pour plus d'informations, reportez-vous à la section Noeud KNN dans le chapitre 16 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Modèles d'association

Les *modèles d'association* recherchent des modèles dans vos données où une ou plusieurs entités (telles que des événements, des achats ou des attributs) sont associées à une ou plusieurs autres entités. Les modèles établissent des ensembles de règles qui définissent ces relations. Ici, les champs au sein des données peuvent se comporter à la fois comme des entrées et comme des cibles. Vous ne pouvez pas découvrir ces associations manuellement, mais des algorithmes de règles font bien plus et plus rapidement, de sorte que vous êtes en mesure d'explorer des modèles plus complexes. Les modèles Apriori et Carma sont des exemples de l'utilisation de tels algorithmes. Un autre type de modèle d'association est un modèle de détection de séquences qui recherche des motifs séquentiels dans des données à structure temporelle.

Les modèles d'association sont particulièrement utiles pour prévoir des résultats multiples ; par exemple, les clients qui ont acheté le produit X ont également acheté les produits Y et Z. Les modèles d'association associent une conclusion particulière (telle que la décision d'acheter un produit) à un ensemble de conditions. L'avantage des algorithmes de règles d'association par rapport aux algorithmes d'arbre décision standard (C5.0 et Arbre C&RT) est le fait qu'il puisse

exister des associations entre tous les attributs. Un algorithme d'arbre décision peut construire une règle uniquement avec une seule conclusion. En revanche, les algorithmes d'association tentent d'en trouver plusieurs, chaque règle pouvant avoir une conclusion différente.

Noeuds Association



Le noeud Apriori extrait des données un ensemble de règles et retient les règles contenant la plus grande quantité d'informations. Le noeud Apriori fournit cinq méthodes de sélection de règles et utilise un modèle d'indexation sophistiqué pour traiter efficacement les volumes de données importants. Pour les problèmes importants, l'apprentissage du noeud Apriori est généralement plus rapide ; il n'existe aucune limite quant au nombre de règles pouvant être conservées et il peut prendre en charge des règles faisant l'objet de 32 pré-conditions. Le noeud Apriori exige que les champs d'entrée et de sortie soient tous catégoriels, mais fournit de meilleures performances car il est optimisé de ce type de données. [Pour plus d'informations, reportez-vous à la section Noeud Apriori dans le chapitre 12 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le modèle CARMA extrait un ensemble de règles des données sans que vous ayez à définir les champs d'entrée ou les champs cible. Au contraire du noeud Apriori le noeud CARMA offre des paramètres de création pour la prise en charge de la règle (à la fois pour les antécédents et les conséquences), plutôt qu'une simple prise en charge d'antécédents. Cela signifie que les règles générées peuvent être utilisées dans un grand nombre d'applications, par exemple pour rechercher une liste des produits ou des services (antécédents) dont la conséquence correspond à l'élément que vous souhaitez promouvoir à l'occasion de cette période de congés. [Pour plus d'informations, reportez-vous à la section Noeud CARMA dans le chapitre 12 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)



Le noeud Séquence recherche des règles d'association dans des données dotées d'une dimension temporelle. Une séquence est une liste de jeux d'éléments ayant tendance à survenir dans un ordre prévisible. Par exemple, un client qui achète un rasoir et une lotion après-rasage achètera vraisemblablement de la crème à raser. Le noeud Séquence est basé sur l'algorithme de règles d'association CARMA, qui utilise une méthode efficace de double lecture pour rechercher des séquences. [Pour plus d'informations, reportez-vous à la section Noeud Séquence dans le chapitre 12 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Modèles de segmentation

Les *modèles de segmentation* divisent les données en segments, ou classes, d'enregistrement ayant des profils similaires de champs d'entrée. Comme ils ne s'occupent que des champs d'entrée, les modèles de segmentation n'ont aucun concept de champs de sortie ou cible. Parmi les exemples de modèles de segmentation, on trouve les réseaux Kohonen, la classification K-Means, la classification en deux étapes et la détection d'anomalies.

Les modèles de Segmentation (appelés aussi « modèles de classification ») sont utiles dans les cas où le résultat précis est inconnu (par exemple, lorsque vous identifiez de nouveaux types de fraude ou lorsque vous identifiez des groupes d'intérêt dans votre clientèle). Les modèles de classification se chargent essentiellement d'identifier des groupes d'enregistrements similaires et de répertorier les enregistrements en fonction du groupe auquel ils appartiennent. Cette opération peut s'effectuer sans la connaissance préalable des groupes et de leurs caractéristiques, et distingue

les modèles de classification non supervisée des autres techniques de modélisation par le fait qu'aucun champ de sortie ni champ cible n'est prédéfini pour le modèle à prévoir. Il n'y a pas de réponse vraie ou fausse pour ces modèles. Leur valeur est déterminée par leur capacité à capturer des groupements intéressants dans les données et ils fournissent des descriptions utiles de ces mêmes groupements. Les modèles de classification non supervisée sont souvent utilisés pour créer des classes ou des segments qui sont ensuite utilisés en tant qu'entrées dans les analyses suivantes (par exemple, par la segmentation de clients potentiels dans des sous-groupes homogènes).

Noeuds Segmentation



Le noeud Classification automatique évalue et compare les modèles de classification identifiant des groupes d'enregistrements ayant des caractéristiques similaires. Le noeud fonctionne de la même manière que les autres noeuds de modélisation automatiques, vous permettant de tester plusieurs combinaisons d'options en une seule modélisation. Les modèles peuvent être comparés à l'aide de mesures de bases permettant d'essayer de filtrer et de classer l'utilité des modèles de classification et de fournir une mesure en fonction de l'importance de champs particuliers. [Pour plus d'informations, reportez-vous à la section Noeud Classification automatique dans le chapitre 5 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud K-Means classe l'ensemble de données dans différents groupes (ou classes). La méthode définit un nombre de classes fixe, affecte à plusieurs reprises des enregistrements à des classes et ajuste les centres de classe, jusqu'à ce que le modèle ne puisse plus être amélioré. Au lieu de tenter de prédire un résultat, le modèle k -means utilise un processus connu sous le nom d'apprentissage non supervisé pour découvrir des tendances dans l'ensemble de champs d'entrée. [Pour plus d'informations, reportez-vous à la section Noeud K-Means dans le chapitre 11 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud Kohonen génère un type de réseau de neurones qui peut être utilisé pour classer les données en groupes distincts. Lorsque l'apprentissage du réseau est terminé, les enregistrements similaires doivent être regroupés dans la connexion de sortie, tandis que les enregistrements différents sont à l'opposé. Vous pouvez étudier le nombre d'observations capturées par chaque unité du nœud de modèle afin d'identifier les unités fortes. Vous pouvez ainsi vous faire une idée du nombre de classes approprié. [Pour plus d'informations, reportez-vous à la section Noeud Kohonen dans le chapitre 11 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud TwoStep utilise une méthode de classification non supervisée en deux étapes. La première étape consiste en une exploration des données visant à compresser les données d'entrée brutes en sous-classes plus faciles à manipuler. Au cours de la seconde étape, l'utilisation d'une méthode de classification hiérarchique permet de fusionner progressivement les sous-classes en classes de plus en plus importantes. La technique TwoStep a l'avantage d'évaluer automatiquement le nombre de classes optimal pour les données d'apprentissage. Il peut prendre en charge de manière efficace des types de champ mixtes et des ensembles de données volumineux. [Pour plus d'informations, reportez-vous à la section Noeud Classification TwoStep dans le chapitre 11 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)



Le noeud Détection des anomalies identifie les observations inhabituelles, ou valeurs éloignées, qui ne se conforment pas aux patrons de données « normales ». Il vous permet d'identifier les valeurs éloignées même si celles-ci ne correspondent pas aux patrons connus précédemment et même si vous ne savez pas exactement ce que vous recherchez. [Pour plus d'informations, reportez-vous à la section Noeud Détection des anomalies dans le chapitre 4 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Exploration des bases de données des modèles

SPSS Modeler prend en charge l'intégration des outils de Data mining et de modélisation disponibles auprès des fournisseurs de base de données, notamment Oracle Data Miner, IBM DB2 InfoSphere Warehouse et Microsoft Analysis Services. Dans la base de données, vous pouvez créer, stocker et déterminer le score des modèles qui proviennent tous de l'application SPSS Modeler. Pour plus d'informations, consultez le *Guide d'exploration de base de données SPSS Modeler*, disponible sur le DVD du produit.

Modèles IBM SPSS Statistics

Si vous avez installé une copie sous licence de IBM® SPSS® Statistics sur votre ordinateur, vous pouvez accéder et exécuter certains programmes SPSS Statistics depuis SPSS Modeler pour créer et évaluer des modèles. [Pour plus d'informations, reportez-vous à la section Noeuds IBM SPSS Statistics - Présentation dans le chapitre 8 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)

Informations supplémentaires

Une documentation détaillée sur les algorithmes de modélisation est également disponible. Pour plus d'informations, consultez le *Guides des algorithmes de SPSS Modeler*, disponible sur le DVD du produit.

Exemples de Data mining

Pour en savoir plus sur le Data mining, le meilleur moyen est de commencer par un exemple. Un certain nombre d'exemples d'application sont disponibles dans le *Guide des applications IBM® SPSS® Modeler*, qui fournit des présentations brèves et ciblées de méthodes et de techniques de modélisation spécifiques. [Pour plus d'informations, reportez-vous à la section Exemples d'application dans le chapitre 1 sur p. 4.](#)

Création de flux

Présentation de la création de flux

Le Data mining avec IBM® SPSS® Modeler est basé sur l'exécution de données via une série de noeuds, appelée **flux**. Cette série de noeuds représente les opérations à réaliser sur les données et les liens entre les noeuds indiquent la direction du flux de données. Généralement, un flux de données est utilisé pour lire des données dans SPSS Modeler, effectuer leur exécution par des manipulations diverses et envoyer ces dernières vers une destination, telle qu'un tableau ou un visualiseur.

Par exemple, supposons que vous souhaitiez ouvrir une source de données, ajouter un nouveau champ, sélectionner des enregistrements en fonction des valeurs contenues dans ce nouveau champ, puis afficher les résultats dans un tableau. Dans ce cas, votre flux de données serait composé de quatre noeuds :



Un noeud Délimité défini de façon à pouvoir lire les données à partir de la source de données.



Un noeud Calculer qui permet d'ajouter le nouveau champ calculé à l'ensemble de données.



Un noeud Sélectionner que vous utilisez pour configurer des critères de sélection destinés à exclure certains enregistrements du flux de données.



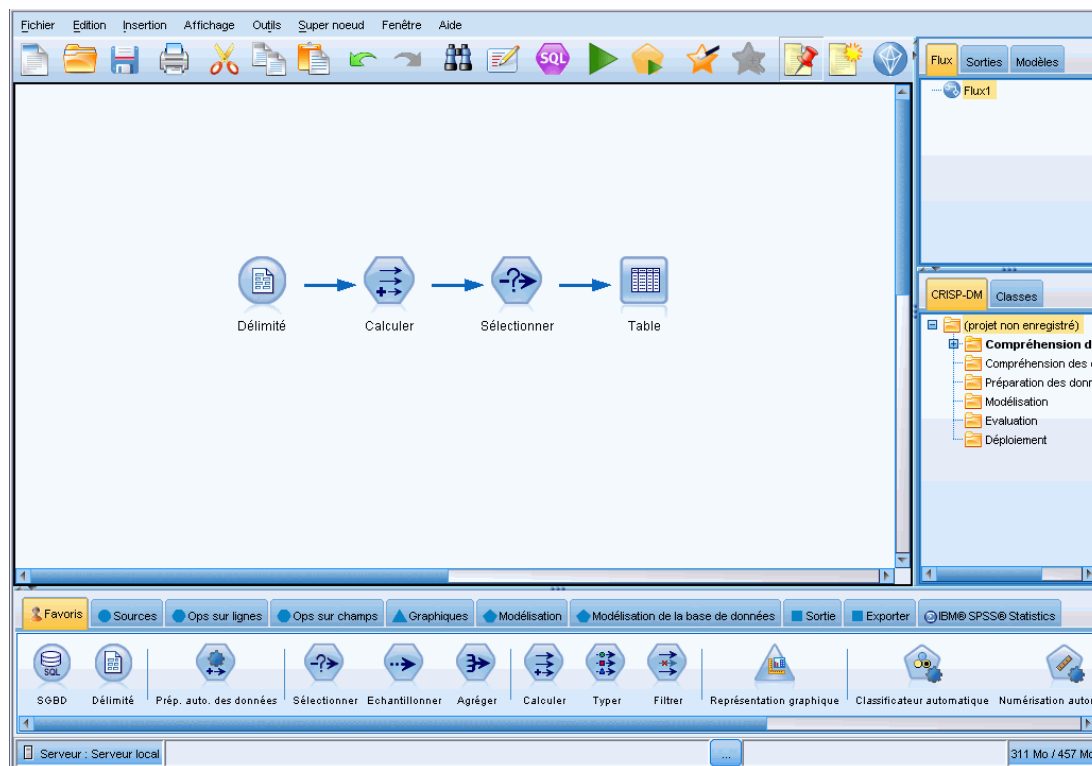
Un noeud Table qui permet d'afficher les résultats de vos manipulations à l'écran.

Création de flux de données

L'interface unique de IBM® SPSS® Modeler vous permet d'explorer les données visuellement en travaillant avec des graphiques de flux de données. Au niveau le plus élémentaire, vous pouvez créer un flux de données en utilisant la procédure suivante :

- Ajoutez des noeuds dans l'espace de travail de flux.
- Connectez les noeuds de façon à former un flux.
- Définissez toutes les options de noeud ou de flux.
- Exécutez le flux.

Figure 5-1
Exemple de flux terminé dans l'espace de travail de flux



Cette section contient des informations plus détaillées sur l'utilisation des noeuds pour créer des flux de données plus complexes. Elle aborde également les options et paramètres des noeuds et des flux. Pour obtenir des exemples pas à pas de création de flux en utilisant les données fournies avec SPSS Modeler (dossier *Demos* du répertoire d'installation du programme), reportez-vous à [Exemples d'application sur p. 4](#).

Utilisation des noeuds

Dans IBM® SPSS® Modeler, les noeuds permettent d'explorer des données. Les différents noeuds dans l'espace de travail représentent différents objets et actions. La palette située au bas de la fenêtre SPSS Modeler contient tous les noeuds pouvant être utilisés pour la création des flux.

Il existe plusieurs types de noeuds. Les **Noeuds source** introduisent les données dans le flux et se trouvent dans l'onglet Sources de la palette des noeuds. Les **Noeuds d'exécution** exécutent des opérations sur des enregistrements et des champs de données individuels et se trouvent dans les onglets de la palette Ops sur lignes et Ops sur champs. Les **noeuds de sortie** produisent plusieurs résultats pour les données, les diagrammes et les résultats des modèles et sont inclus dans les onglets Graphiques, Sortie et Exporter de la palette de noeuds. Les **noeuds de modélisation** utilisent des algorithmes statistiques pour créer des nuggets de modèle et se trouvent dans l'onglet Modélisation, et (s'il est activé) dans l'onglet Modélisation de base de données; de la palette des noeuds. [Pour plus d'informations, reportez-vous à la section Palette de noeuds dans le chapitre 3 sur p. 15.](#)

Ces noeuds doivent être connectés pour former des flux qui, une fois exécutés, vous permettent de visualiser des relations et de tirer des conclusions. Les flux sont similaires à des scripts (vous pouvez les enregistrer et les réutiliser avec d'autres fichiers de données).

Un noeud exécutable qui traite des données du flux est appelé **noeud terminal**. Un noeud de modélisation ou de sortie est un noeud terminal s'il se trouve à la fin d'un flux ou d'une branche de flux. Vous ne pouvez pas connectez davantage de noeuds à un noeud terminal.

Remarque : Vous pouvez personnaliser la palette de noeuds. [Pour plus d'informations, reportez-vous à la section Personnalisation de la palette Noeuds dans le chapitre 12 sur p. 242.](#)

Ajout de noeuds à un flux

Pour ajouter des noeuds à un flux à partir de la palette de noeuds, vous disposez de plusieurs méthodes :

- Double-cliquez sur un noeud de la palette. *Remarque* : un double-clic sur un noeud permet de le relier automatiquement au flux en cours. [Pour plus d'informations, reportez-vous à la section Connexion de noeuds dans un flux sur p. 40.](#)
- Faites glisser un noeud à partir de la palette vers l'espace de travail de flux.
- Cliquez sur un noeud de la palette, puis cliquez sur l'espace de travail de flux.
- Choisissez une option appropriée à partir du menu Insérer de IBM® SPSS® Modeler.

Une fois que vous avez ajouté un noeud à l'espace de travail de flux, double-cliquez sur le noeud pour afficher sa boîte de dialogue. Les options disponibles dépendent du type de noeud que vous ajoutez. Pour obtenir des informations sur les commandes de cette boîte de dialogue, cliquez sur le bouton Aide.

Suppression de noeuds

Pour supprimer un noeud du flux de données, cliquez dessus, puis appuyez sur la touche Suppr, ou cliquez dessus avec le bouton droit de la souris et sélectionnez Supprimer dans le menu.

Connexion de noeuds dans un flux

Pour que les noeuds ajoutés à l'espace de travail de flux forment un flux de données, ils doivent être reliés. Les connexions entre les noeuds indiquent la direction des données transmises d'une opération à la suivante. Il existe plusieurs façons de relier des noeuds pour former un flux : en double-cliquant, en utilisant le bouton central de la souris ou manuellement.

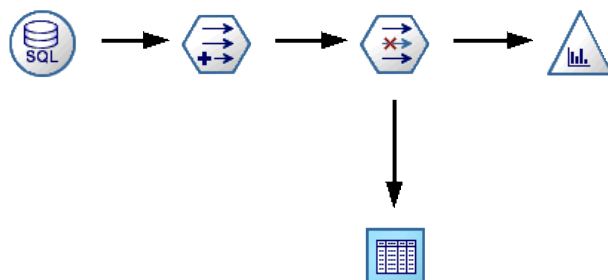
Pour ajouter et relier des noeuds en double-cliquant

La façon la plus simple de former un flux consiste à double-cliquer sur les noeuds de la palette. Cette méthode permet de relier automatiquement le nouveau noeud au noeud sélectionné dans l'espace de travail de flux. Par exemple, si l'espace de travail contient un noeud SGBD, vous pouvez le sélectionner et double-cliquer sur le noeud à ajouter depuis la palette (par exemple, un noeud Calculer). Cette action relie automatiquement le noeud Calculer au noeud SGBD existant. Vous pouvez répéter ce processus jusqu'à ce que vous ayez connecté un noeud terminal, comme

un noeud Histogramme ou Table. A ce stade, tout nouveau noeud sera connecté au dernier noeud non terminal en amont.

Figure 5-2

Flux créé en double-cliquant sur les noeuds à partir des palettes

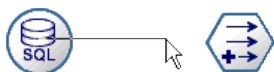


Pour relier des noeuds en utilisant le bouton central de la souris

Dans l'espace de travail de flux, cliquez sur un noeud et faites-le glisser vers un autre noeud en utilisant le bouton central de la souris (si votre souris ne dispose pas d'un bouton central, vous pouvez simuler l'action de ce bouton en appuyant sur la touche Alt tout en faisant glisser la souris d'un noeud vers un autre).

Figure 5-3

Utilisation du bouton central de la souris pour relier des noeuds



Pour relier des noeuds manuellement

Si votre souris ne dispose pas d'un bouton central et que vous préférez relier les noeuds manuellement, vous pouvez utiliser le menu contextuel d'un noeud pour le relier à un autre noeud qui figure déjà dans l'espace de travail.

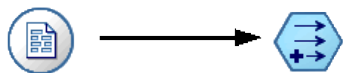
- ▶ Faites un clic droit sur le noeud à partir duquel vous souhaitez démarrer la connexion. Cette action ouvre le menu du noeud.
- ▶ Dans le menu, cliquez sur Connecter.
- ▶ L'icône de connexion apparaît sur le noeud de départ et sur le curseur. Cliquez sur le second noeud dans l'espace de travail pour relier les deux noeuds.

Figure 5-4

Connexion de noeuds en utilisant l'option Connecter du menu contextuel.



Figure 5-5
Noeuds reliés



Certaines règles doivent être respectées pour la connexion des noeuds. Vous recevrez un message d'erreur si vous tentez d'effectuer les types de connexion suivants :

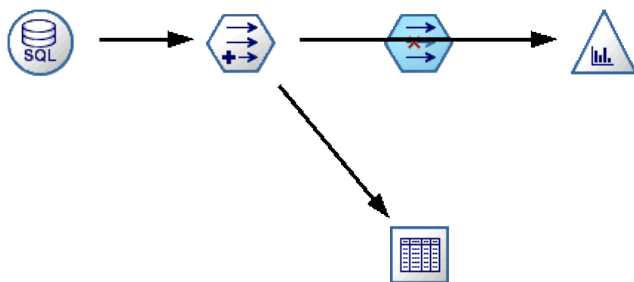
- Connexion vers un noeud source
- Connexion à partir d'un noeud terminal
- Noeud ayant plus de connexions d'entrée que le nombre maximal autorisé
- Connexion de deux noeuds déjà connectés
- Circularité (renvoi des données au noeud dont elles sont issues)

Contournement de noeuds dans un flux

Lorsque vous contournez un noeud du flux de données, toutes ses connexions d'entrée et de sortie sont remplacées par des connexions directes entre ses noeuds d'entrée et ses noeuds de sortie. Si le noeud ne dispose pas à la fois de connexions d'entrée et de sortie, toutes ses connexions sont alors supprimées sans être redirigées.

Par exemple, un flux qui calcule un nouveau champ, filtre des champs, puis explore les résultats dans un histogramme et un tableau. Si vous souhaitez également voir le même graphique et le même tableau de données *avant* le filtrage des champs, vous pouvez ajouter de nouveaux noeuds Histogramme et Table au flux ou contourner le noeud Filtrer. Lorsque vous contournez le noeud Filtrer, les connexions au graphique et au tableau passent directement par le noeud Calculer. Le noeud Filtrer est déconnecté du flux.

Figure 5-6
Contournement d'un noeud Filtrer précédemment connecté



Pour contourner un noeud

- Dans l'espace de travail de flux, double-cliquez sur le noeud à contourner à l'aide du bouton central de la souris. Vous pouvez également utiliser Alt+double-clic.

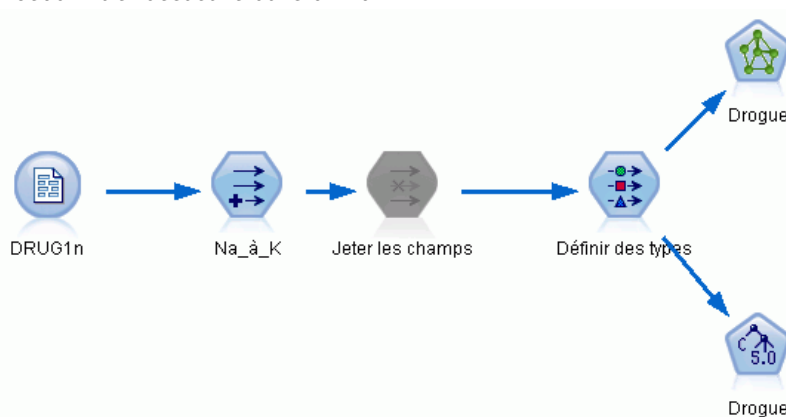
Remarque : vous pouvez annuler cette action en cliquant sur Annuler dans le menu Edition ou en appuyant sur Ctrl+Z.

Désactivation de noeuds dans un flux

Les noeuds d'exécution ayant une entrée unique au sein de flux peuvent être désactivés, ce qui permet d'ignorer le noeud au cours de l'exécution du flux. Cela vous évite d'avoir à supprimer ou à contourner le noeud et signifie que vous pouvez le laisser connecté aux noeuds restants. Vous pouvez toujours ouvrir et modifier les paramètres du noeud ; cependant les modifications ne prennent pas effet tant que vous n'avez pas de nouveau activé le noeud.

Par exemple, vous pouvez avoir un flux qui filtre plusieurs champs puis qui crée des modèles avec l'ensemble de données réduit. Si vous souhaitez créer aussi les mêmes modèles *sans* que les champs ne soient filtrés, afin de vérifier s'ils améliorent les résultats du modèle, vous pouvez désactiver le noeud Filtrer. Lorsque vous désactivez le noeud Filtrer, les connexions aux noeuds de modélisation sont directement transmises du noeud Calculer au noeud typer.

Figure 5-7
Noeud Filtrer désactivé dans un flux



Pour désactiver un noeud

- ▶ Dans l'espace de travail de flux, faites un clic droit sur le noeud à désactiver.
- ▶ Cliquez sur Désactiver le noeud dans le menu contextuel.

Vous pouvez aussi cliquer sur Noeud > Désactiver le noeud à partir du menu Edition. Si vous souhaitez de nouveau inclure le noeud dans le flux, cliquez sur Activer le noeud de la même manière.

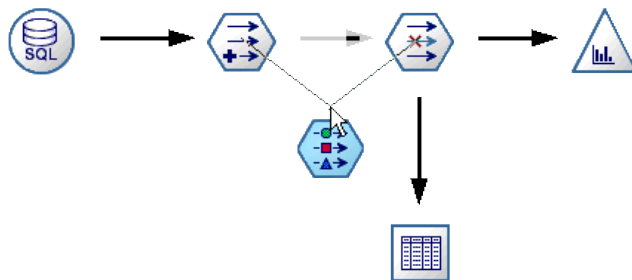
Remarque : vous pouvez annuler cette action en cliquant sur Annuler dans le menu Edition ou en appuyant sur Ctrl+Z.

vous pouvez annuler cette action en cliquant sur Annuler dans le menu Edition ou en appuyant sur Ctrl+Z.

Ajout de noeuds à des connexions existantes

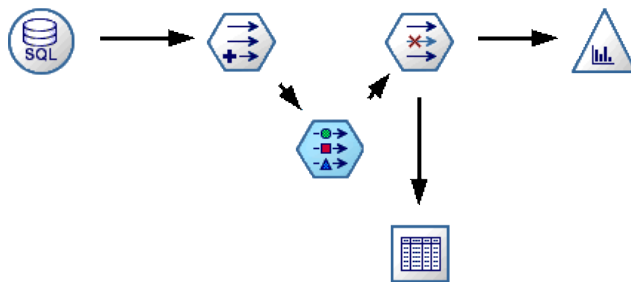
Vous pouvez ajouter un nouveau noeud entre deux noeuds connectés en faisant glisser la flèche qui les relie.

Figure 5-8
Connexion d'un nouveau noeud entre deux noeuds connectés



- ▶ Avec le bouton central de la souris, cliquez sur la flèche de connexion dans laquelle vous souhaitez insérer le noeud et faites-la glisser. Vous pouvez également simuler l'action du bouton du milieu de la souris en maintenant la touche Alt enfoncée pendant que vous cliquez et effectuez le déplacement.

Figure 5-9
Permet de créer un flux



- ▶ Faites glisser la connexion jusqu'au noeud à inclure et relâchez le bouton de la souris.

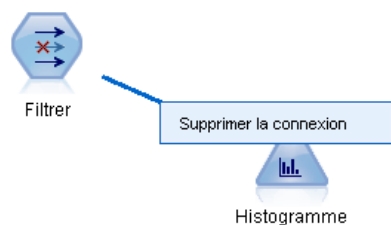
Remarque : vous pouvez supprimer les nouvelles connexions du noeud et restaurer les connexions initiales en **contournant** le noeud.

Suppression de connexions entre des noeuds

Pour supprimer la connexion entre deux noeuds :

- ▶ Faites un clic droit sur la flèche de connexion.
- ▶ Dans le menu, cliquez sur Supprimer la connexion.

Figure 5-10
Suppression de la connexion entre des noeuds dans un flux



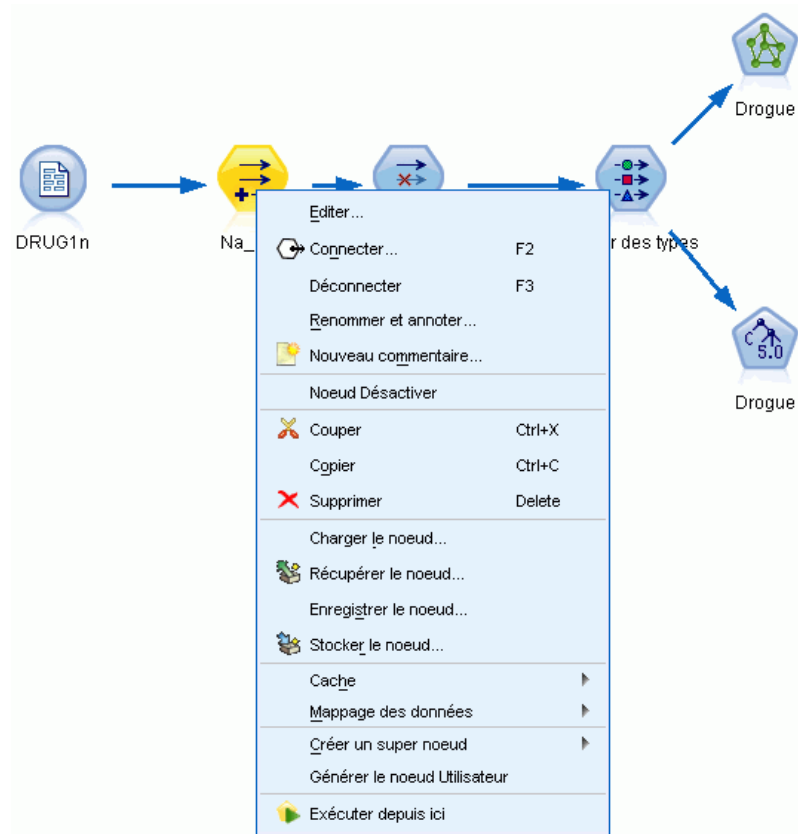
Pour supprimer toutes les connexions en direction et en provenance d'un noeud, procédez comme suit :

- Sélectionnez le noeud et appuyez sur F3.
- Sélectionnez le noeud puis, dans le menu principal, cliquez sur :
Edition > Noeud > Déconnecter

Définition des options des noeuds

Une fois que vous avez créé et connecté les noeuds, vous pouvez les personnaliser de plusieurs façons. Cliquez sur le noeud avec le bouton droit de la souris, puis sélectionnez l'une des options de menu.

Figure 5-11
Options du menu contextuel pour les noeuds



- Cliquez sur Editer pour ouvrir la boîte de dialogue correspondant au noeud sélectionné.
- Cliquez sur Connecter pour connecter manuellement deux noeuds.
- Cliquez sur l'option Déconnecter pour supprimer tous les liens provenant et partant du noeud.
- Cliquez sur l'option Renommer et annoter pour faire apparaître l'onglet Annotations de la boîte de dialogue de modification.
- Cliquez sur Nouveau commentaire pour ajouter un commentaire relatif au noeud. [Pour plus d'informations, reportez-vous à la section Ajout de commentaires et d'annotations à des noeuds et à des flux sur p. 70.](#)
- Cliquez sur Désactiver le noeud pour « masquer » le noeud au cours du traitement. Pour rendre le noeud visible de nouveau pour le traitement, cliquez sur Activer le noeud. [Pour plus d'informations, reportez-vous à la section Désactivation de noeuds dans un flux sur p. 43.](#)
- Cliquez sur Couper ou Supprimer pour supprimer les noeuds sélectionnés de l'espace de travail de flux. *Remarque* : cliquer sur Couper vous permet de coller des noeuds, mais pas l'option Supprimer.
- Cliquez sur l'option Copier le noeud pour effectuer une copie du noeud sans ses connexions. Il peut être ajouté à un nouveau flux ou à un flux existant.
- Cliquez sur l'option Charger le noeud pour ouvrir un noeud enregistré et charger ses options dans le noeud sélectionné. *Remarque* : les noeuds doivent être du même type.

- Cliquez sur Extraire le noeud pour extraire le noeud à partir d'un IBM® SPSS® Collaboration and Deployment Services Repository connecté. [Pour plus d'informations, reportez-vous à la section Accès aux objets du IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 5 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2*.](#)
- Cliquez sur Enregistrer le noeud pour enregistrer les détails du noeud dans un fichier. Vous ne pouvez charger les détails du noeud que dans un autre noeud du même type.
- Cliquez sur Stocker le noeud pour stocker le noeud sélectionné dans un IBM SPSS Collaboration and Deployment Services Repository connecté. [Pour plus d'informations, reportez-vous à la section Accès aux objets du IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 5 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2*.](#)
- Cliquez sur Cache pour faire apparaître les options de mise en cache du noeud sélectionné.
- Cliquez sur Mappage des données pour faire apparaître les options de mappage des données sur une nouvelle source ou les options de spécifications des champs obligatoires.
- Cliquez sur Créer un super noeud pour faire apparaître les options de création d'un super noeud dans le flux en cours. [Pour plus d'informations, reportez-vous à la section Création de super noeuds dans le chapitre 9 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)
- Cliquez sur Générer le noeud Utilisateur pour remplacer le noeud sélectionné. Les exemples générés par ce noeud auront les mêmes champs que le noeud actuel. [Pour plus d'informations, reportez-vous à la section Noeud Utilisateur dans le chapitre 2 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)
- Cliquez sur Exécuter à partir d'ici pour exécuter tous les noeuds terminaux situés en aval du noeud sélectionné.

Options de mise en cache des noeuds

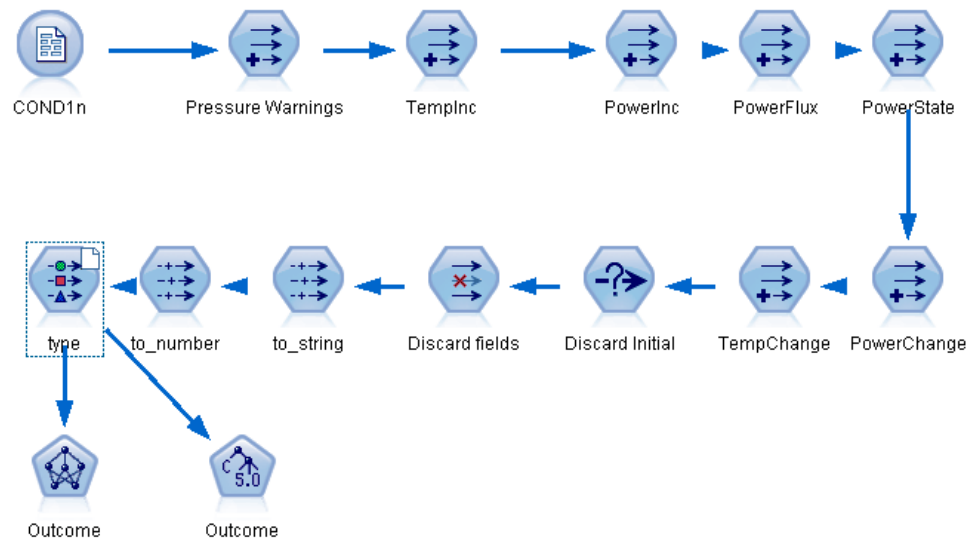
Pour optimiser l'exécution du flux, vous pouvez définir un **cache** sur n'importe quel noeud non terminal. Lorsque vous définissez un cache sur un noeud, le cache est alimenté par les données passant par le noeud lors de l'exécution suivante du flux de données. Les données sont alors lues à partir du cache (stocké sur le disque dans un répertoire temporaire) et non à partir de la source de données.

La mise en cache est plus utile après une opération de longue durée telle que le tri, la fusion ou l'agrégation. Par exemple, supposez que vous ayez un noeud source configuré pour la lecture des données relatives aux ventes à partir de la base de données et un noeud Agréger qui fournit des résultats de ventes par zone géographique. Vous pouvez définir un cache sur le noeud Agréger plutôt que sur le noeud source pour que le cache stocke uniquement les données agrégées et non l'ensemble des données.

Remarque : La mise en cache au niveau des noeuds source, qui stocke simplement une copie des données originales telle qu'elle est lue dans IBM® SPSS® Modeler, n'améliorera pas les performances dans la plupart des cas.

Lorsque la mise en cache d'un noeud est activée, une petite icône en forme de document apparaît dans l'angle supérieur droit. Une fois les données mises en cache au niveau du noeud, l'icône de document devient verte.

Figure 5-12
Mise en cache au niveau du noeud Typer pour stocker les champs nouvellement calculés



Pour activer un cache

- ▶ Dans l'espace de travail de flux, cliquez avec le bouton droit de la souris sur le noeud, puis cliquez sur Cache dans le menu.
- ▶ Dans le sous-menu de mise en cache, cliquez sur Activer.
- ▶ Vous pouvez désactiver le cache en cliquant sur le noeud avec le bouton droit de la souris et en cliquant sur Désactiver dans le sous-menu de mise en cache.

Mise en mémoire cache des noeuds dans une base de données

Dans le cas des flux exécutés dans une base de données, les données peuvent être mises en mémoire cache en milieu de flux dans une table temporaire de la base de données, plutôt que dans le système de fichiers. Vous pouvez combiner cela à la fonction d'optimisation SQL et accroître ainsi les performances de manière significative. Par exemple, la sortie d'un flux qui fusionne plusieurs tables afin de créer une vue de Data mining peut être mise en mémoire cache et réutilisée selon les besoins. La génération automatique du code SQL pour tous les noeuds en aval peut améliorer encore les performances.

Lors de l'utilisation de la mise en mémoire cache de la base de données avec des chaînes d'une longueur supérieure à 256 caractères, vérifiez qu'il existe un noeud Typer en amont depuis la mise en mémoire cache du noeud et que les valeurs des champs sont lues. Cela permet de garantir que la colonne correspondante dans la table temporaire est définie sur la largeur appropriée pour la prise en charge des chaînes.

Pour tirer profit de la mise en mémoire cache de base de données, vous devez activer les fonctions d'optimisation SQL et de mise en mémoire cache de base de données. Les paramètres d'optimisation du serveur annulent ceux définis pour le client. [Pour plus d'informations, reportez-vous à la section Définition des options d'optimisation dans le chapitre 12 sur p. 237.](#)

Une fois la mise en mémoire cache de base de données activée, il vous suffit de cliquer à l'aide du bouton droit de la souris sur un noeud non terminal afin de mettre les données en mémoire cache à ce stade du flux ; le cache est automatiquement créé dans la base de données lors de l'exécution suivante du flux. Si la mise en mémoire cache de base de données ou l'optimisation SQL n'est pas activée, le cache est écrit dans le système de fichiers.

Remarque : Les bases de données suivantes prennent en charge les tables temporaires à des fins de mise en mémoire cache : DB2, Netezza, Oracle, SQL Server et Teradata. Les autres bases de données utilisent une table standard pour la mise en mémoire cache de base de données. Si nécessaire, le code SQL peut être personnalisé pour des bases de données spécifiques : contactez l'assistance technique.

Pour vider un cache

Lorsque le cache d'un noeud est vide, une icône en forme de document blanc apparaît sur le noeud. Lorsque le cache est saturé, l'icône en forme de document devient entièrement verte. Pour remplacer le contenu du cache, vous devez au préalable vider le cache, puis réexécuter le flux de données de façon à le remplir de nouveau.

- ▶ Dans l'espace de travail de flux, cliquez avec le bouton droit de la souris sur le noeud, puis cliquez sur Cache dans le menu.
- ▶ Dans le sous-menu de mise en cache, cliquez sur Vider.

Pour enregistrer un cache

Vous pouvez enregistrer le contenu d'un cache dans un fichier de données IBM® SPSS® Statistics (*.sav). Vous pouvez ensuite recharger le fichier en tant que cache ou définir un noeud utilisant le fichier de cache comme source de données. Vous pouvez également charger un cache que vous avez enregistré pour un autre projet.

- ▶ Dans l'espace de travail de flux, cliquez avec le bouton droit de la souris sur le noeud, puis cliquez sur Cache dans le menu.
- ▶ Dans le sous-menu de mise en cache, cliquez sur Enregistrer le cache.
- ▶ Dans la boîte de dialogue Enregistrer le cache, choisissez l'emplacement d'enregistrement du fichier de cache.
- ▶ Entrez un nom dans la zone de texte Fichier.
- ▶ Assurez-vous que l'extension *.sav est sélectionnée dans la liste Type de fichiers, puis cliquez sur Enregistrer.

Pour charger un cache

Si vous avez enregistré un fichier de cache avant de le supprimer du noeud, vous pouvez le recharger.

- ▶ Dans l'espace de travail de flux, cliquez avec le bouton droit de la souris sur le noeud, puis cliquez sur Cache dans le menu.

- Dans le sous-menu de mise en cache, cliquez sur Charger le cache.
- Dans la boîte de dialogue Charger le cache, sélectionnez l'emplacement du fichier de cache et cliquez sur le bouton Charger.

Aperçu des données dans les noeuds

Pour vous assurer que les données seront modifiées comme vous le désirez lors de la création du flux, vous pouvez envoyer vos données dans un noeud Table à chaque étape importante. Pour vous éviter cela, vous pouvez générer un aperçu de chaque noeud qui affiche un échantillon des données qui seront créées, ce qui permet de gagner du temps passé à la création de chaque noeud.

Pour des noeuds en amont d'un nugget de modèle, l'aperçu affiche les champs d'entrée ; pour un nugget de modèles ou des noeuds en aval du nugget (sauf pour les noeuds terminaux), l'aperçu affiche les champs d'entrée et les champs générés.

Le nombre de lignes affichées par défaut est de 10, mais vous pouvez le modifier dans les propriétés du flux. [Pour plus d'informations, reportez-vous à la section Définition des options de flux sur p. 51.](#)

Figure 5-13
Aperçu des données à partir d'un nugget de modèle

	year_built	volume_interior	volume_other	lot_size	taxable_value	\$XR-taxable_value	\$XRE-taxable_value
1	979	166	11	100	90500	105523.184	1526.604
2	988	603	73	497	420000	355497.802	14929.477
3	987	303	75	91	152500	147713.631	1733.849
4	926	228	12	55	92000	69313.952	8116.212
5	988	666	145	441	390000	367551.491	16353.591
6	970	563	60	800	435000	395173.708	4759.398
7	902	355	41	225	130000	109865.220	3694.157
8	972	468	78	625	380500	334714.268	6367.557
9	986	315	14	107	158000	145149.106	1263.839
10	975	282	27	141	134000	138747.302	1892.730

Dans le menu Générer, vous pouvez créer plusieurs types de noeuds. [Pour plus d'informations, reportez-vous à la section Navigateur du noeud Table dans le chapitre 6 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Verrouillage des noeuds

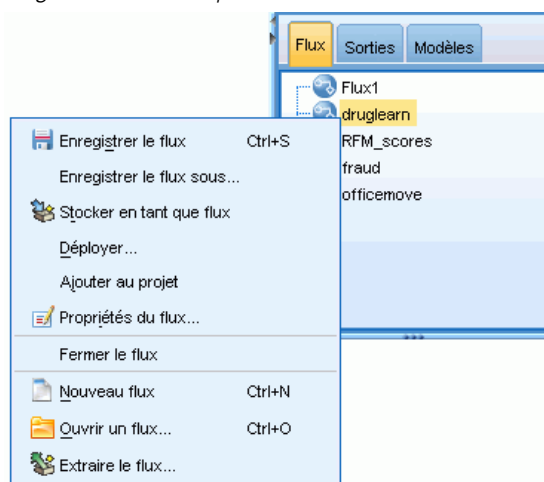
Pour empêcher les autres utilisateurs de modifier les paramètres d'un ou de plusieurs noeuds dans un flux, vous pouvez encapsuler le ou les noeuds dans un type de noeud spécial appelé super noeud, puis verrouiller le super noeud en appliquant une protection par mot de passe. [Pour plus d'informations, reportez-vous à la section Verrouillage des super noeuds dans le chapitre 9 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Utilisation des flux

Une fois les noeuds source, exécution et terminal connectés dans l'espace de travail de flux, le flux est créé. En tant qu'ensembles de noeuds, les flux peuvent être enregistrés, annotés et ajoutés à des projets. Vous pouvez également définir de nombreuses options de flux, telles que l'optimisation, les options de date et d'heure, les paramètres et les scripts. Ces propriétés seront abordées dans les rubriques suivantes.

Dans IBM® SPSS® Modeler, vous pouvez utiliser et modifier plusieurs flux de données en même temps. La partie droite de la fenêtre principale contient le panneau des Gestionnaires qui vous aide à parcourir les flux, les résultats et les modèles actuellement ouverts. Si vous ne pouvez pas consulter le panneau des Gestionnaires, cliquez sur Gestionnaires dans le menu Affichage, puis cliquez sur l'onglet Flux.

Figure 5-14
Onglet Flux dans le panneau des Gestionnaires avec les options du menu contextuel



Dans cet onglet, vous pouvez effectuer les opérations suivantes :

- Accéder aux flux.
- Enregistrer les flux.
- Enregistrer les flux dans le projet en cours.
- Fermer les flux.
- Ouvrir de nouveaux flux.
- Stocker et extraire les flux d'un référentiel IBM SPSS Collaboration and Deployment Services (si vous en disposez). [Pour plus d'informations, reportez-vous à la section A propos du IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 159.](#)

Dans l'onglet Flux, cliquez avec le bouton droit de la souris sur un flux pour accéder à ces options.

Définition des options de flux

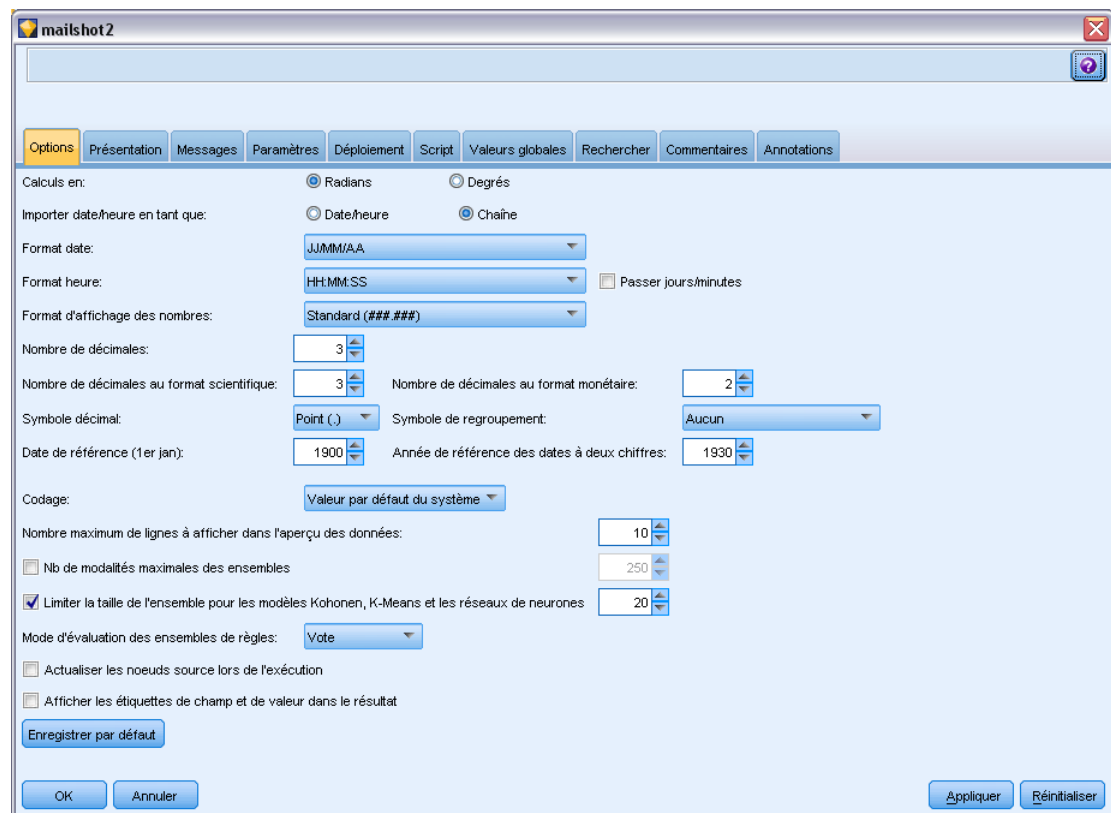
Vous pouvez spécifier un certain nombre d'options pour le flux en cours. La plupart de ces options s'appliquent aux expressions CLEM.

Pour définir les options de flux

- ▶ Dans le menu Fichier, cliquez sur Propriétés du flux (ou sélectionnez le flux dans l'onglet Flux du panneau des Gestionnaires, faites un clic droit puis cliquez sur Propriétés du flux dans le menu contextuel).
- ▶ Cliquez sur l'onglet Options.

Sinon, dans le menu Outils, cliquez sur :
Propriétés du flux > Options

Figure 5-15
Onglet Options dans la boîte de dialogue des propriétés du flux



Calculs en. Sélectionnez Radians ou Degrés comme unité de mesure à utiliser dans les expressions CLEM trigonométriques.

Importer date/heure en tant que. Indiquez si vous utilisez le stockage date/heure pour les champs date/heure ou si vous importez ces champs en tant que variables de chaîne.

Format date. Sélectionnez le format de date à utiliser pour les champs de stockage de date ou lorsque les chaînes sont interprétées comme des dates par les fonctions de date CLEM.

Format heure. Sélectionnez le format d'heure à utiliser pour les champs de stockage d'heure ou lorsque les chaînes sont interprétées comme des heures par les fonctions d'heure CLEM.

Passer jours/minutes. En ce qui concerne les formats temporels, indiquez si les différences horaires négatives doivent être interprétées comme se référant au jour précédent ou à l'heure précédente.

Format d'affichage des nombres. Vous pouvez sélectionner les formats d'affichage standard (#####.###), scientifique (#.###E+##) ou monétaire (### ## €).

Nombre de décimales (au format standard, scientifique ou monétaire). Pour les formats d'affichage des nombres, indique le nombre de décimales à utiliser lorsque des nombres réels sont affichés ou imprimés. Cette option apparaît séparément pour chaque format d'affichage.

Symbole décimal. Sélectionnez une virgule (,) ou un point (.) en tant que séparateur décimal.

Symbole de regroupement. Pour les formats d'affichage des nombres, sélectionnez le symbole permettant de regrouper des valeurs (par exemple, l'espace dans 3 000 00). Vous avez le choix entre les options suivantes : aucun, point, virgule, espace et paramètres régionaux définis (auquel cas la valeur par défaut des paramètres régionaux actuels est utilisée).

Date de référence (1er jan). Sélectionnez la date de référence (toujours le 1er janvier) utilisée par les fonctions de date CLEM opérant avec une date unique.

Année de référence des dates à deux chiffres : Spécifiez l'année pivot pour ajouter les chiffres correspondant au siècle lorsque l'année est indiquée uniquement par deux chiffres. Par exemple, spécifiez 1930 comme année pivot suppose que 05/11/02 est dans l'année 2002. Le même paramètre utilise le 20ème siècle pour les dates après le 30 ; ainsi il sera supposé que la date 05/11/73 est en 1973.

Codage. Spécifiez la méthode par défaut du flux pour le codage du texte. (*Remarque* : s'applique à variable. Noeud source Fichier et noeud Export Fichier plat uniquement. Aucun autre noeud n'utilise ce paramètre ; la majorité des fichiers de données contiennent des informations de codage intégrées.) Vous pouvez choisir la valeur par défaut du système ou UTF-8. La valeur par défaut du système est indiquée dans le Panneau de configuration de Windows ou sur le serveur si le système est exécuté en mode réparti. [Pour plus d'informations, reportez-vous à la section Prise en charge d'Unicode dans IBM SPSS Modeler dans l'annexe B sur p. 268.](#)

Nombre maximum de lignes à afficher dans l'aperçu des données. Spécifiez le nombre de lignes à afficher lorsqu'un aperçu des données est demandé pour un noeud. [Pour plus d'informations, reportez-vous à la section Aperçu des données dans les noeuds sur p. 50.](#)

Nombre maximum de membres pour les champs nominaux. Sélectionnez cette option pour spécifier le nombre maximal de membres pour les champs nominaux (ensemble) au-delà duquel le type du champ devient **Sans type**. Cette option peut être utile lorsque vous utilisez des champs nominaux volumineux. *Remarque* : Lorsque le niveau de mesure d'un champ est configuré sur **Sans type**, son rôle est automatiquement configuré sur **Aucun**. Cela signifie que les champs ne sont pas disponibles pour la modélisation.

Limiter la taille de l'ensemble pour les modèles Kohonen, K-Means et les réseaux de neurones. Sélectionnez cette option pour spécifier le nombre maximal de membres des champs nominaux utilisés dans les réseaux de neurones, les réseaux Kohonen et les modèles K-Means. La taille d'ensemble par défaut est 20. Si un nombre est supérieur à cette valeur, le champ est ignoré et un avertissement apparaît fournissant des informations sur le champ en question.

Mode d'évaluation des ensembles de règles. Détermine le mode d'évaluation des modèles des ensembles de règles. Par défaut, les ensembles de règles utilisent la valeur Vote pour combiner les prévisions à partir de règles individuelles et déterminer la prévision définitive. Pour que les ensembles de règles utilisent par défaut la première règle, sélectionnez Premier résultat. [Pour plus d'informations, reportez-vous à la section Nugget Modèle d'ensemble de règles dans le chapitre 6 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#) Notez que cette option ne s'applique pas aux modèles Liste de décision, qui utilisent toujours la première correspondance telle que définie par l'algorithme.

Rafraîchir les noeuds source lors de l'exécution. Sélectionnez cette option pour rafraîchir automatiquement tous les noeuds source lors de l'exécution du flux actuel. Cette action s'apparente à celle obtenue en cliquant sur le bouton Rafraîchir d'un noeud source, à la différence que cette option rafraîchit automatiquement tous les noeuds source du flux actuel (à l'exception des noeuds Utilisateur).

Remarque : cette option permet de vider les caches des noeuds en aval même si les données n'ont pas été modifiées. Toutefois, le vidage ne se produit qu'une fois par exécution du flux, ce qui signifie que vous pouvez quand même utiliser les caches en aval comme stockage temporaire lors d'une exécution unique. Par exemple, supposons que vous ayez défini un cache de milieu de flux après une opération de calcul complexe, et que vous disposez de plusieurs graphiques et rapports liés en aval de ce noeud Calculer. Lors de l'exécution du flux, le cache situé au niveau du noeud Calculer est vidé, puis de nouveau rempli, mais uniquement pour le premier graphique ou rapport. Les noeuds terminaux suivants lisent les données du cache du noeud Calculer.

Afficher les étiquettes de champ et de valeur dans le résultat. Affiche les étiquettes de champ et de valeur dans les tableaux, graphiques et autres sorties. S'il n'existe pas d'étiquette, ce sont les noms de champ et les valeurs de données qui apparaissent. Les étiquettes sont désactivées par défaut. Toutefois, vous pouvez activer ou désactiver chacune d'elles dans les autres composants de IBM® SPSS® Modeler. Vous pouvez également afficher les étiquettes dans la fenêtre de sortie via un bouton bascule disponible dans la barre d'outils.

Figure 5-16

Icône de la barre d'outils permettant d'activer ou de désactiver les étiquettes de champ et de valeur



Enregistrer comme paramètres par défaut. Les options spécifiées s'appliquent uniquement au flux en cours. Cliquez sur ce bouton pour définir ces options comme paramètre par défaut de tous les flux.

Définition des options pour la présentation du flux

Dans l'onglet Présentation de la boîte de dialogue des propriétés du flux, vous pouvez définir plusieurs options concernant l'affichage et l'utilisation de l'espace de travail de flux.

Pour définir les options de présentation

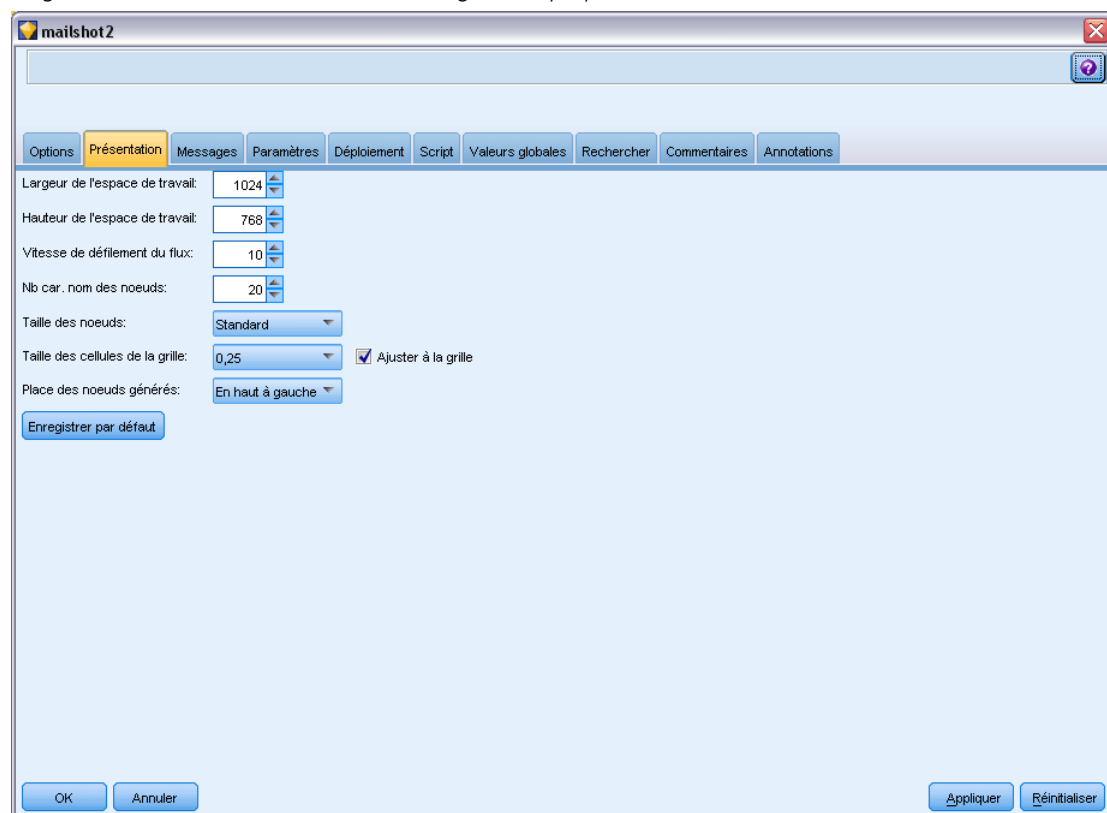
- Dans le menu Fichier, cliquez sur Propriétés du flux (ou sélectionnez le flux dans l'onglet Flux du panneau des Gestionnaires, faites un clic droit puis cliquez sur Propriétés du flux dans le menu contextuel).

- Cliquez sur l'onglet Présentation.

Sinon, dans le menu Outils, cliquez sur :
Propriétés du flux > Présentation

Figure 5-17

Onglet Présentation dans la boîte de dialogue des propriétés du flux



Largeur de l'espace de travail. Précisez la largeur de l'espace de travail de flux en pixels.

Hauteur de l'espace de travail . Précisez la hauteur de l'espace de travail de flux en pixels.

Vitesse de défilement du flux. Spécifiez la vitesse de défilement de l'espace de travail du flux pour contrôler la vitesse à laquelle défile le panneau de l'espace de travail de flux lorsque vous faites glisser un noeud d'un endroit à l'autre dans l'espace de travail. Plus les chiffres sont élevés, plus la vitesse de défilement est rapide.

Nb car. nom des noeuds. Indiquez la limite en nombre de caractères pour le nom des noeuds de l'espace de travail de flux.

Taille des noeuds. Indiquez si vous souhaitez afficher des icônes de noeud de grande ou de petite taille dans l'espace de travail de flux.

Taille des cellules de la grille. Sélectionnez la taille de cellule de grille dans la liste. Ce nombre est utilisé pour l'alignement des noeuds dans l'espace de travail de flux via une grille invisible. La taille de cellule de grille par défaut est de 0,25.

Ajuster à la grille. Sélectionnez cette option pour aligner les icônes sur une grille invisible (option sélectionnée par défaut).

Emplacement d'icône généré. Dans l'espace de travail, choisissez où placer les icônes des noeuds générés à partir des nuggets de modèle. L'emplacement par défaut se situe en haut à gauche.

Enregistrer comme paramètres par défaut. Les options spécifiées s'appliquent uniquement au flux en cours. Cliquez sur ce bouton pour définir ces options comme paramètre par défaut de tous les flux.

Affichage des messages d'opération de flux

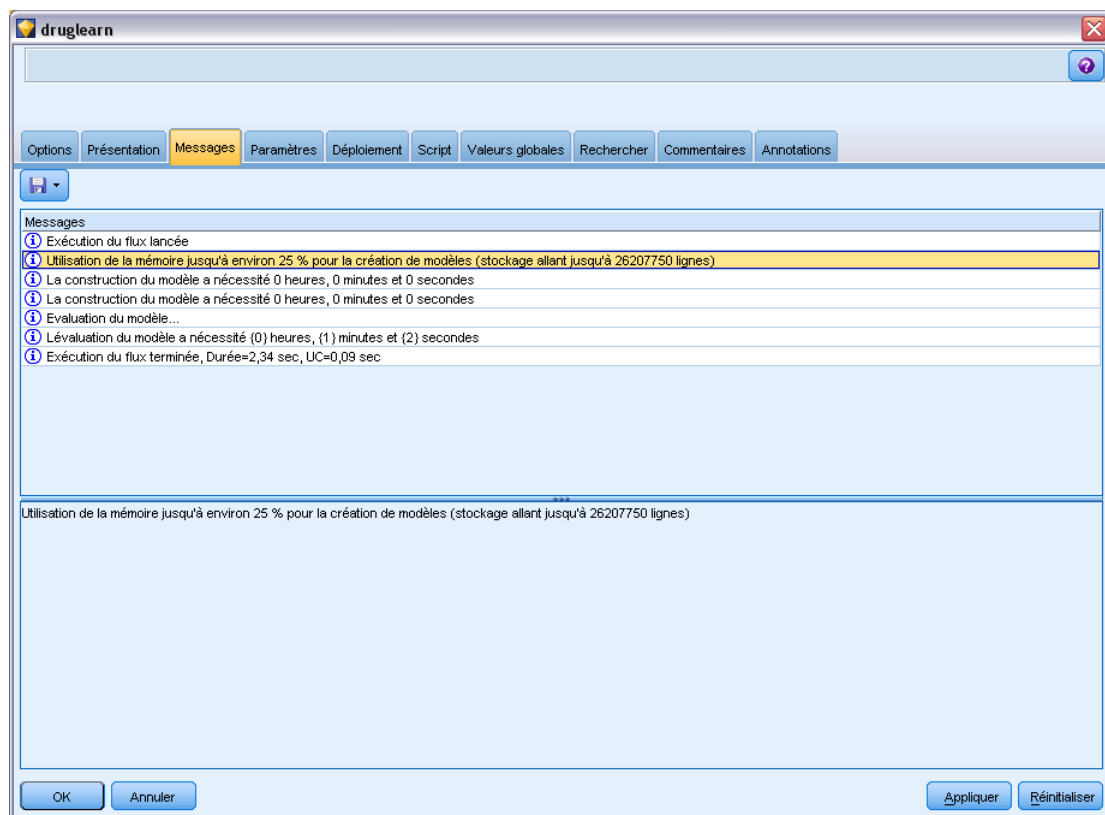
Les messages concernant les opérations de flux comme l'exécution, l'optimisation, la durée de création et d'évaluation d'un modèle peuvent être facilement consultés grâce à l'onglet Messages de la boîte de dialogue des propriétés du flux. Les messages d'erreur sont également indiqués dans ce tableau.

Pour afficher les messages de flux

- ▶ Dans le menu Fichier, cliquez sur Propriétés du flux (ou sélectionnez le flux dans l'onglet Flux du panneau des Gestionnaires, faites un clic droit puis cliquez sur Propriétés du flux dans le menu contextuel).
- ▶ Cliquez sur l'onglet Messages.

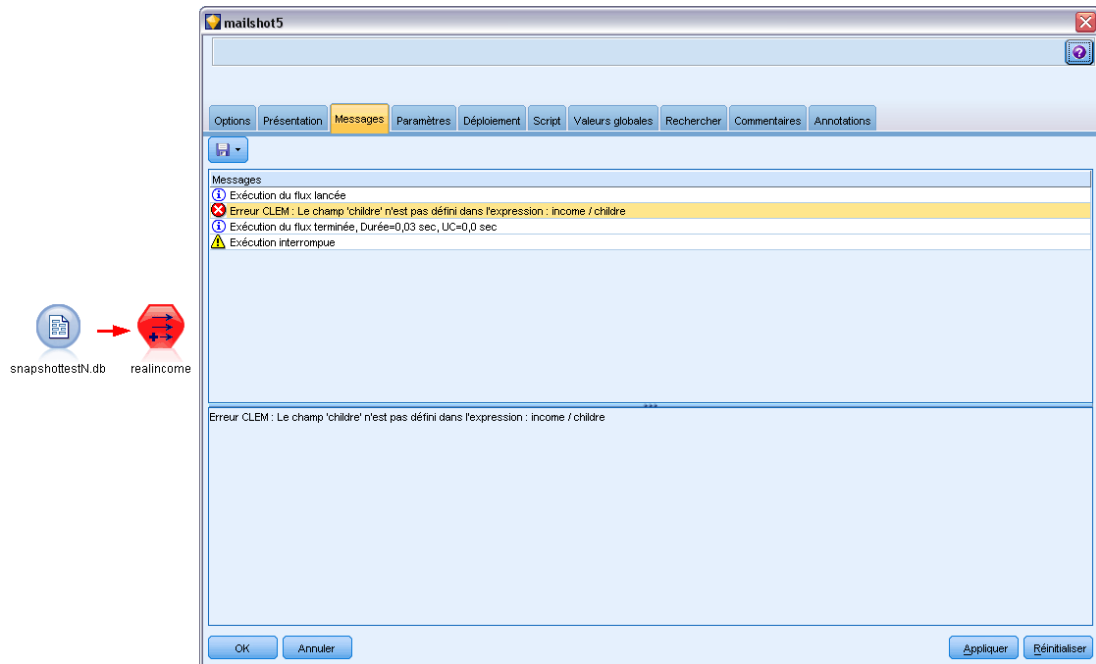
Sinon, dans le menu Outils, cliquez sur :
Propriétés du flux > Messages

Figure 5-18
Onglet Messages dans la boîte de dialogue des propriétés du flux



Les messages d'erreur sont affichés ici, avec les messages concernant les opérations sur le flux. Lorsque l'exécution du flux est interrompue à cause d'une erreur, cette boîte de dialogue apparaît dans l'onglet Messages et affiche le message d'erreur. En outre, le noeud comportant les erreurs est mis en évidence en rouge dans l'espace de travail de flux.

Figure 5-19
Exécution du flux avec indication d'erreur



Si les options de consignation et d'optimisation SQL sont activées dans la boîte de dialogue Options utilisateur, les informations relatives au code SQL généré apparaissent également. [Pour plus d'informations, reportez-vous à la section Définition des options d'optimisation dans le chapitre 12 sur p. 237.](#)

Vous pouvez enregistrer les messages indiqués ici se rapportant à un flux en cliquant sur Enregistrer les messages dans la liste déroulante du bouton d'enregistrement (à gauche, juste en-dessous de l'onglet Messages). Vous pouvez également effacer tous les messages d'un flux donné en cliquant sur Supprimer tous les messages dans la liste du bouton d'enregistrement.

Définition des paramètres de flux et de session

Vous pouvez définir des paramètres afin de les utiliser dans les expressions CLEM et pour la génération de scripts. Il s'agit, en effet, de variables définies par l'utilisateur qui sont enregistrées et conservées dans le flux, la session ou le super noeud en cours ; elles sont accessibles à partir de l'interface utilisateur, ainsi que via la fonction de génération de scripts. Si vous enregistrez un flux, par exemple, les paramètres définis pour ce dernier sont également enregistrés. (Cela permet de les différencier des variables de script locales qui ne peuvent être utilisées que dans le script dans lequel elles sont déclarées.) Dans le cadre de la génération de scripts, les paramètres sont souvent intégrés dans une expression CLEM et la valeur du paramètre est spécifiée dans le script.

La portée d'un paramètre dépend de l'endroit où ce dernier est défini :

- Les paramètres de flux peuvent être définis dans un script de flux ou dans la boîte de dialogue des propriétés du flux ; ils sont disponibles pour tous les noeuds du flux. Ils sont affichés sur la liste des paramètres dans le Générateur de formules.
- Les paramètres de session peuvent être définis dans un script autonome ou dans la boîte de dialogue des paramètres de session. Ils sont disponibles pour tous les flux utilisés dans la session actuelle (tous les flux répertoriés dans l'onglet Flux du panneau des gestionnaires).

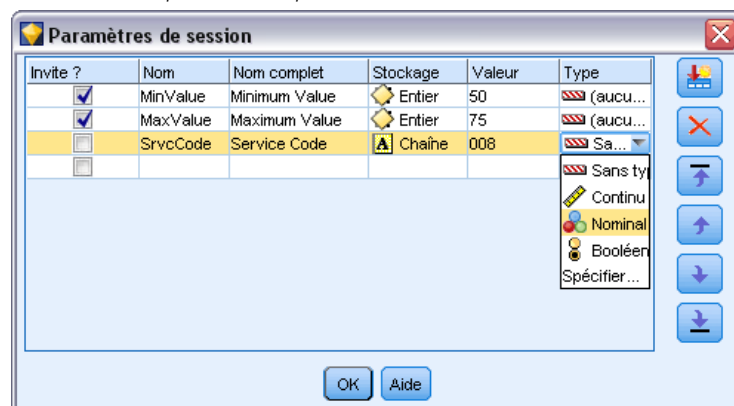
Il est également possible de définir des paramètres pour les super noeuds ; dans ce cas, les paramètres ne sont visibles que par les noeuds encapsulés au sein de ce super noeud. [Pour plus d'informations, reportez-vous à la section Définitions des paramètres de super noeud dans le chapitre 9 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

[Pour plus d'informations sur la définition et l'utilisation de paramètres dans les scripts, reportez-vous à Paramètres de flux, de session et de super noeud au le chapitre 3 .](#)

Pour définir des paramètres de flux et de session via l'interface utilisateur

- Pour définir des paramètres de flux, dans le menu principal, cliquez sur : Outils > Propriétés du flux > Paramètres
- Pour définir des paramètres de session, cliquez sur Définir les paramètres de session dans le menu Outils.

Figure 5-20
Définition des paramètres pour la session



Invite?. Cochez cette case si vous souhaitez que l'utilisateur soit invité à saisir une valeur pour ce paramètre au moment de l'exécution.

Nom. Les noms des paramètres sont répertoriés ici. Vous pouvez créer un paramètre en entrant un nom dans ce champ. Par exemple, pour créer un paramètre relatif à la température minimale, vous pouvez saisir Valeur min.. N'insérez pas le préfixe \$P-, qui indique un paramètre dans les expressions CLEM. Ce nom est également utilisé pour l'affichage dans le Générateur de formules de CLEM.

Nom complet. Répertorie le nom descriptif de chaque paramètre créé.

Stockage. Sélectionnez le type de stockage dans la liste. Le stockage indique le mode de stockage des valeurs de données dans le paramètre. Par exemple, si vous utilisez des valeurs commençant par des zéros à conserver (comme 008), vous devez sélectionner Chaîne comme type de stockage. Sinon, les zéros seront supprimés de la valeur. Les types de stockage disponibles sont les suivants : chaîne, entier, réel, temps, date et horodatage. Pour les paramètres de date, les valeurs doivent être définies à l'aide de la notation standard ISO telle qu'elle est présentée dans le paragraphe suivant.

Valeur. Indique la valeur actuelle du paramètre sélectionné. Modifiez ce paramètre selon les besoins. Pour les paramètres de date, les valeurs doivent être définies à l'aide de la notation standard ISO (soit, YYYY-MM-DD). Toute date définie dans un autre format est refusée.

Type (facultatif). Si vous prévoyez de déployer le flux vers une application externe, sélectionnez un niveau de mesure dans la liste. Sinon, il est conseillé de laisser la colonne *Type* en l'état. Si vous souhaitez spécifier des contraintes de valeur pour le paramètre, telles que des limites supérieures et inférieures d'un intervalle numérique, sélectionnez Spécifier dans la liste.

Vous ne pouvez définir les options de noms longs, de stockage et de type pour les paramètres que dans l'interface utilisateur. Il est impossible de définir ces options à l'aide de scripts.

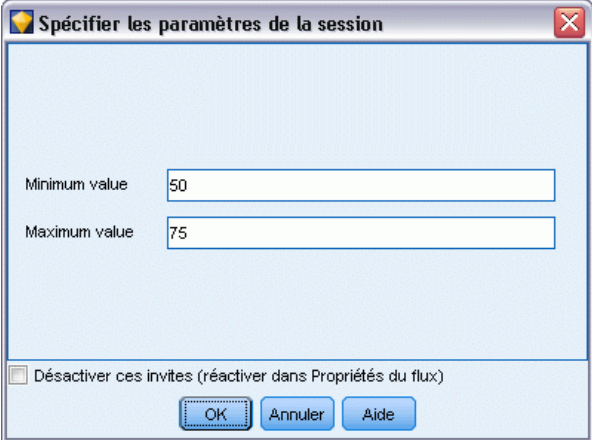
Cliquez sur les flèches à droite pour déplacer le paramètre sélectionné vers le haut ou le bas de la liste des paramètres disponibles. Utilisez le bouton de suppression (indiqué par un *X*) pour supprimer le paramètre sélectionné.

Spécification des invites d'exécution pour les valeurs de paramètre

Si vous possédez des flux où il est possible que vous soyez amené à entrer différentes valeurs pour le même paramètre en différentes occasions, vous pouvez spécifier des invites d'exécution pour un ou plusieurs flux ou des valeurs de paramètre de session.

Figure 5-21

Invite d'exécution pour les valeurs de paramètre



Spécifier les paramètres de la session

Minimum value 50

Maximum value 75

Désactiver ces invites (réactiver dans Propriétés du flux)

OK Annuler Aide

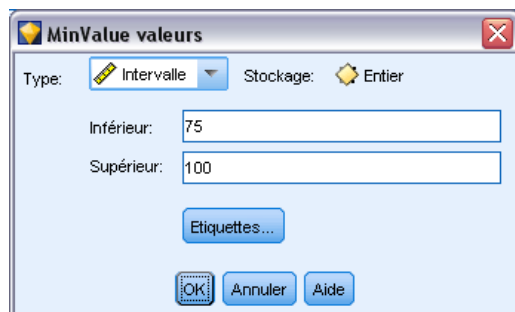
Paramètres. (Facultatif) Saisissez une valeur pour le paramètre, ou conservez la valeur par défaut s'il y en a une.

Désactiver ces invites. Cochez cette case si vous ne souhaitez pas que ces invites apparaissent lorsque vous exécutez le flux. Vous pouvez les réafficher en cochant la case Invite ? dans la boîte de dialogue Propriétés des flux ou Propriétés de la session où les paramètres ont été définis. [Pour plus d'informations, reportez-vous à la section Définition des paramètres de flux et de session sur p. 58.](#)

Spécification des contraintes des valeurs d'un type de paramètre

Vous pouvez faire en sorte que des contraintes de valeur pour un paramètre soient disponibles au cours du déploiement du flux dans une application externe lisant les flux de modélisation des données. Cette boîte de dialogue vous permet de spécifier les valeurs mises à la disposition des utilisateurs externes exécutant le flux. Dans la boîte de dialogue, les contraintes de valeur varient de manière dynamique en fonction du type de données. Les options présentées ici sont identiques à celles proposées pour les valeurs du noeud Typier.

Figure 5-22
Spécification des valeurs disponibles d'un paramètre



Type. Affiche le type de niveau de mesure actuellement sélectionné. Vous pouvez modifier cette valeur pour indiquer la façon dont vous souhaitez utiliser les paramètres dans IBM® SPSS® Modeler.

Stockage. Affiche le type de stockage, s'il est connu. Les types de stockage ne sont pas affectés par le niveau de mesure (Continu, Nominal ou Booléen) choisi pour travailler dans SPSS Modeler. Vous pouvez changer de type de stockage dans l'onglet Paramètres principal.

La partie inférieure de la boîte de dialogue change de manière dynamique en fonction du niveau de mesure sélectionné dans le champ Type.

Niveaux de mesure continus

Inférieur. Indiquez la limite inférieure des valeurs de paramètre.

Supérieur. Indiquez la limite supérieure des valeurs de paramètre.

Étiquettes : Vous pouvez spécifier des étiquettes pour toutes les valeurs d'un champ d'intervalle. Cliquez sur le bouton Étiquettes pour ouvrir une boîte de dialogue distincte permettant de spécifier les étiquettes de valeur.

Niveaux de mesure nominaux

Valeurs : Cette option vous permet de spécifier les valeurs d'un paramètre utilisé en tant que champ nominal. Ces valeurs ne sont pas introduites automatiquement dans le flux SPSS Modeler, mais utilisées dans une liste déroulante destinée aux applications de déploiement externes. A l'aide des flèches et du bouton Supprimer, vous pouvez modifier, réorganiser ou supprimer les valeurs existantes.

Niveaux de mesure booléens

Vrai. Spécifiez la valeur booléenne du paramètre lorsque la condition est respectée.

Faux. Spécifiez la valeur booléenne du paramètre lorsque la condition n'est pas respectée.

Étiquettes : Vous pouvez spécifier des étiquettes pour les valeurs d'un champ booléen.

Options de déploiement des flux

L'onglet Déploiement de la boîte de dialogue Propriétés du flux vous permet de spécifier des options pour le déploiement du flux en tant que scénario dans IBM® SPSS® Collaboration and Deployment Services afin de rafraîchir un modèle, de programmer une tâche automatisée ou une autre utilisation par IBM® SPSS® Decision Management ou Predictive Analytics 5.x. Pour être déployés, tous les flux nécessitent une branche de scoring avant de pouvoir être déployés. Les exigences et options supplémentaires dépendent du type de déploiement. [Pour plus d'informations, reportez-vous à la section Stockage et et déploiement des objets IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 161.](#)

Affichage des valeurs globales des flux

A l'aide de l'onglet Valeurs globales de la boîte de dialogue des propriétés du flux, vous pouvez afficher les valeurs globales définies pour le flux en cours. Les valeurs globales sont créées à l'aide d'un noeud V. globales (Valeurs globales). Elles permettent de déterminer des statistiques, telles que la moyenne, la somme ou l'écart-type des champs sélectionnés. [Pour plus d'informations, reportez-vous à la section Noeud V. globales \(Valeurs globales\) dans le chapitre 6 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

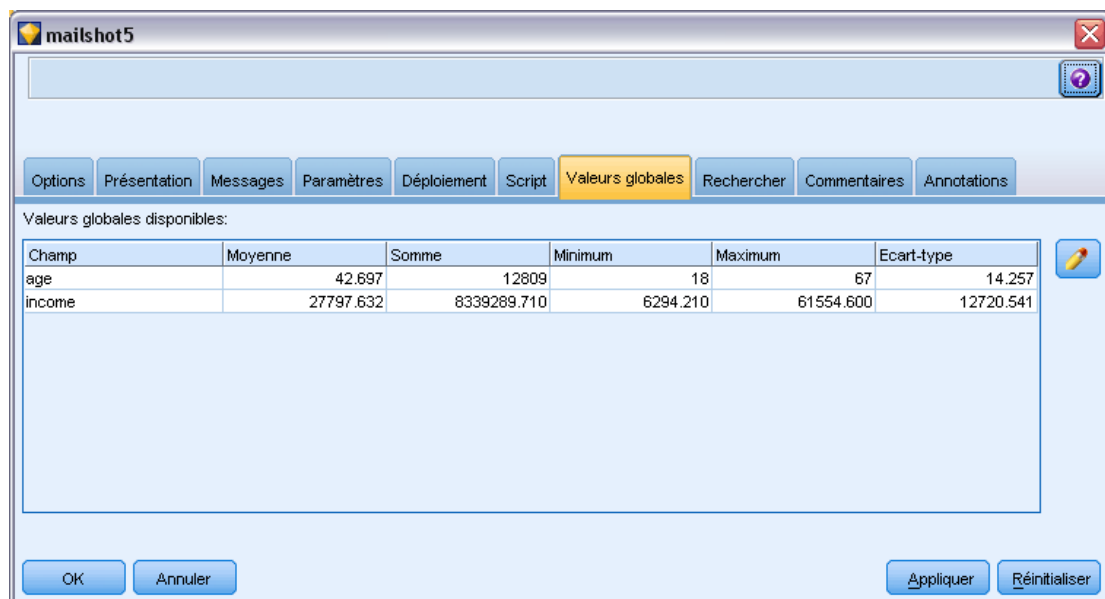
Une fois le noeud V. globales (Valeurs globales) exécuté, ces valeurs sont alors disponibles pour diverses utilisations dans les opérations de flux. [Pour plus d'informations, reportez-vous à la section Fonctions globales dans le chapitre 8 sur p. 155.](#)

Pour afficher les valeurs globales d'un flux

- ▶ Dans le menu Fichier, cliquez sur Propriétés du flux (ou sélectionnez le flux dans l'onglet Flux du panneau des Gestionnaires, faites un clic droit puis cliquez sur Propriétés du flux dans le menu contextuel).
- ▶ Cliquez sur l'onglet Globales.

Sinon, dans le menu Outils, cliquez sur :
Propriétés du flux > Valeurs globales

Figure 5-23
Affichage des valeurs globales disponibles pour le flux



Valeurs globales disponibles. Les valeurs globales disponibles sont répertoriées dans ce tableau. Vous ne pouvez pas éditer les valeurs globales à ce niveau. Toutefois, vous pouvez supprimer toutes les valeurs globales pour un flux en vous servant du bouton Effacer toutes les valeurs situé à droite du tableau.

Recherche de noeuds dans un flux

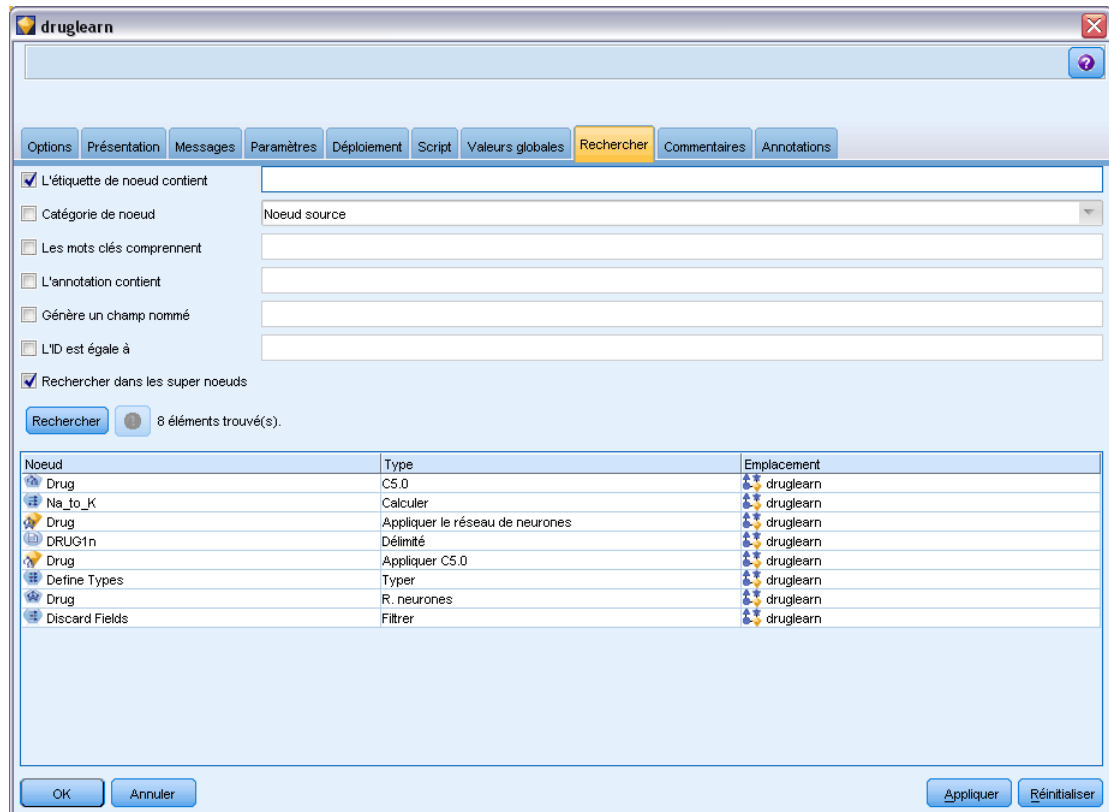
Vous pouvez rechercher des noeuds dans un flux en spécifiant un certain nombre de critères, comme le nom, la catégorie ou l'identificateur de ces noeuds. Cette option est particulièrement utile pour les flux complexes qui contiennent un grand nombre de noeuds.

Pour rechercher des noeuds dans un flux

- ▶ Dans le menu Fichier, cliquez sur Propriétés du flux (ou sélectionnez le flux dans l'onglet Flux du panneau des Gestionnaires, faites un clic droit puis cliquez sur Propriétés du flux dans le menu contextuel).
- ▶ Cliquez sur l'onglet Rechercher.

Sinon, dans le menu Outils, cliquez sur :
Propriétés du flux > Rechercher

Figure 5-24
Recherche de noeuds dans un flux



Vous pouvez spécifier plusieurs options afin de limiter la recherche, mais la recherche par ID de noeud (à l'aide du champ ID égal à) exclut les autres options.

L'étiquette de noeud contient. Cochez cette case et saisissez la totalité ou une partie de l'étiquette du noeud pour rechercher un noeud spécifique. Les recherches ne sont pas sensibles à la casse, et plusieurs mots peuvent être traités comme un seul extrait de texte.

Catégorie de noeud. Cochez cette case et cliquez sur une catégorie de la liste pour rechercher un type de noeud spécifique. Noeud d'exécution signifie un noeud de l'onglet Ops sur lignes ou Ops sur champs de la palette des noeuds ; Appliquer le noeud du modèle se rapporte à un nugget de modèle.

Les mots clés comprennent. Cochez cette case et saisissez un ou plusieurs mots-clés complets pour rechercher des noeuds qui contiennent ce texte dans le champ Mots-clés dans l'onglet Annotations de la boîte de dialogue Noeud. Le texte des mots-clés que vous saisissez doit correspondre parfaitement. Séparez plusieurs mots-clés par des points-virgules pour rechercher des alternatives (par exemple, saisissez proton;neutron pour rechercher tous les noeuds contenant l'un ou l'autre de ces mots-clés. [Pour plus d'informations, reportez-vous à la section Annotations sur p. 78.](#)

L'annotation contient. Cochez cette case et saisissez un ou plusieurs mots pour rechercher des noeuds qui contiennent ce texte dans la zone de texte principale de l'onglet Annotations de la boîte de dialogue Noeud. Les recherches ne sont pas sensibles à la casse, et plusieurs mots peuvent être traités comme un seul extrait de texte. [Pour plus d'informations, reportez-vous à la section Annotations sur p. 78.](#)

Génère un champ nommé. Cochez cette case et saisissez le nom d'un champ généré (par exemple, drug-\$C). Vous pouvez utiliser cette option pour rechercher des noeuds de modélisation qui génèrent un champ particulier. Saisissez un seul nom de champ, qui doit correspondre parfaitement.

ID est égal à. Cochez cette case et saisissez l'ID du noeud pour rechercher un noeud particulier avec cet identificateur (sélectionner cette option désactive toutes les options précédentes). Les ID de noeud sont attribués par le système lorsqu'un noeud est créé et peuvent être utilisés pour référencer un noeud à des fins de générations de scripts ou d'automatisation. Saisissez un seul ID de noeud, qui doit correspondre parfaitement. [Pour plus d'informations, reportez-vous à la section Annotations sur p. 78.](#)

Rechercher dans les super noeuds. Cette case est cochée par défaut, ce qui signifie que la recherche est effectuée sur des noeuds situés à la fois à l'intérieur et à l'extérieur des super noeuds. Décochez cette case si vous souhaitez effectuer une recherche sur les noeuds à l'extérieur des super noeuds uniquement, au niveau supérieur du flux.

Chercher. Après avoir spécifié toutes les options désirées, cliquez sur ce bouton pour commencer la recherche.

Les noeuds correspondants aux options spécifiées sont répertoriés dans la partie inférieure de la boîte de dialogue. Sélectionnez un noeud dans la liste pour le mettre en surbrillance dans l'espace de travail du flux.

Changement de nom des flux

A l'aide de l'onglet Annotations de la boîte de dialogue des propriétés du flux, vous pouvez ajouter des annotations descriptives à un flux et donner un nom personnalisé à ce dernier. Ces options sont utiles plus particulièrement lorsque vous générez des rapports sur les flux ajoutés au panneau de projet. [Pour plus d'informations, reportez-vous à la section Annotations sur p. 78.](#)

Descriptions de flux

Pour chaque flux que vous créez, IBM® SPSS® Modeler crée une description de flux contenant des informations concernant le contenu du flux. Ceci peut s'avérer utile si vous essayez de vérifier ce que réalise un flux mais que vous n'avez pas installé SPSS Modeler, par exemple lorsque vous accédez à un flux via IBM® SPSS® Collaboration and Deployment Services.

Figure 5-25
Ouverture de la section de la description de flux

IBM SPSS Modeler	
Document de flux	newschancart_commented
Créé le :	27 janvier 2006 01:17
Créé par :	llanders
Dernier enregistrement le :	28 mai 2010 20:51
Dernier enregistrement par :	djones

Contenu

[Description & Commentaire](#)
[Informations de scoring](#)
[Informations de modélisation](#)

Description & Commentaire

Commentaire
This stream predicts whether customers are likely to subscribe to our interactive news service, based on characteristics of customers who have already subscribed
This is the source data about our existing customers
This is where we tell the stream what we are trying to predict

La description du flux s'affiche sous la forme d'un document HTML se composant de plusieurs sections.

Informations générales sur le flux

Cette section contient le nom du flux ainsi que des détails sur le moment de la création du flux et de son dernier enregistrement.

Description et commentaires

Cette section comprend :

- toutes les annotations du flux (reportez-vous à [Annotations sur p. 78](#))
- tous les commentaires non connectés à des noeuds spécifiques
- tous les commentaires connectés à des noeuds dans les branches de modélisation et de scoring du flux

Informations de scoring

Cette section contient des informations sous divers en-têtes en rapport avec la branche de scoring du flux.

- **Commentaires** : comprend des commentaires liés uniquement à des noeuds dans la branche de scoring.
- **Entrées**. Répertorie les champs d'entrée, ainsi que leur type de stockage (par exemple, chaîne, entier, réel, etc.).
- **Sorties**. Répertorie les champs de sortie, y compris les champs supplémentaires générés par le noeud de modélisation, ainsi que leur type de stockage.
- **Paramètres**. Répertorie tous les paramètres en rapport avec la branche de scoring du flux et ceux qui peuvent être affichés ou modifiés à chaque fois que le modèle est évalué. Ces paramètres sont identifiés lorsque vous cliquez sur le bouton Paramètres de scoring dans l'onglet Déploiement de la boîte de dialogue Propriétés du flux.
- **Noeud Modèle**. Affiche le nom et le type du modèle (par exemple, R. neurones, Arbre C&R, etc.). Il s'agit du nugget de modèle sélectionné pour le champ noeud Modèle dans l'onglet Déploiement de la boîte de dialogue Propriétés du flux.
- **Détails des modèles**. Affiche des détails relatifs au nugget de modèle identifié sous l'en-tête précédent. Lorsque c'est possible, l'importance des variables prédictives et des graphiques d'évaluation du modèle sont inclus.

Informations de modélisation

Contient des informations concernant la branche modélisation du flux.

- **Commentaires** : Répertorie tous les commentaires ou annotations connectés aux noeuds de la branche de modélisation.
- **Entrées**. Répertorie les champs d'entrée ainsi que leur rôle dans la branche de modélisation (sous la forme d'une valeur de rôle de champ, par exemple, Entrée, Cible, Division, etc.).
- **Paramètres**. Répertorie tous les paramètres en rapport avec la branche de modélisation du flux et ceux qui peuvent être affichés ou modifiés à chaque fois que le modèle est mis à jour. Ces paramètres sont identifiés lorsque vous cliquez sur le bouton Paramètres du concepteur de modèle sur l'onglet Déploiement de la boîte de dialogue Propriétés du flux.
- **Noeud de modélisation**. Affiche le nom et le type du noeud de modélisation utilisé pour générer ou mettre à jour le modèle.

Aperçu des descriptions de flux

Vous pouvez afficher le contenu de la description d'un flux dans un navigateur Web en cliquant sur une option dans la boîte de dialogue Propriétés du flux. Le contenu de la description dépend des options que vous avez spécifiées dans l'onglet Déploiement de la boîte de dialogue. [Pour plus d'informations, reportez-vous à la section Options de déploiement des flux dans le chapitre 9 sur p. 188.](#)

Pour afficher une description de flux :

- ▶ Dans le menu IBM® SPSS® Modeler principal, cliquez sur :
Outils > Propriétés du flux > Déploiement
- ▶ Définissez le type de déploiement, le noeud de scoring désigné et tous les paramètres de scoring.
- ▶ Si le type de déploiement est Rafraîchissement de modèle, vous pouvez aussi sélectionner :
 - un noeud de modélisation et tous les paramètres de création de modèle
 - un nugget de modèle sur la branche de scoring du flux
- ▶ Cliquez sur le bouton Aperçu de la description du flux.

Exportation des descriptions de flux

Vous pouvez exporter le contenu de la description de flux dans un fichier HTML.

Pour exporter une description de flux :

- ▶ Dans le menu principal, cliquez sur :
Fichier > Exporter la description de flux
- ▶ Saisissez un nom pour le fichier HTML et cliquez sur Enregistrer.

Exécution des flux

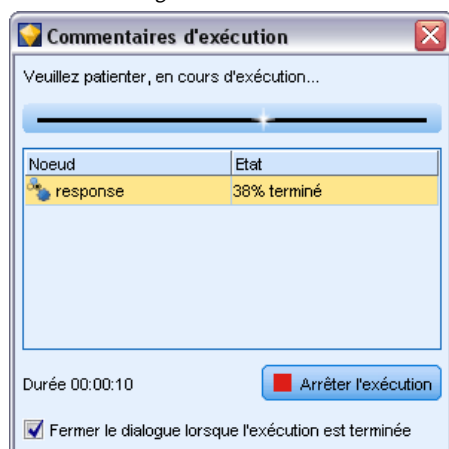
Après avoir spécifié les options nécessaires pour les flux et connecté les noeuds requis, vous pouvez exécuter le flux en exécutant les données via les noeuds du flux. Il existe plusieurs manières d'exécuter un flux dans IBM® SPSS® Modeler. Vous pouvez :

- Cliquez sur Exécuter dans le menu Outils.
- Cliquez sur un des boutons Exécuter... de la barre d'outils. Ces boutons vous permettent d'exécuter la totalité du flux ou simplement le noeud terminal sélectionné. [Pour plus d'informations, reportez-vous à la section Barre d'outils IBM SPSS Modeler dans le chapitre 3 sur p. 19.](#)
- Exécutez un flux de données unique en cliquant sur un noeud terminal à l'aide du bouton droit de la souris et en cliquant sur Exécuter dans le menu contextuel.
- Exécutez une partie d'un flux de données en cliquant sur un noeud non terminal à l'aide du bouton droit de la souris et en cliquant sur Exécuter à partir d'ici dans le menu contextuel. Cette action permet de n'exécuter que les opérations après le noeud sélectionné.

Pour arrêter l'exécution d'un flux en cours, vous pouvez cliquer sur le bouton d'arrêt rouge dans la barre d'outils ou cliquer sur Arrêter l'exécution du menu Outils.

Si un flux prend plus de trois secondes à s'exécuter, la boîte de dialogue Commentaires d'exécution s'affiche pour indiquer la progression.

Figure 5-26
Boîte de dialogue Commentaires d'exécution



Certains noeuds disposent d'affichages supplémentaires fournissant davantage d'informations sur l'exécution du flux. Vous pouvez les afficher en sélectionnant la ligne correspondante dans la boîte de dialogue. La première ligne est automatiquement sélectionnée.

Utilisation des modèles

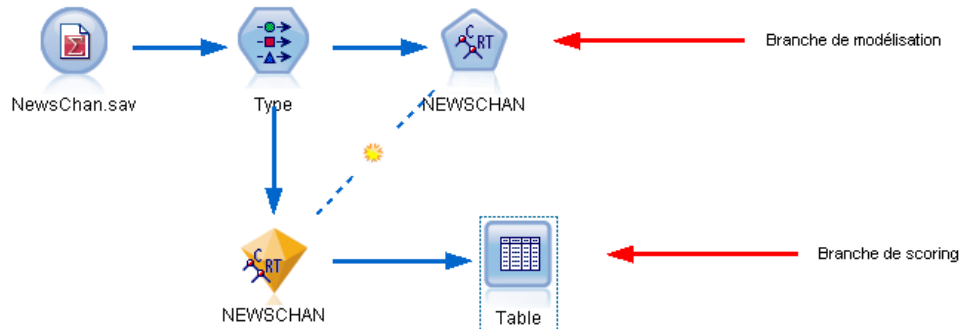
Si un flux comprend un noeud de modélisation (c'est à dire, d'un onglet Modélisation ou Base de données d'une palette de noeuds), un **nugget de modèle** est créé lors de l'exécution du flux. Un nugget de modèle est un conteneur pour un **modèle**, c'est-à-dire l'ensemble des règles, des formules ou des équations qui vous permettent de générer des prédictions concernant des données source et qui se trouve au cœur de l'analyse prédictive.

Figure 5-27
Nugget de modèle



Lorsque vous êtes parvenu à exécuter un noeud de modélisation, un nugget de modèle correspondant est placé dans l'espace de travail du flux, où il est représenté par une icône en forme de diamant doré (d'où le nom « nugget » (pépite en français)). Vous pouvez ouvrir le nugget et parcourir son contenu pour afficher les détails du modèle. Pour afficher les prédictions, vous attachez et exécutez un ou plusieurs noeuds terminaux, le résultat à partir duquel les prédictions sont présentées sous une forme lisible.

Figure 5-28
Branches de modélisation et de scoring dans un flux



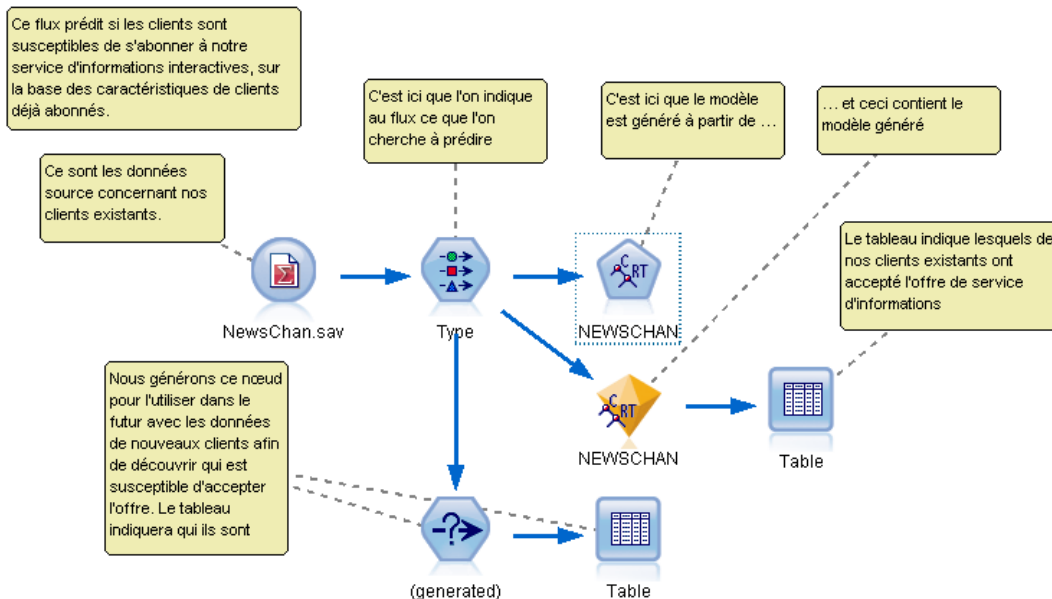
Un flux de modélisation standard se compose de deux branches. La **branche de modélisation** contient le noeud de modélisation ainsi que les noeuds source et d'exécution qui le précèdent. La **branche de scoring** est créée lorsque vous exécutez le noeud de modélisation et qu'il contient le nugget de modèle et le noeud terminal ou les noeuds que vous utilisez pour afficher les prédictions.

Pour plus d'informations, reportez-vous au guide des *noeuds de modélisation de IBM® SPSS® Modeler*.

Ajout de commentaires et d'annotations à des noeuds et à des flux

Il se peut que vous ayez à décrire un flux à d'autres personnes de votre entreprise. Pour vous aider, vous pouvez joindre des explications aux flux, noeuds et nuggets de modèle.

Figure 5-29
Flux avec commentaires ajoutés



Les autres peuvent ensuite afficher ces commentaires à l'écran et vous pouvez imprimer une image du flux qui inclut ces commentaires.

Vous pouvez lister tous les commentaires d'un flux ou d'un super noeud, modifier l'ordre des commentaires dans la liste, modifier le texte du commentaire et modifier la couleur du premier plan ou de l'arrière-plan d'un commentaire. [Pour plus d'informations, reportez-vous à la section Liste des commentaires sur un flux sur p. 76.](#)

Vous pouvez également ajouter des remarques sous la forme d'annotations de texte aux flux, noeuds et nuggets au moyen de l'onglet Annotations d'une boîte de dialogue Propriétés du flux, d'une boîte de dialogue Noeud ou d'une fenêtre de nugget de modèle. Ces remarques sont visibles uniquement lorsque l'onglet Annotations est ouvert, mais les annotations de flux peuvent être affichées sous forme de commentaires à l'écran. [Pour plus d'informations, reportez-vous à la section Annotations sur p. 78.](#)


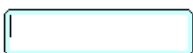


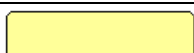
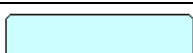
Commentaires

Les commentaires prennent la forme de zones de texte dans lesquelles vous pouvez saisir n'importe quelle taille de texte et ajouter tous les commentaires souhaités. Un commentaire peut être autonome (non attaché à un objet de flux) ou il peut être connecté à un ou plusieurs noeuds ou nuggets de modèle dans le flux. Les commentaires autonomes sont généralement utilisés pour décrire l'objectif général du flux ; les commentaires connectés décrivent le noeud ou le nugget auquel ils sont attachés. Les noeuds et les nuggets peuvent avoir plusieurs commentaires attachés, et le flux peut contenir n'importe quel nombre de commentaires autonomes.

Remarque : Vous pouvez également afficher les annotations de flux sous forme de commentaires à l'écran, bien que ceux-ci ne puissent pas être joints aux noeuds ou aux nuggets. [Pour plus d'informations, reportez-vous à la section Conversion d'annotations en commentaires sur p. 77.](#)

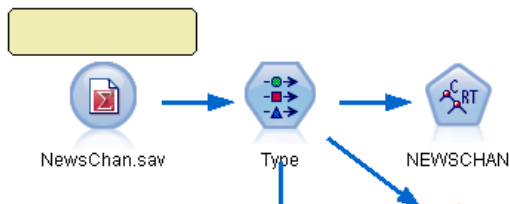
L'apparence de la zone de texte change pour indiquer le mode de commentaire actuel (ou l'annotation sous forme de commentaire), comme l'indique le tableau suivant :

Table 5-1
Modes des zones de texte de commentaires et d'annotations

Zone de texte des commentaires	Zone de texte des annotations	Mode	Indique	Obtenu par...
		Edition	Le commentaire est ouvert pour l'édition.	Création d'un nouveau commentaire ou annotation ou double-clic sur un commentaire ou une annotation existante.
		Dernière sélection	Il est possible de déplacer, redimensionner ou de supprimer un commentaire.	Cliquez sur l'arrière-plan du flux après l'édition ou cliquez une fois sur un commentaire ou une annotation existante.
		Affichage	l'édition est terminée.	Cliquez sur un autre noeud, commentaire ou annotation après l'édition.

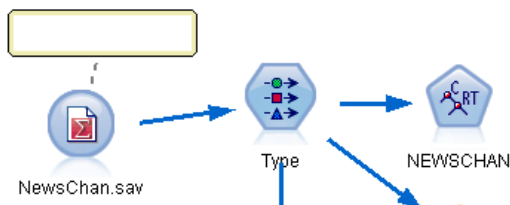
Lorsque vous créez un nouveau commentaire autonome, il apparaît initialement en haut à gauche de l'espace de travail du flux.

Figure 5-30
Nouveau commentaire autonome



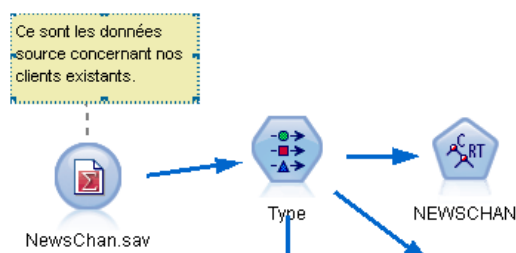
Si vous attachez un commentaire à un noeud ou un nugget, le commentaire est initialement affiché au-dessus de l'objet de flux auquel il est attaché.

Figure 5-31
Nouveau commentaire attaché au noeud



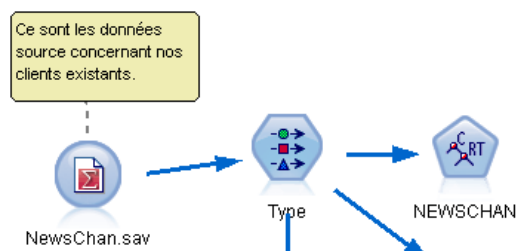
La zone de texte est en blanc pour indiquer que du texte peut être saisi. Une fois que vous avez saisi le texte, cliquez hors de la zone de texte. L'arrière-plan du commentaire devient jaune pour indiquer que la saisie de texte est terminée. Le commentaire reste sélectionné, vous permettant de le déplacer, de le redimensionner ou de le supprimer.

Figure 5-32
Commentaire en mode d'édition



Lorsque vous cliquez de nouveau, la bordure se transforme en lignes pleines pour indiquer que l'édition est terminée.

Figure 5-33
Commentaire terminé



En double-cliquant sur un commentaire, vous faites passer la zone de texte en mode d'édition : l'arrière-plan devient blanc et le texte du commentaire peut être édité.

Vous pouvez également attacher des commentaires aux super noeuds. [Pour plus d'informations, reportez-vous à la section Annotation et changement de nom des super noeuds dans le chapitre 9 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Opérations impliquant des commentaires

Vous pouvez exécuter plusieurs opérations sur les commentaires. Vous pouvez :

- ajouter un commentaire autonome
- attacher un commentaire à un noeud ou un nugget
- éditer un commentaire
- redimensionner un commentaire
- déplacer un commentaire
- déconnecter un commentaire
- supprimer un commentaire
- afficher ou masquer tous les commentaires sur un flux

Pour ajouter un commentaire autonome

- ▶ Assurez-vous que rien n'est sélectionné dans le flux.
- ▶ Effectuez l'une des actions suivantes :
 - Dans le menu principal, cliquez sur :
Insertion > Nouveau commentaire
 - Cliquez avec le bouton droit de la souris sur l'arrière-plan du flux et cliquez sur Nouveau commentaire dans le menu contextuel.
 - Cliquez sur le bouton de Nouveau commentaire de la barre d'outils.
- ▶ Saisissez le texte du commentaire (ou coller le texte dans le presse-papiers).
- ▶ Cliquez sur un noeud dans le flux pour enregistrer le commentaire.

Pour attacher un commentaire à un noeud ou un nugget

- ▶ Sélectionnez un ou plusieurs noeuds ou nuggets dans l'espace de travail de flux.
- ▶ Effectuez l'une des actions suivantes :
 - Dans le menu principal, cliquez sur :
Insertion > Nouveau commentaire
 - Cliquez avec le bouton droit de la souris sur l'arrière-plan du flux et cliquez sur Nouveau commentaire dans le menu contextuel.
 - Cliquez sur le bouton de Nouveau commentaire de la barre d'outils.
- ▶ Entrez le texte du commentaire.
- ▶ Cliquez sur un autre noeud dans le flux pour enregistrer le commentaire.
Vous pouvez également :
- ▶ Insérer un commentaire autonome (voir section précédente).
- ▶ Effectuez l'une des actions suivantes :
 - Sélectionnez le commentaire, appuyez sur F2, puis sélectionnez le noeud ou le nugget.
 - Sélectionnez le noeud ou le nugget, appuyez sur F2, puis sélectionnez le commentaire.
 - (Souris à trois boutons uniquement) Déplacez le pointeur de la souris sur le commentaire, maintenez enfoncé le bouton central, faites glisser le pointeur de la souris sur le noeud ou le nugget et relâchez le bouton de la souris.

Pour attacher un commentaire à un noeud ou un nugget supplémentaire

Si un commentaire est déjà attaché à un noeud ou un nugget, ou s'il se trouve actuellement au niveau du flux et que vous voulez l'attacher à un noeud ou un nugget supplémentaire, effectuez l'une des actions suivantes :

- Sélectionnez le commentaire, appuyez sur F2, puis sélectionnez le noeud ou le nugget.

- Sélectionnez le noeud ou le nugget, appuyez sur F2, puis sélectionnez le commentaire.
- (Souris à trois boutons uniquement) Déplacez le pointeur de la souris sur le commentaire, maintenez enfoncé le bouton central, faites glisser le pointeur de la souris sur le noeud ou le nugget et relâchez le bouton de la souris.

Pour modifier un commentaire existant

- ▶ Effectuez l'une des actions suivantes :
 - Double-cliquez sur la zone de texte du commentaire.
 - Sélectionnez la zone de texte et appuyez sur Entrée.
 - Faites un clic droit sur la zone de texte pour afficher son menu et choisissez Modifier.
- ▶ Modifiez le texte du commentaire. Vous pouvez utiliser les touches de raccourci standard de Windows lors de l'édition, par exemple, Ctrl+C pour copier du texte. Les autres options existantes pendant la modification sont répertoriées dans le menu contextuel du commentaire.
- ▶ Cliquez une fois hors de la zone de texte pour afficher les commandes de redimensionnement, puis à nouveau pour terminer le commentaire.

Pour redimensionner une zone de texte de commentaire

- ▶ Sélectionnez le commentaire pour afficher les commandes de redimensionnement.
- ▶ Cliquez sur une commande et faites-la glisser pour redimensionner la zone.
- ▶ Cliquez hors de la zone de texte pour enregistrer la modification.

Pour déplacer un commentaire existant

Si vous souhaitez déplacer un commentaire mais pas ses objets attachés (le cas échéant), effectuez l'une des actions suivantes :

- Déplacez le pointeur de la souris sur le commentaire, maintenez enfoncé le bouton gauche de la souris et faites glisser le commentaire vers sa nouvelle position.
- Sélectionnez le commentaire, maintenez enfoncée la touche Alt et déplacez le commentaire à l'aide des flèches.

Si vous souhaitez déplacer un commentaire avec des noeuds ou des nuggets auxquels le commentaire est attaché :

- ▶ Sélectionnez tous les objets à déplacer.
- ▶ Effectuez l'une des actions suivantes :
 - Déplacez le pointeur de la souris sur l'un des objets, maintenez enfoncé le bouton gauche de la souris, et faites glisser les objets vers leur nouvelle position.
 - Sélectionnez un des objets, maintenez enfoncée la touche Alt, et déplacez les objets à l'aide des flèches.

Pour déconnecter un commentaire d'un noeud ou d'un nugget

- ▶ Sélectionnez au moins un commentaire à déconnecter.
- ▶ Effectuez l'une des actions suivantes :
 - Appuyez sur la touche F3.
 - Faites un clic droit sur un commentaire sélectionné et cliquez sur Déconnecter sur son menu.

Pour supprimer un commentaire

- ▶ Sélectionnez au moins un commentaire à supprimer.
- ▶ Effectuez l'une des actions suivantes :
 - Appuyez sur la touche Supprimer.
 - Faites un clic droit sur un commentaire sélectionné et cliquez sur Supprimer sur son menu.

Si le commentaire était attaché à un noeud ou un nugget, la ligne de liaison est supprimée également.

Si le commentaire était à l'origine un flux ou une annotation de super noeud ayant été converti en commentaire indépendant, ce commentaire est supprimé de l'espace de travail mais son texte est conservé dans l'onglet Annotations du flux ou du super noeud.

Pour afficher ou masquer des commentaires sur un flux

- ▶ Effectuez l'une des actions suivantes :
 - Dans le menu principal, cliquez sur :
Affichage > Commentaires
 - Cliquez sur le bouton Afficher/Masquer les commentaires de la barre d'outils.

Liste des commentaires sur un flux

Vous pouvez afficher une liste de tous les commentaires qui ont été faits pour un flux ou un super noeud particulier.

Dans la liste, vous pouvez

- modifier l'ordre des commentaires
- Modifier le texte du commentaire
- Modifier la couleur du premier plan ou de l'arrière-plan d'un commentaire

Liste de commentaires

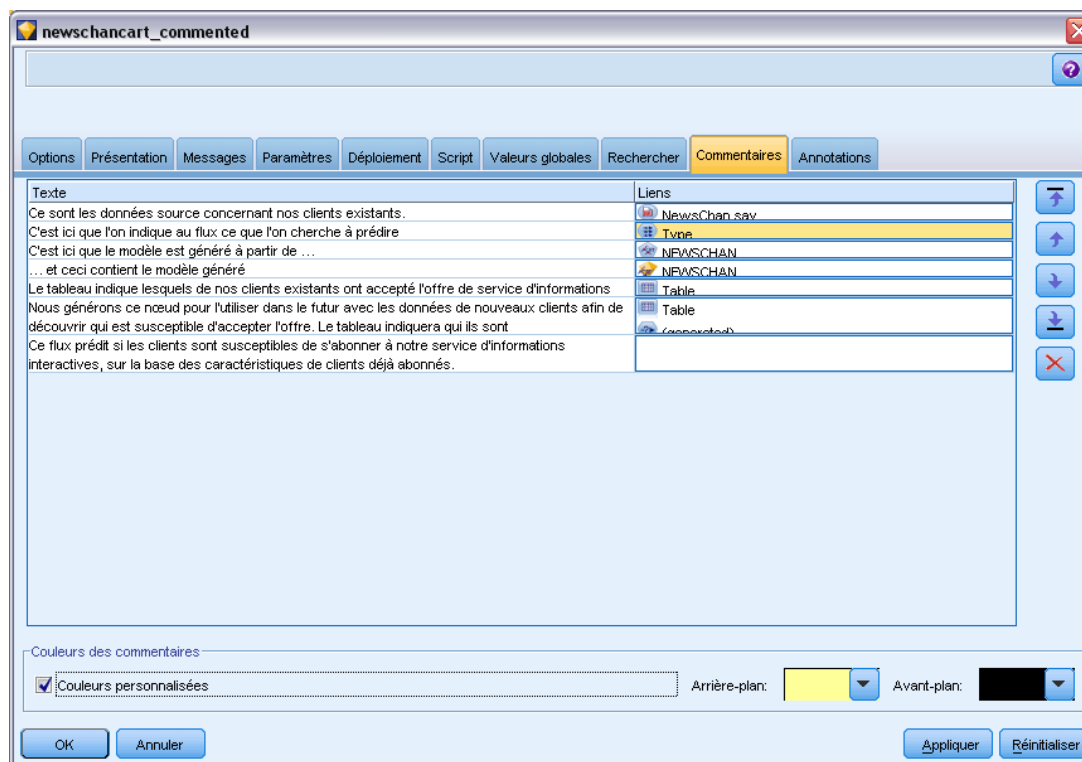
Pour lister des commentaires effectués sur un flux, procédez de l'une des manières suivantes :

- Dans le menu principal, cliquez sur :
Outils > Propriétés du flux > Commentaires

- Cliquez avec le bouton droit de la souris dans le panneau des gestionnaires et cliquez sur Propriétés du flux, puis sur Commentaires.
- Cliquez avec le bouton droit de la souris sur l'arrière-plan d'un flux de l'espace de travail et cliquez sur Propriétés du flux, puis sur Commentaires.

Figure 5-34

Liste des commentaires sur un flux



Texte : Le texte du commentaire. Double-cliquez sur le texte pour modifier le champ en une zone de texte modifiable.

Liens. Le nom du noeud auquel le commentaire est attaché. Si ce champ est vide, le commentaire s'applique au flux.

Boutons de positionnement. Ils permettent de déplacer un commentaire sélectionné vers le haut ou vers le bas dans la liste.

Couleurs de commentaire. Pour modifier la couleur de premier plan ou d'arrière-plan d'un commentaire, sélectionnez le commentaire, cochez la case Couleurs personnalisées, puis sélectionnez une couleur dans la liste Arrière-plan ou Premier plan (ou les deux). Cliquez sur Appliquer, puis cliquez sur l'arrière-plan du flux pour visualiser l'effet de la modification. Cliquez sur OK pour enregistrer cette modification.

Conversion d'annotations en commentaires

Il est possible de convertir des annotations faites sur des flux ou des super noeuds en commentaires.

Dans le cas de flux, l'annotation est convertie en un commentaire autonome (c'est-à-dire qu'il n'est joint à aucun noeud) dans l'espace de travail du flux.

Lorsqu'une annotation de super noeud est convertie en un commentaire, celui-ci n'est pas joint au super noeud dans l'espace de travail du flux, mais il est visible lorsque vous effectuez un zoom avant dans le super noeud.

Pour convertir une annotation de flux en un commentaire

- ▶ Cliquez sur Propriétés du flux dans le menu Outils. (Vous pouvez également cliquer avec le bouton droit de la souris sur un flux dans le panneau des gestionnaires et cliquer sur Propriétés du flux.)
- ▶ Cliquez sur l'onglet Annotations.
- ▶ Cochez la case Afficher une annotation comme un commentaire.
- ▶ Cliquez sur OK.

Pour convertir une annotation de super noeud en un commentaire

- ▶ Double-cliquez sur l'icône du super noeud dans l'espace de travail.
- ▶ Cliquez sur l'onglet Annotations.
- ▶ Cochez la case Afficher une annotation comme un commentaire.
- ▶ Cliquez sur OK.

Annotations

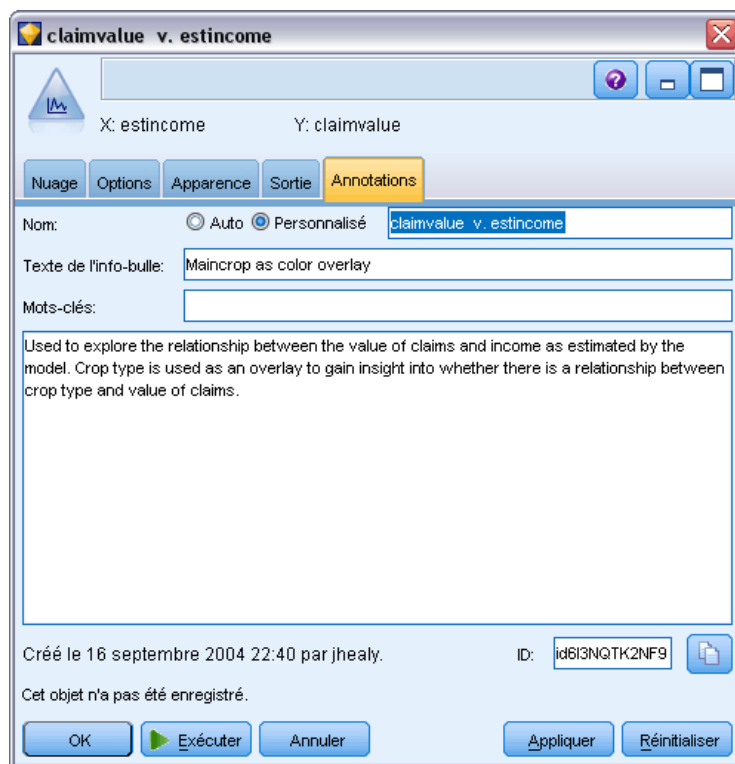
Il est possible d'annoter des noeuds, des flux et des modèles de différentes manières. Vous pouvez ajouter des annotations descriptives et spécifier un nom personnalisé. Ces options sont utiles plus particulièrement lorsque vous générez des rapports sur les flux ajoutés au panneau de projet. Pour les noeuds et les nuggets de modèle, vous pouvez également ajouter une info-bulle afin de pouvoir faire la différence entre les noeuds similaires dans l'espace de travail de flux.

Ajout d'annotations

Lorsque vous éditez un noeud ou un nugget de modèle, une boîte de dialogue à onglets apparaît ; elle contient l'onglet Annotations qui permet de définir différentes options d'annotation. Vous pouvez également ouvrir directement l'onglet Annotations.

- ▶ Pour annoter un noeud ou un nugget, cliquez dessus avec le bouton droit de la souris dans l'espace de travail de flux, puis cliquez sur Renommer et annoter. La boîte de dialogue d'édition s'ouvre et l'onglet Annotations s'affiche.
- ▶ Pour annoter un flux, cliquez sur Propriétés du flux dans le menu Outils. (Vous pouvez également cliquer avec le bouton droit de la souris sur un flux dans le panneau des gestionnaires et cliquer sur Propriétés du flux.) Cliquez sur l'onglet Annotations.

Figure 5-35
Options de l'onglet Annotations



Nom. Sélectionnez **Personnalisé** pour modifier le nom généré automatiquement ou pour créer un nom unique pour le noeud selon son affichage dans l'espace de travail.

Texte de l'info-bulle. (Uniquement pour les noeuds et les nuggets de modèle) Saisissez le texte pour l'info-bulle dans l'espace de travail du flux. Cette option s'avère particulièrement utile lorsque vous travaillez avec plusieurs noeuds similaires.

Mots-clés. Permet d'indiquer les mots-clés à utiliser dans les rapports de projet, ou lors de la recherche de noeuds dans un flux ou du suivi des objets stockés dans le référentiel (reportez-vous à A propos du IBM SPSS Collaboration and Deployment Services Repository sur p. 159). Vous pouvez séparer les mots-clés par des points-virgules, comme l'illustre l'exemple suivant : `revenu; type de culture; valeur de la demande`. Les espaces utilisés au début et à la fin de chaque mot-clé sont supprimés (par exemple, `revenu ; type de culture` génère les mêmes résultats que `revenu;type de culture`). (Toutefois, les espaces utilisés à l'intérieur des mots-clés ne sont pas supprimés. Par exemple, l'occurrence `type de culture` dotée d'un espace et l'occurrence `type de culture` incluant deux espaces sont différentes.)

La zone de texte principale peut être utilisée pour entrer des annotations longues concernant les opérations ou les décisions relatives au noeud. Par exemple, lorsque vous partagez et réutilisez des flux, il est utile de garder une trace des décisions, telles que l'utilisation d'un noeud **Filtrer** pour supprimer un champ comportant de nombreux blancs. Les informations saisies lors de l'annotation sont stockées avec le noeud. Vous pouvez également inclure ces annotations dans un rapport de projet créé à partir du panneau de projet. [Pour plus d'informations, reportez-vous à la section Introduction aux projets dans le chapitre 11 sur p. 214.](#)

Afficher une annotation comme un commentaire (Uniquement pour des annotations sur un flux ou un super noeud) Cochez cette case pour convertir l'annotation en un commentaire autonome qui sera visible sur l'espace de travail du flux. [Pour plus d'informations, reportez-vous à la section Ajout de commentaires et d'annotations à des noeuds et à des flux sur p. 70.](#)

ID. Affiche un ID unique qui peut être utilisé pour référencer le noeud à des fins de génération de scripts ou d'automatisation. Cette valeur est générée automatiquement lorsque le noeud est créé et elle ne sera pas modifiée. Notez également que, pour éviter toute confusion avec la lettre « O », les zéros ne sont pas utilisés dans les ID de noeud. Utilisez le bouton Copier, situé à droite, pour copier et coller l'ID dans des scripts ou ailleurs, selon les besoins. [Pour plus d'informations, reportez-vous à la section Référence aux noeuds dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Enregistrement des flux de données

Après avoir créé un flux, vous pouvez l'enregistrer pour un usage ultérieur.

Pour enregistrer un flux

- ▶ Dans le menu Fichier, cliquez sur Enregistrer le flux ou sur Enregistrer le flux sous.
- ▶ Dans la boîte de dialogue Enregistrer, sélectionnez le répertoire dans lequel enregistrer le fichier de flux.
- ▶ Entrez le nom du flux dans la zone de texte Fichier.
- ▶ Sélectionnez Ajouter au projet pour ajouter le flux enregistré au projet en cours.

Cliquez sur Enregistrer pour stocker le flux avec l'extension **.str* dans le répertoire spécifié.

Fichiers de sauvegarde automatique. Lorsqu'un flux est enregistré, la version précédemment enregistrée du fichier est conservée automatiquement en tant que copie de sauvegarde. Un tiret est ajouté au nom du fichier (par exemple, *monflux.str-*). Pour restaurer cette version sauvegardée, supprimez simplement le tiret et rouvrez le fichier.

Enregistrement des états

Outre les flux, vous pouvez enregistrer les **états** qui incluent le diagramme de flux affiché et tout nugget de modèle que vous avez créé (répertorié dans l'onglet Modèles du panneau des gestionnaires).

Pour enregistrer un état

- ▶ Dans le menu Fichier, cliquez sur :
Etat > Enregistrer l'état ou Enregistrer l'état sous
- ▶ Dans la boîte de dialogue Enregistrer, sélectionnez le répertoire dans lequel enregistrer le fichier d'état.

Cliquez sur Enregistrer pour stocker l'état avec l'extension **.cst* dans le répertoire spécifié.

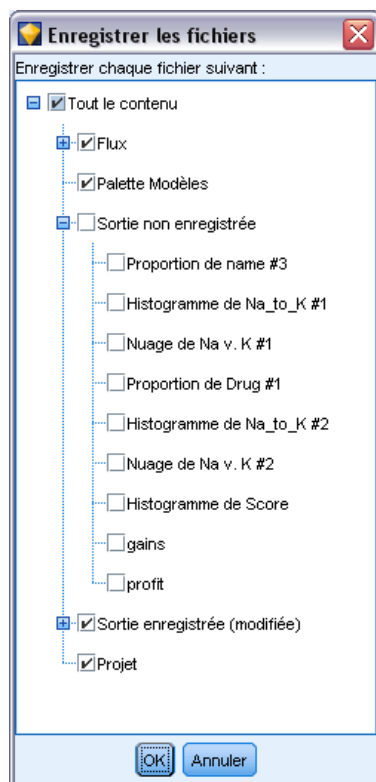
Enregistrement de noeuds

Vous pouvez également enregistrer un noeud précis en cliquant dessus avec le bouton droit de la souris dans l'espace de travail de flux et en cliquant sur Enregistrer le noeud dans le menu contextuel. Utilisez l'extension de fichier *.nod.

Enregistrement de plusieurs objets de flux

Si vous fermez IBM® SPSS® Modeler alors que plusieurs objets n'ont pas été enregistrés (comme des flux, des projets ou des nuggets de modèles), le système vous invite à les enregistrer avant que vous quittiez définitivement le logiciel. Si vous choisissez d'enregistrer les éléments, une boîte de dialogue apparaît. Elle contient des options permettant d'enregistrer chaque objet.

Figure 5-36
Enregistrement de plusieurs objets



- ▶ Cochez les cases correspondant aux objets à enregistrer.
- ▶ Cliquez sur OK pour enregistrer chaque objet à l'emplacement requis.

Une boîte de dialogue Enregistrer standard apparaît alors pour chaque objet. Une fois l'enregistrement terminé, vous quittez automatiquement l'application.

Enregistrement des résultats

Les tableaux, graphiques et rapports générés depuis les noeuds de sortie IBM® SPSS® Modeler peuvent être enregistrés au format d'objet de sortie (*.cou).

- ▶ Lorsque vous affichez la sortie à enregistrer, dans les menus de la fenêtre de sortie, cliquez sur : Fichier > Enregistrer
- ▶ Indiquez les nom et emplacement du fichier de sortie.
- ▶ Si vous le souhaitez, sélectionnez Ajouter le fichier au projet dans la boîte de dialogue d'enregistrement pour inclure le fichier dans le projet actuel. [Pour plus d'informations, reportez-vous à la section Introduction aux projets dans le chapitre 11 sur p. 214.](#)

Vous pouvez également cliquer avec le bouton droit de la souris sur un objet de sortie répertorié dans le panneau des gestionnaires et sélectionner Enregistrer dans le menu contextuel.

Codage et décodage des informations

Lorsque vous enregistrez un flux, un noeud, un projet, un fichier de sortie ou un nugget de modèle, vous pouvez le coder afin d'empêcher toute personne non autorisée de l'utiliser. Pour ce faire, sélectionnez une option supplémentaire à l'enregistrement et ajoutez un mot de passe à l'élément enregistré. Ce codage peut être paramétré pour n'importe quel élément que vous enregistrez et lui ajoute une sécurité supplémentaire ; ce type de codage est différent du codage SSL utilisé si vous transmettez des fichiers entre IBM® SPSS® Modeler et IBM® SPSS® Modeler Server.

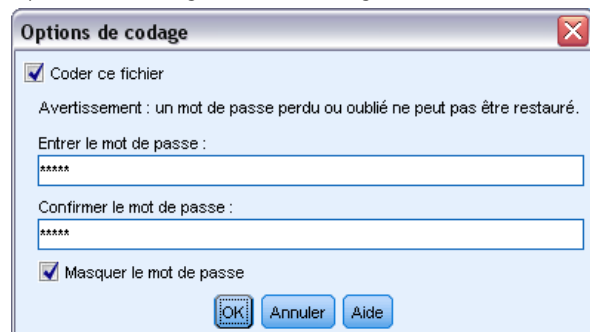
Lorsque vous essayez d'ouvrir un élément codé, une invite vous demande d'entrer le mot de passe. Une fois que vous avez saisi le mot de passe approprié, l'élément est automatiquement décodé et s'ouvre comme d'habitude.

Pour coder un élément

- ▶ Dans la boîte de dialogue Enregistrer de l'élément à coder, cliquez sur Options. La boîte de dialogue Options de codage apparaît.

Figure 5-37

Options de codage lors de l'enregistrement d'un fichier



- ▶ Sélectionnez Coder ce fichier.
- ▶ Pour plus de sécurité, vous pouvez également sélectionner Masquer le mot de passe. Tout ce que vous saisissez apparaît alors sous la forme d'une série de points.

- ▶ Entrez le mot de passe. *Avertissement* : Si vous oubliez le mot de passe, le fichier ou le modèle ne peut pas être ouvert.
- ▶ Si vous avez sélectionné Masquer le mot de passe, saisissez de nouveau le mot de passe pour confirmer que vous l'avez correctement entré.
- ▶ Cliquez sur OK pour revenir à la boîte de dialogue Enregistrer.

Remarque : Si vous enregistrez une copie d'un élément protégé par un code, le nouvel élément est automatiquement enregistré dans un format codé en utilisant le mot de passe d'origine, sauf si vous modifiez les paramètres de la boîte de dialogue Options de codage.

Chargement de fichiers

Vous pouvez recharger un certain nombre d'objets enregistrés dans IBM® SPSS® Modeler :

- Flux (.str)
- Etats (.cst)
- Modèles (.gm)
- Palette de modèles (.gen)
- Noeuds (.nod)
- Sortie (.cou)
- Projets (.cpj)

Ouverture de nouveaux fichiers

Il est possible de charger les flux directement à partir du menu Fichier.

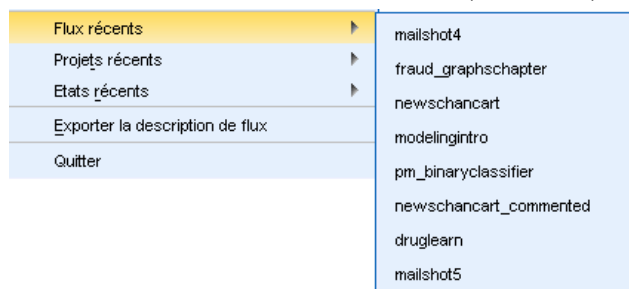
- ▶ Dans le menu Fichier, cliquez sur Ouvrir un flux.

Tous les autres types de fichier peuvent être ouverts via les sous-menus du menu Fichier. Par exemple, pour charger un modèle, dans le menu Fichier, cliquez sur :
Modèles > Ouvrir un modèle ou Charger la palette Modèles

Ouverture de fichiers récemment utilisés

Pour charger rapidement les derniers fichiers utilisés, utilisez les options figurant en bas du menu Fichier.

Figure 5-38
Ouverture de fichiers récemment utilisés à partir des options du menu Fichier



Sélectionnez Flux récents, Projets récents ou Etats récents pour développer la liste des derniers fichiers utilisés.

Mappage de flux de données

L'outil de mappage permet de connecter une nouvelle source de données à un flux existant. Cet outil configure la connexion et vous aide à définir la manière dont les noms des champs de la nouvelle source remplaceront ceux du flux existant. Au lieu de recréer intégralement un flux de données pour une nouvelle source de données, vous pouvez simplement établir une connexion avec un flux existant.

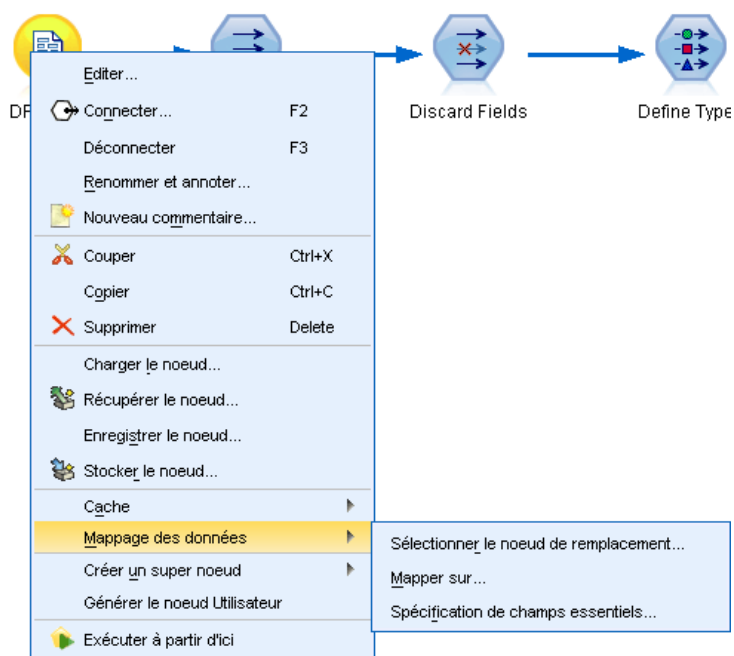
L'outil de mappage des données vous permet de relier deux fragments de flux et de vous assurer que le nom de tous les champs (essentiels) correspondent parfaitement. En résumé, le mappage des données équivaut simplement à créer un noeud Filtrer qui fait correspondre les champs appropriés en leur donnant un nouveau nom.

Il existe deux méthodes équivalentes pour mapper les données :

Sélectionner le noeud de remplacement. Cette méthode débute par le noeud à remplacer. D'abord, faites un clic droit sur le noeud à remplacer ; puis, avec l'option Mappage de données > Sélectionner le noeud de remplacement dans le menu contextuel, sélectionnez le noeud avec lequel le remplacer.

Mapper sur. Cette méthode débute par le noeud à introduire dans le flux. D'abord, faites un clic droit sur le noeud à introduire ; puis, avec l'option Mappage de données > Mapper sur dans le menu contextuel, sélectionnez le noeud auquel le joindre. Cette méthode est particulièrement utile pour le mappage sur un noeud terminal. *Remarque* : il est impossible d'effectuer des mappages sur des noeuds Fusionner ou Ajouter. En revanche, vous pouvez utiliser la méthode de connexion standard pour relier le flux au noeud Fusionner.

Figure 5-39
Sélection des options de mappage des données



Le mappage des données est étroitement lié à la création de flux. Si vous essayez de vous connecter à un noeud qui comporte déjà une connexion, vous aurez la possibilité de remplacer la connexion existante ou d'effectuer le mappage sur ce noeud.

Mappage des données sur un modèle

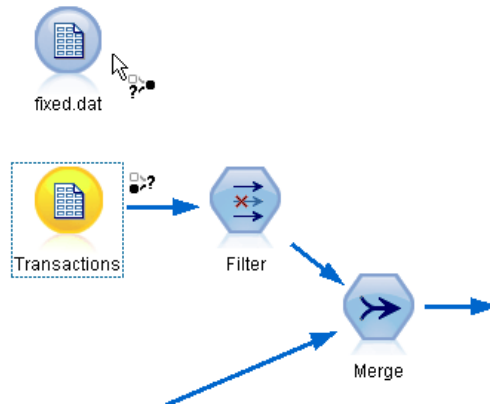
Pour remplacer la source de données d'un modèle de flux par un nouveau noeud source en apportant vos propres données dans IBM® SPSS® Modeler, vous devez utiliser l'option Sélectionner le noeud de remplacement du menu contextuel Mappage des données. Cette option est disponible pour tous les noeuds à l'exception de Fusionner, Agréger et de tous les noeuds terminaux. L'utilisation de l'outil de mappage de données pour effectuer cette action permet de vous assurer que les champs correspondent parfaitement entre les opérations du flux existantes et la nouvelle source de données. La procédure suivante vous fournira une présentation du processus de mappage de données.

Etape 1 : Spécification des champs essentiels dans le noeud source d'origine. Pour que les opérations de flux soient correctement exécutées, vous devez spécifier des champs essentiels. [Pour plus d'informations, reportez-vous à la section Spécification de champs essentiels sur p. 87.](#)

Etape 2 : Ajout d'une nouvelle source de données à l'espace de travail de flux. En vous servant de l'un des noeuds source, introduisez les nouvelles données de remplacement.

Etape 3 : Remplacement du noeud source du modèle. A l'aide de l'option Mappage des données du menu contextuel du noeud source du modèle, cliquez sur Sélectionner le noeud de remplacement, puis sélectionnez le noeud source pour les données de remplacement.

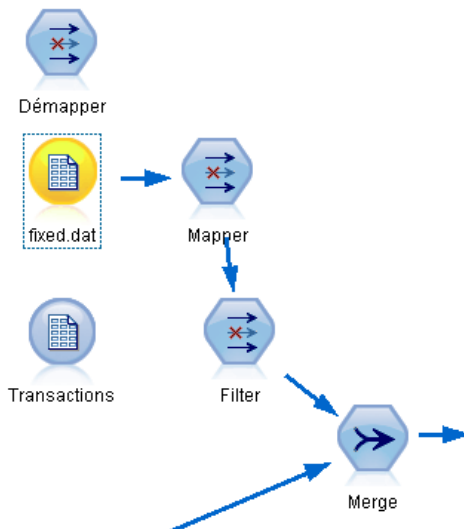
Figure 5-40
Sélection d'un noeud source de remplacement



Etape 4 : Vérification des champs mappés. Dans la boîte de dialogue qui apparaît, vérifiez que le logiciel mappe correctement les champs de la source de données de remplacement sur le flux. Tous les champs essentiels non mappés sont affichés en rouge. Ces champs sont utilisés dans les opérations de flux et doivent être remplacés par un champ similaire de la nouvelle source de données afin que les opérations en aval fonctionnent correctement. [Pour plus d'informations, reportez-vous à la section Vérification des champs mappés sur p. 88.](#)

Une fois que vous avez utilisé la boîte de dialogue pour vous assurer que tous les champs essentiels étaient correctement mappés, l'ancienne source de données est déconnectée. La nouvelle source de données est alors connectée au flux grâce à un noeud Filtrer appelé *Mapper*. Ce noeud Filtrer dirige le mappage réel des champs du flux. Un noeud Filtrer appelé *Démapper* est également inclus dans l'espace de travail de flux. Le noeud Filtrer appelé *Démapper* peut être utilisé pour annuler le mappage des noms de champs si vous l'ajoutez au flux. Le mappage sera annulé, mais vous devrez éditer les noeuds terminaux en aval pour resélectionner les champs et les superpositions.

Figure 5-41
Nouvelle source de données mappée avec succès sur le flux

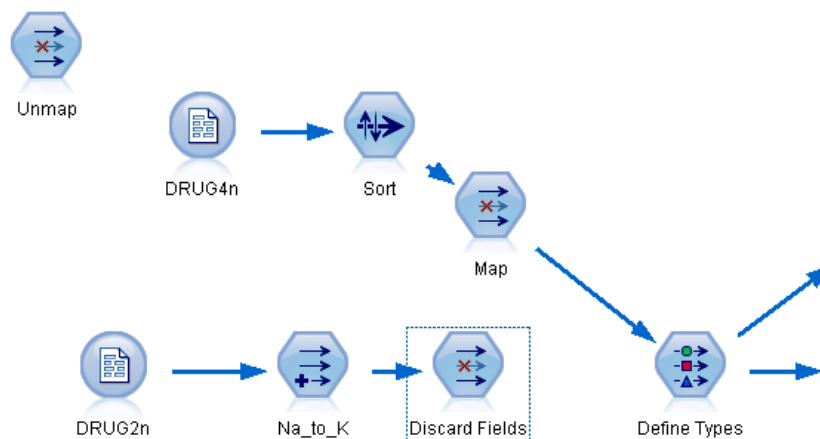


Mappage entre flux

Comme pour la connexion de noeuds, cette méthode de mappage des données n'exige pas que vous définissiez des champs essentiels au préalable. Avec cette méthode, vous pouvez simplement relier un flux à l'autre via l'option Mapper sur du menu contextuel Mappage des données. Ce type de mappage de données est utile pour le mappage à des noeuds terminaux, ainsi que pour la copie et le collage entre flux. *Remarque* : lorsque vous utilisez l'option Mapper sur, vous ne pouvez pas effectuer de mappage vers les noeuds Fusionner et Ajouter, et tous les types de noeud source.

Figure 5-42

Mappage d'un flux depuis son noeud Trier vers le noeud Typer d'un autre flux



Pour mapper des données entre des flux

- ▶ Cliquez avec le bouton droit de la souris sur le noeud que vous souhaitez utiliser pour la connexion au nouveau flux.
- ▶ Dans le menu, cliquez sur :
Mappage des données > Mapper sur
- ▶ Utilisez le curseur pour sélectionner un noeud de destination dans le flux cible.
- ▶ Dans la boîte de dialogue qui apparaît, assurez-vous que les champs correspondent parfaitement, puis cliquez sur OK.

Spécification de champs essentiels

Lorsque vous effectuez un mappage sur un flux existant, les champs essentiels sont spécifiés par l'auteur du flux. Ces champs essentiels indiquent si un champ particulier est utilisé dans les opérations en aval. Par exemple, le flux existant peut créer un modèle qui utilise un champ appelé *Attrition*. Dans ce flux, *Attrition* est un champ essentiel requis pour la création du modèle. De même, les champs utilisés dans les noeuds de manipulation, tels que les noeuds Calculer, sont requis pour le calcul du nouveau champ. En définissant explicitement ces champs comme obligatoires, vous avez la certitude qu'ils seront mappés sur les champs appropriés contenus dans la nouvelle source de données. Si des champs obligatoires ne sont pas mappés, un message d'erreur

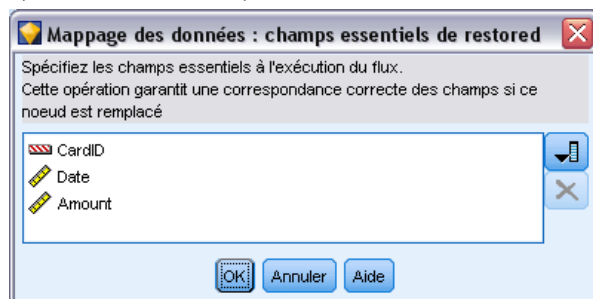
est envoyé. Si certains noeuds de manipulation ou de sortie vous semblent inutiles, vous pouvez les supprimer du flux et retirer les champs correspondants de la liste des champs obligatoires.

Pour définir des champs essentiels

- ▶ Cliquez avec le bouton droit de la souris sur le noeud source du modèle de flux à remplacer.
- ▶ Dans le menu, cliquez sur :
Mappage des données > Spécifier des champs essentiels

Figure 5-43

Spécification de champs essentiels

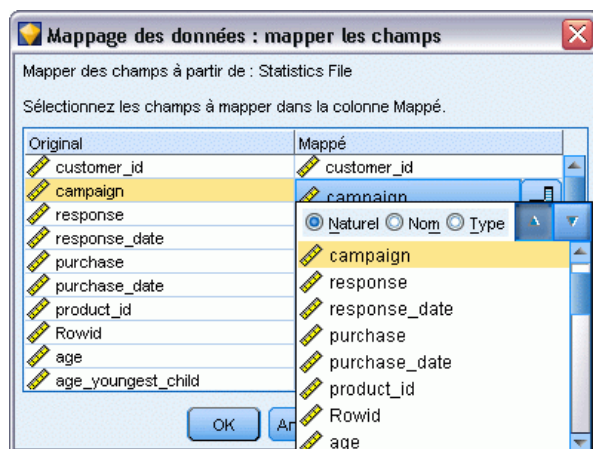


- ▶ A l'aide du sélecteur de champs, vous pouvez ajouter des champs à la liste ou en supprimer. Pour ouvrir le sélecteur de champs, cliquez sur l'icône située à droite de la liste des champs.

Vérification des champs mappés

Une fois que vous avez choisi le point à partir duquel un flux (ou une source) de données sera mappé sur un autre flux (une autre source), une boîte de dialogue apparaît. Elle vous permet de sélectionner les champs à mapper ou de vous assurer que le mappage par défaut du système est correct. Si des champs essentiels ont été définis pour le flux ou la source de données et qu'ils ne présentent aucune correspondance, ils sont affichés en rouge. Tous les champs non mappés de la source de données seront inclus tels quels dans le noeud Filtrer, mais vous pouvez également mapper des champs non essentiels.

Figure 5-44
Sélection des champs pour le mappage



Original. Répertorie tous les champs du modèle ou du flux existant (tous les champs présents plus en aval). Les champs provenant de la nouvelle source de données seront mappés sur les champs existants.

Mappé. Affiche la liste des champs sélectionnés à mapper sur les champs du modèle. Il s'agit des champs dont les noms doivent être modifiés pour correspondre aux champs d'origine utilisés dans les opérations de flux. Cliquez sur la cellule du tableau correspondant à un champ pour activer la liste des champs disponibles.

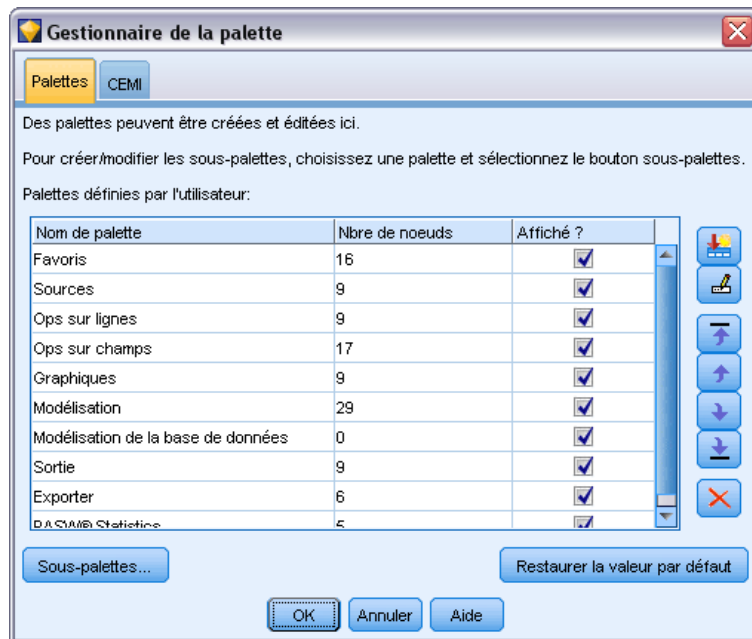
Avant toute opération de mappage, il est conseillé d'étudier la source de données avec attention, afin d'identifier précisément les champs à mapper. Par exemple, vous pouvez utiliser l'onglet Types du noeud source pour consulter le récapitulatif des données source.

Astuces et raccourcis

Pour une utilisation rapide et simple, familiarisez-vous avec les astuces et raccourcis suivants :

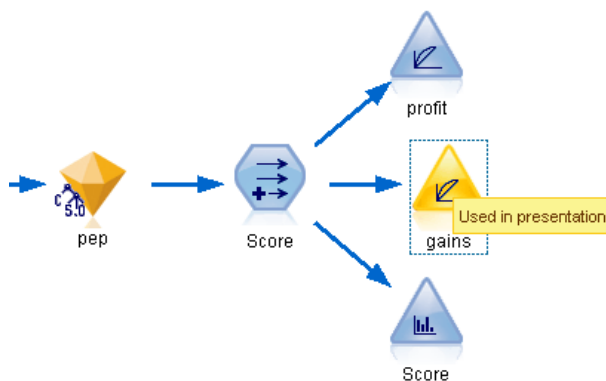
- **Créez des flux rapidement à l'aide d'un double-clic.** Double-cliquez simplement sur un noeud de la palette pour l'ajouter et le connecter au flux en cours.
- **Utilisez des combinaisons de touches pour sélectionner des noeuds en aval.** Appuyez sur Ctrl+Q et sur Ctrl+W pour activer ou désactiver la sélection de tous les noeuds en aval.
- **Utilisez des touches de raccourci pour connecter et déconnecter les noeuds.** Lorsqu'un noeud est sélectionné dans l'espace de travail, appuyez sur la touche F2 pour activer une connexion, sur la touche Tab pour accéder au noeud souhaité et sur la touche Maj+Espace pour mettre un terme à la connexion. Appuyez sur la touche F3 pour déconnecter toutes les entrées et sorties du noeud sélectionné.
- **Personnalisez l'onglet Palette de noeuds avec vos noeuds favoris.** Dans le menu Outils, cliquez sur Gérer les palettes pour ouvrir une boîte de dialogue vous permettant d'ajouter, de supprimer ou de déplacer des noeuds affichés dans la Palette de noeuds.

Figure 5-45
Gestionnaire de la palette



- **Renommez les noeuds et ajoutez des info-bulles.** Chaque boîte de dialogue de noeud contient un onglet Annotations. Il vous permet de spécifier un nom personnalisé pour chaque noeud de l'espace de travail et d'ajouter des info-bulles pour l'organisation de votre flux. Vous pouvez également inclure des annotations longues pour suivre la progression, enregistrer les détails du processus, et indiquer toute décision commerciale prise ou à prendre.

Figure 5-46
Info-bulle et nom de noeud personnalisé



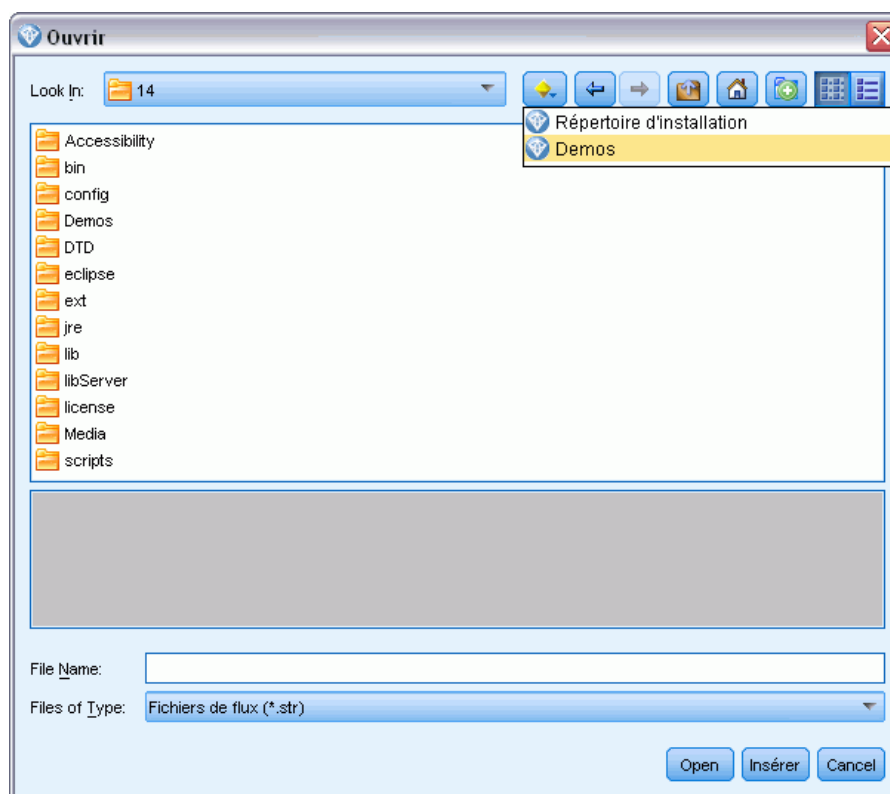
- **Insérez des valeurs automatiquement dans une expression CLEM.** Le Générateur de formules, accessible à partir d'un grand nombre de boîtes de dialogue (telles que celles des noeuds Calculer et Remplacer), vous permet d'insérer automatiquement des valeurs de champ dans une expression CLEM. Dans le Générateur de formules, cliquez sur le bouton des valeurs pour sélectionner les valeurs de champ.

Figure 5-47
Bouton Valeurs



- **Accès rapide aux fichiers.** Lorsque vous recherchez des fichiers dans une boîte de dialogue Ouvrir, utilisez la liste Fichier (bouton représentant un diamant jaune) pour accéder aux répertoires précédemment utilisés et aux répertoires IBM® SPSS® Modeler par défaut. Utilisez les boutons Suivant et Précédent pour parcourir les répertoires utilisés.

Figure 5-48
Sélection du dossier Demos dans la liste des répertoires récemment consultés



- **Réduisez l'encombrement de la fenêtre de sortie.** Vous pouvez fermer et supprimer rapidement une sortie à l'aide du bouton X rouge situé dans le coin supérieur droit de toutes les fenêtres de sortie. Vous pouvez ainsi conserver uniquement les résultats prometteurs ou intéressants dans l'onglet Sorties du panneau des gestionnaires.

Un ensemble complet de raccourcis clavier est disponible pour le logiciel. [Pour plus d'informations, reportez-vous à la section Accessibilité via le clavier dans l'annexe A sur p. 257.](#)

Savez-vous que vous pouvez...

- faire glisser et sélectionner un groupe de noeuds sur l'espace de travail à l'aide de la souris.
- copier et coller des noeuds d'un flux à l'autre.

- accéder à l'aide à partir de chaque boîte de dialogue et fenêtre de sortie.
- obtenir de l'aide sur le processus CRISP-DM (Cross-Industry Standard Process for Data Mining). (Dans le menu Aide, cliquez sur Aide CRISP-DM.)

Traitement des valeurs manquantes

Description des valeurs manquantes

Au cours de l'étape de préparation des données de Data mining, vous aurez souvent à remplacer des valeurs manquantes. Les **valeurs manquantes** sont des valeurs de l'ensemble de données qui sont inconnues, non recueillies ou entrées de façon incorrecte. En général, ces valeurs ne sont pas valides pour leur champ. Par exemple, le champ *Sexe* doit contenir les valeurs *M* et *F*. Si les valeurs *Y* ou *Z* apparaissent dans ce champ, il y a de fortes chances qu'elles ne soient pas valides et, par conséquent, elles devraient être considérées comme des blancs. De même, une valeur négative pour le champ *Age* ne signifie rien et devrait également être considérée comme blanc. Il arrive fréquemment que de telles valeurs soient entrées volontairement ou ne soient pas renseignées au cours d'un questionnaire pour indiquer une absence de réponse. Il est recommandé d'apporter une attention particulière à ces blancs pour déterminer si une absence de réponse, telle que le refus d'indiquer son âge, est un facteur dans la prévision d'un résultat spécifique.

Certaines méthodes de modélisation traitent mieux les données manquantes que d'autres. Par exemple, C5.0 et Apriori parviennent à analyser les valeurs qui sont explicitement déclarées « manquantes » dans un noeud Typer. D'autres méthodes de modélisation rencontrent des problèmes de traitement des données manquantes, entraînant ainsi des durées d'apprentissage prolongées et des modèles moins précis.

Plusieurs types de valeur manquante sont reconnus par IBM® SPSS® Modeler :

- **Valeurs nulles ou manquantes système.** Ces valeurs sont des valeurs « non-chaîne » qui ne sont pas renseignées dans la base de données ou dans le fichier source, et qui n'ont pas été spécifiquement définies comme « manquantes » dans un noeud source ou un noeud Typer. Les valeurs manquantes système sont affichées sous la forme \$null\$. Les chaînes vides ne sont pas considérées comme des valeurs nulles dans SPSS Modeler, même si elles peuvent être traitées comme telles par certaines bases de données.
- **Chaînes vides et espaces blancs.** Les chaînes vides et les espaces blancs (chaînes sans caractère visible) sont traités différemment des valeurs nulles. Dans la plupart des cas, les chaînes vides sont considérées comme des espaces blancs. Par exemple, si vous choisissez de traiter les espaces blancs comme blancs dans un noeud source ou un noeud Typer, ce paramètre s'applique également aux chaînes vides.
- **Valeurs manquantes définies par l'utilisateur ou vides.** Ces valeurs sont des valeurs, telles que inconnu, 99, ou -1, qui sont explicitement définies comme manquantes dans un noeud source ou Typer. Vous pouvez également, si vous le souhaitez, préciser si les valeurs nulles et les espaces blancs doivent être traités comme des blancs ; un traitement spécial leur est alors appliqué et ils sont exclus de la plupart des calculs. Par exemple, vous pouvez utiliser la fonction @BLANK pour traiter comme des blancs ces valeurs, ainsi que d'autres types de valeur manquante. [Pour plus d'informations, reportez-vous à la section Utilisation de la boîte de dialogue Valeurs dans le chapitre 4 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Figure 6-1
Spécification des valeurs manquantes pour une variable continue

Lecture de données mixtes. Au cours de la lecture des champs de stockage numérique (entier, nombre réel, heure, horodatage ou date), toutes les valeurs non numériques sont définies comme étant *nulles* ou *manquantes dans le système*. En effet, contrairement à certaines applications, les types de stockage mixtes au sein d'un champne sont pas autorisés. Pour éviter ce type de problème, faites en sorte que les champs comportant des données mixtes soient lus en tant que chaînes ; pour cela, modifiez le type de stockage dans le noeud source ou dans l'application externe. [Pour plus d'informations, reportez-vous à la section Définition du stockage et du formatage des champs dans le chapitre 2 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)

Lecture de chaînes vides issues d'Oracle. Lorsque vous lisez une base de données Oracle ou que vous écrivez dedans, souvenez-vous que, contrairement à SPSS Modeler et à la plupart des autres bases de données, Oracle traite et stocke les valeurs de chaîne vides comme des valeurs nulles. Autrement dit, les mêmes données extraites d'une base de données Oracle, ou d'un fichier ou d'une autre base de données peuvent se comporter différemment, et donc renvoyer des résultats différents.

Traitement des valeurs manquantes

Vous devez décider du traitement à appliquer aux valeurs manquantes en fonction de vos connaissances métier ou de votre domaine. Afin de réduire la durée d'apprentissage et d'améliorer la précision, il est recommandé de supprimer les blancs de l'ensemble de données. Par ailleurs, la présence de blancs peut mener à de nouvelles opportunités commerciales ou de nouvelles

idées. Pour déterminer la technique la plus appropriée, prenez en considération les aspects suivants de vos données :

- Taille de l'ensemble de données
- Nombre de champs contenant des blancs
- Quantité d'informations manquantes

En règle générale, vous pouvez appliquer deux méthodes :

- Vous pouvez exclure des champs ou des enregistrements avec des valeurs manquantes.
- Vous pouvez attribuer, remplacer ou forcer les valeurs manquantes avec un grand nombre de méthodes.

Ces deux méthodes peuvent être en grande partie automatisées grâce au noeud Audit données. Par exemple, vous pouvez générer un noeud Filtrer qui exclut les champs contenant trop de valeurs manquantes pour être utiles à la modélisation et générer un super noeud qui attribue les valeurs manquantes à l'un des champs ou à tous les champs restants. Voilà où la puissance réelle de l'audit intervient, vous permettant non seulement d'évaluer l'état actuel de vos données, mais également d'agir sur la base de cette évaluation. [Pour plus d'informations, reportez-vous à la section Préparation des données pour l'analyse \(Audit données\) dans le chapitre 8 dans *Guide des applications de IBM SPSS Modeler 14.2*.](#)

Traitement des enregistrements contenant des valeurs manquantes

Si la majorité des valeurs manquantes sont concentrées dans un nombre limité d'enregistrements, vous pouvez simplement les exclure. Par exemple, une banque conserve en général des enregistrements complets et détaillés sur ses clients prêteurs. Si, toutefois, la banque restreint moins le nombre de prêts accordés à son propre personnel, les données recueillies pour les prêts au personnel comporteraient plusieurs champs vides. Dans ce cas, deux options de traitement des valeurs manquantes sont à votre disposition :

- Vous pouvez utiliser un noeud Sélectionner pour supprimer les enregistrements relatifs au personnel.
- Si l'ensemble de données est volumineux, vous pouvez supprimer tous les enregistrements contenant des blancs. [Pour plus d'informations, reportez-vous à la section Sélection d'enregistrements contenant des valeurs manquantes dans le chapitre 6 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)

Traitement des champs contenant des valeurs manquantes

Si la majorité des valeurs manquantes est concentrée dans un petit nombre de champs, vous pouvez les traiter au niveau du champ plutôt qu'au niveau de l'enregistrement. Cette méthode vous permet de tester l'importance relative de certains champs avant de décider d'une méthode de traitement des valeurs manquantes. Si un champ n'est pas significatif en termes de modélisation, il n'est probablement pas utile de le conserver, quel que soit le nombre de valeurs manquantes qu'il comporte.

Par exemple, une société effectuant des études de marché peut recueillir des données à l'aide d'un questionnaire général de 50 questions. Deux des questions abordent l'âge et les idéologies politiques, informations que bon nombre de personnes hésitent à fournir. Dans ce cas, *Age* et *Convictions_politiques* comportent de nombreuses valeurs manquantes.

Niveau de mesure de champ

Pour choisir la méthode à utiliser, vous devez également tenir compte du niveau de mesure de champs contenant des valeurs manquantes.

Champs numériques. Pour les champs de type numérique, tels que les champs *Continus*, éliminez toutes les valeurs non numériques avant de construire un modèle. En effet, de nombreux modèles ne fonctionneront pas si des blancs sont inclus dans les champs numériques.

Champs catégoriels. Pour les champs catégoriels, tels que les champs *Nominaux* et *Booléens*, il n'est pas nécessaire de modifier les valeurs manquantes, bien que les modifier augmenterait la précision du modèle. Par exemple, un modèle qui utilise le champ *Sexe* fonctionnera toujours avec des valeurs insignifiantes, telles que *Y* et *Z*, mais le fait de supprimer toutes les valeurs autres que *M* et *F* amplifiera la précision du modèle.

Filtrage ou suppression de champs

Pour filtrer les champs présentant un nombre trop élevé de valeurs manquantes, vous disposez de plusieurs options :

- Vous pouvez utiliser le noeud Audit données pour filtrer les champs en fonction de la qualité. [Pour plus d'informations, reportez-vous à la section Filtrage de champs contenant des données manquantes dans le chapitre 6 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)
- Vous pouvez utiliser un noeud Sélection de fonction pour filtrer les champs dont le nombre de valeurs manquantes est supérieur à un pourcentage défini et les classer en fonction de l'importance par rapport à une cible donnée. [Pour plus d'informations, reportez-vous à la section Noeud Sélection de fonction dans le chapitre 4 dans Noeuds de modélisation IBM SPSS Modeler 14.2.](#)
- Au lieu de supprimer ces champs, vous pouvez utiliser un noeud Typer pour définir leur rôle sur la valeur Aucun. Ainsi, les champs resteront dans l'ensemble de données mais seront supprimés du processus de modélisation.

Attribution ou remplacement de valeurs manquantes

Lorsque le nombre des valeurs manquantes est faible, il peut être utile d'insérer des valeurs à la place des blancs. Vous pouvez le faire à partir du rapport Audit données, qui permet, selon les besoins, de définir des options pour des champs particuliers, puis de générer un super noeud qui attribue des valeurs selon un certain nombre de méthodes. Il s'agit de la méthode la plus souple, qui permet également de définir le traitement d'un grand nombre de champs dans un seul noeud.

Les méthodes suivantes sont disponibles pour attribuer des valeurs manquantes :

Fixe. Remplacement par une valeur fixe (soit la moyenne du champ, soit la moitié de l'intervalle, soit une constante que vous indiquez).

Aléatoire. Remplacement par une valeur aléatoire fondée sur une loi normale ou uniforme.

Expression. Permet d'indiquer une expression personnalisée. Par exemple, vous pourriez remplacer les valeurs par une variable globale créée par le noeud V. globales.

Algorithme. Remplacement par une valeur prévue par un modèle fondé sur l'algorithme C&RT. Chaque champ auquel une valeur est attribuée à l'aide de cette méthode est associé à un modèle C&RT distinct et à un noeud Remplacer qui remplace les valeurs non renseignées et les valeurs nulles par la valeur prédite par le modèle. Ensuite, un noeud Filtrer est utilisé pour supprimer les champs de prévision générés par le modèle.

Pour plus d'informations, reportez-vous à la section Attribution des valeurs manquantes dans le chapitre 6 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.

Vous pouvez également, pour forcer la valeur de certains champs, utiliser un noeud Typer pour que les types de champ ne couvrent que des valeurs correctes, puis définir la colonne *Vérifier* sur Forcer pour les champs dont les blancs doivent être remplacés. Pour plus d'informations, reportez-vous à la section Noeud Typer dans le chapitre 4 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.

Fonctions CLEM pour les valeurs manquantes

Plusieurs fonctions peuvent être utilisées pour traiter les valeurs manquantes. Les fonctions suivantes sont souvent utilisées dans les noeuds Sélectionner et Remplacer pour supprimer ou renseigner des valeurs manquantes :

- `count_nulls(LIST)`
- `@BLANK(FIELD)`
- `@NULL(FIELD)`
- `undef`

Les fonctions @ peuvent être utilisées en association avec la fonction @FIELD pour identifier la présence de blancs ou de valeurs nulles dans un ou plusieurs champs. Les champs peuvent être simplement signalés lorsqu'ils contiennent des blancs ou des valeurs nulles, ou ils peuvent être renseignés par des valeurs de remplacement ou utilisés dans un certain nombre d'autres opérations.

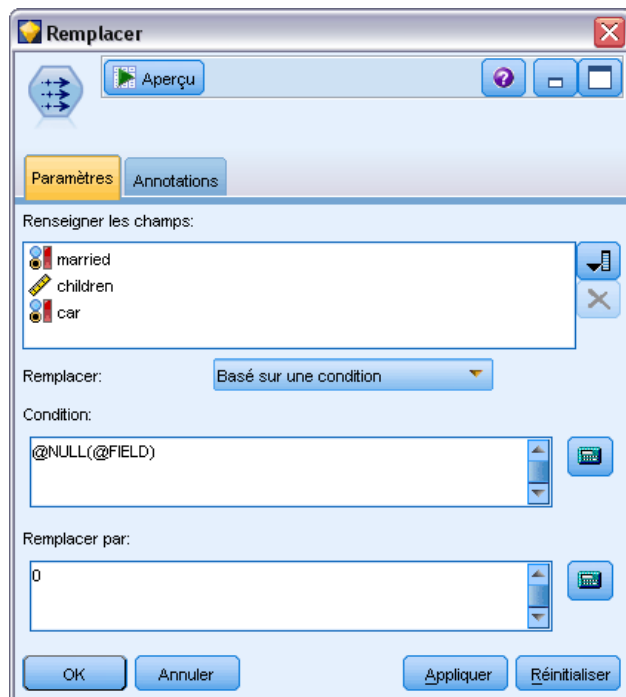
Vous pouvez compter les valeurs nulles dans une liste de champs de la manière suivante :

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Lors de l'utilisation des fonctions acceptant une liste de champs comme entrée, vous pouvez utiliser les fonctions spéciales @FIELDS_BETWEEN et @FIELDS_MATCHING, comme l'illustre l'exemple suivant :

```
count_nulls(@FIELDS_MATCHING('card*'))
```

Figure 6-2
Utilisation d'un noeud Remplacer pour remplacer les valeurs non renseignées par 0 dans les champs sélectionnés



Vous pouvez utiliser la fonction `undef` pour renseigner les champs avec la valeur manquante système, affichée sous la forme `$null$`. Par exemple, pour remplacer une valeur numérique, vous pouvez utiliser une instruction conditionnelle telle que :

```
if not(Age > 17) or not(Age < 66) then undef else Age endif
```

Elle remplace n'importe quel élément non compris dans l'intervalle par une valeur manquante système, affichée sous la forme `$null$`. En utilisant la fonction `not()`, vous pouvez récupérer toutes les autres valeurs numériques, y compris les valeurs négatives. [Pour plus d'informations, reportez-vous à la section Traitement des valeurs nulles et non renseignées dans le chapitre 8 sur p. 156.](#)

Remarque sur la suppression d'enregistrements

Lorsque vous utilisez un noeud Sélectionner pour supprimer des enregistrements, la syntaxe utilise une logique à trois valeurs et inclut automatiquement des valeurs nulles dans les instructions `SELECT`. Pour exclure les valeurs nulles (manquantes système) dans une expression `SELECT`, vous devez le spécifier explicitement en utilisant `and not` dans l'expression. Par exemple, pour sélectionner et inclure tous les enregistrements dans lesquels le type de médicament de prescription est Drug C (Médicament C), utilisez l'instruction `SELECT` suivante :

```
Drug = 'drugC' and not(@NULL(Drug))
```

Dans des cas semblables, les précédentes versions excluaient les valeurs nulles.

Création d'expressions CLEM

A propos de CLEM

Le langage CLEM (Clementine Language for Expression Manipulation) CLEM est efficace pour la compréhension et la manipulation des données circulant au sein des flux IBM® SPSS® Modeler. Les data miners utilisent beaucoup le langage CLEM dans les opérations de flux pour exécuter des tâches aussi simples que le calcul du profit à partir des données de coûts et de revenus, ou aussi complexes que la transformation de données du log Web en un ensemble de champs et d'enregistrements contenant des informations utilisables.

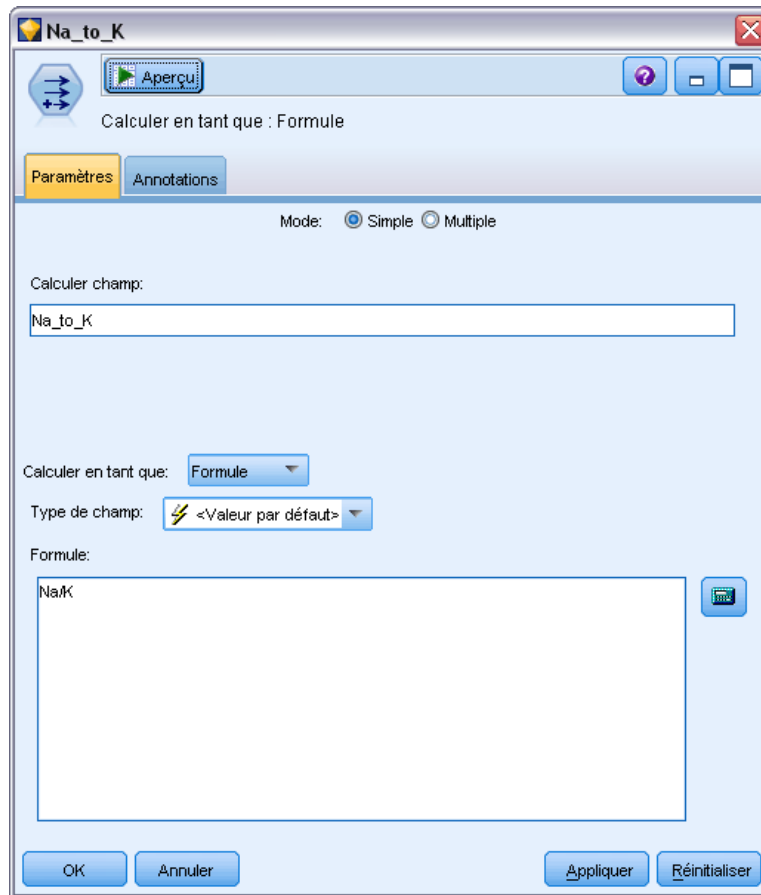
CLEM est utilisé dans SPSS Modeler pour :

- Comparer et évaluer les conditions sur les champs d'enregistrement.
- Calculer des valeurs pour les nouveaux champs.
- Calculer de nouvelles valeurs pour les champs existants.
- Comprendre la séquence des enregistrements.
- Insérer les données des enregistrements dans des rapports.

Scripts. Un sous-ensemble du langage CLEM peut également être utilisé lorsque vous générez des scripts dans l'interface utilisateur. Ceci vous permet d'effectuer la plupart des opérations de manipulation des données traitées de façon automatisée. [Pour plus d'informations, reportez-vous à la section Génération de scripts - Présentation dans le chapitre 2 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2*.](#)

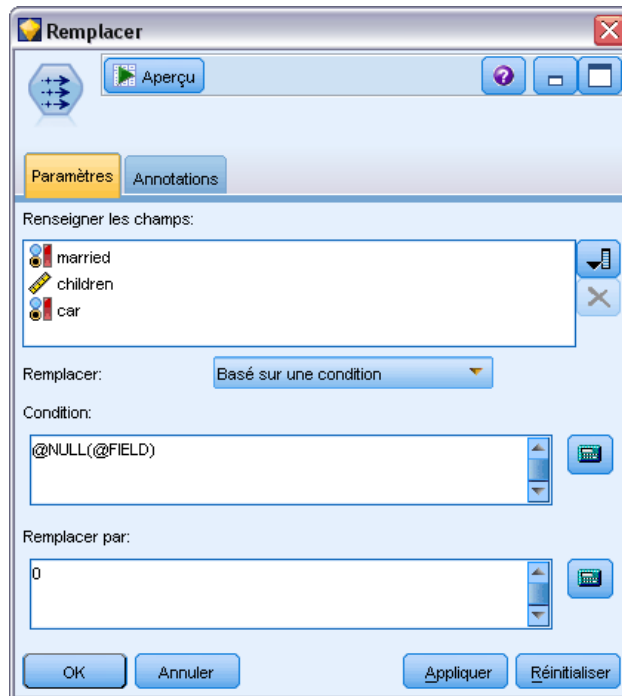
Les expressions CLEM sont indispensables à la préparation des données dans SPSS Modeler et peuvent être utilisées dans un grand nombre de noeuds, des opérations sur les enregistrements et les champs (Sélectionner, Equilibrer, Remplacer) aux graphiques et aux sorties (Analyse, Rapport, Table). Par exemple, vous pouvez utiliser CLEM dans un noeud Calculer pour créer un champ basé sur une formule telle qu'un rapport.

Figure 7-1
Noeud Calculer pour créer un nouveau champ basé sur une formule



Les expressions CLEM peuvent également être utilisées pour les opérations de recherche et de remplacement globales. Par exemple, l'expression @NULL(@FIELD) peut être utilisée dans un noeud Remplacer pour remplacer les **valeurs manquantes du système** par la valeur entière 0. (Pour remplacer les **valeurs manquantes de l'utilisateur**, également appelées blancs, utilisez la fonction @BLANK.)

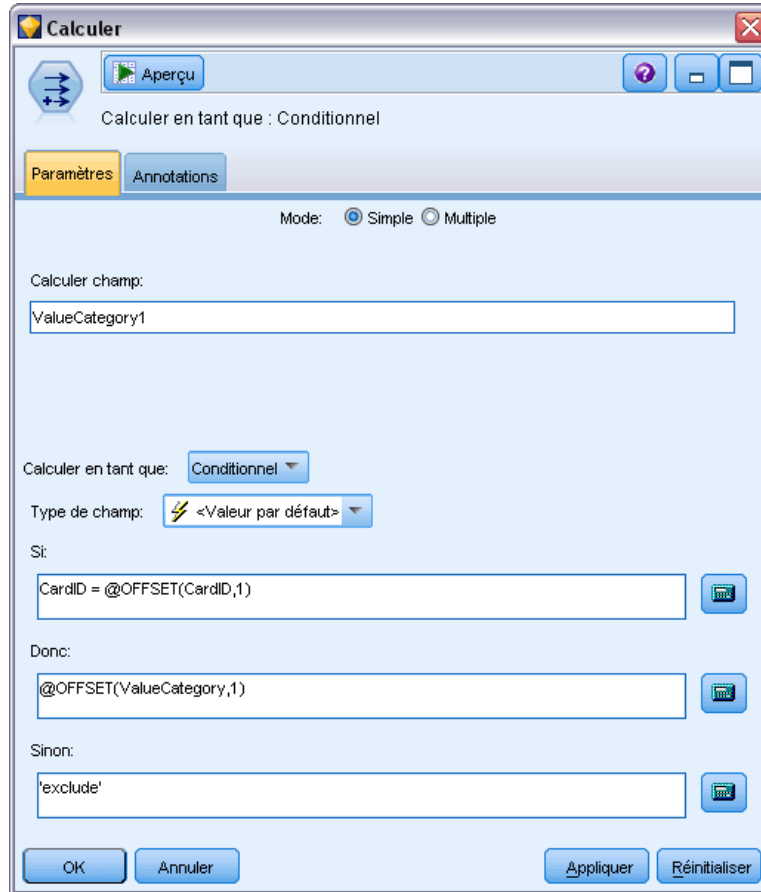
Figure 7-2
Noeud Remplacer remplaçant les valeurs manquantes du système par 0



Des expressions CLEM plus complexes peuvent également être créées. Par exemple, vous pouvez calculer de nouveaux champs en fonction d'un ensemble conditionnel de règles.

Figure 7-3

La valeur Conditionnel effectue le calcul en comparant les valeurs d'un champ à celles des champs qui le précèdent.



CLEM Exemples

Les exemples d'expressions suivants illustrent la syntaxe et les types d'expression à employer dans CLEM.

Expressions simples

Les formules peuvent être très simples, comme celle ci-dessous qui calcule un nouveau champ en fonction des valeurs des champs *After* (Après) et *Before* (Avant) :

$(\text{After} - \text{Before}) / \text{Before} * 100.0$

Les noms de champ ne sont pas entourés de guillemets si vous faites référence à leur valeur.

De même, l'expression suivante renvoie simplement le journal de chaque valeur du champ *salary* (salaire).

log(salary)

Expressions complexes

Les expressions peuvent également être longues et complexes. L'expression suivante renvoie la valeur *true* (*vrai*) si la valeur des deux champs (*\$KX-Kohonen* et *\$KY-Kohonen*) est comprise dans les intervalles spécifiés. Dans cet exemple, les noms de champ sont entourés de guillemets simples car ils contiennent des caractères spéciaux.

```
(' $KX-Kohonen' >= -0.2635771036148072 and ' $KX-Kohonen' <= 0.3146203637123107
and ' $KY-Kohonen' >= -0.18975617885589602 and
' $KY-Kohonen' <= 0.17674794197082522) -> T
```

Diverses fonctions, comme les fonctions de chaîne, nécessitent la saisie de plusieurs paramètres dans une syntaxe correcte. Dans l'exemple suivant, la fonction `subscr` est utilisée pour renvoyer le premier caractère du champ *produce_ID*, et indiquer si un élément est organique, génétiquement modifié ou traditionnel. Les résultats de l'expression sont décrits par `-> `result``.

```
subscr(1,produce_ID) -> `c`
```

De même, l'expression suivante sera :

```
stripchar('3','123') -> `12`
```

Il est important de noter que les caractères sont toujours entourés de guillemets simples inversés.

Combinaison de fonctions dans une expression

Les expressions CLEM sont fréquemment constituées d'une combinaison de fonctions. La fonction suivante combine `subscr` et `lowertoupper` pour renvoyer le premier caractère de *produce_ID* et le convertir en lettre majuscule.

```
lowertoupper(subscr(1,produce_ID)) -> `C`
```

Cette expression peut également être écrite en sténographie :

```
lowertoupper(produce_ID(1)) -> `C`
```

Voici une autre combinaison de fonctions fréquemment utilisée :

```
locchar_back(`n`, (length(web_page)), web_page)
```

Cette expression repère le caractère ``n`` dans les valeurs du champ *web_page* en commençant par le dernier caractère de la valeur, puis en remontant vers le début. En incluant également la fonction `length`, l'expression calcule de manière dynamique la longueur de la valeur actuelle au lieu d'utiliser un nombre statique, comme 7, qui ne fonctionnerait pas pour les valeurs de moins de sept caractères.

Fonctions spéciales

De nombreuses fonctions spéciales (précédées du symbole @) sont disponibles. Voici quelques fonctions couramment utilisées :

@BLANK('referrer ID') -> T

Les fonctions spéciales sont fréquemment utilisées conjointement, méthode souvent utilisée pour signaler les blancs dans plusieurs champs simultanément.

@BLANK(@FIELD)-> T

Pour consulter d'autres exemples, reportez-vous à la documentation CLEM. [Pour plus d'informations, reportez-vous à la section Présentation de la référence CLEM dans le chapitre 8 sur p. 122.](#)

Valeurs et types de données

Comme les formules, les expressions CLEM sont construites à partir de valeurs, de noms de champs, d'opérateurs et de fonctions. L'expression CLEM valide la plus simple est une valeur ou un nom de champ. Exemples de valeurs valides :

3
1.79
'banana'

Exemples de noms de champs :

Product_ID
'\$P-NextField'

Product étant le nom d'un champ issu d'un ensemble de données du panier du marché et *\$P-NextField*, le nom d'un paramètre. La valeur de l'expression est la valeur du champ nommé. En général, les noms de champs commencent par une lettre, et peuvent comporter des chiffres et des traits de soulignement (_). Toutefois, vous n'êtes pas obligé de suivre ces règles si vous placez le nom entre guillemets. Les valeurs CLEM peuvent être les suivantes :

- Chaînes—par exemple, "c1", "Type 2", "a piece of free text"
- Entiers—par exemple, 12, 0, -189
- Nombres réels : par exemple, 12.34, 0.0, -0.0045
- Champs Date/heure : par exemple, 05/12/2002, 12/05/2002, 12/05/02

Vous pouvez également utiliser les éléments suivants :

- Codes de caractère—par exemple, `a` or 3
- Listes d'éléments—par exemple, [1 2 3], ['Type 1' 'Type 2']

Les listes et les codes de caractère ne représentent généralement pas des valeurs de champ. En général, ils jouent le rôle d'arguments de fonctions CLEM.

Règles des guillemets

Même si le logiciel est flexible quant à la détermination des champs, des valeurs, des paramètres et des chaînes utilisés dans une expression CLEM, les règles générales suivantes dressent la liste des conseils à suivre lorsque vous créez des expressions :

- **Chaînes**—utilisez toujours des guillemets doubles lorsque vous écrivez des chaînes ("Type 2" ou "value"). Vous pouvez également utiliser des guillemets simples mais cela risque de créer une confusion avec les champs placés entre guillemets.
- **Caractères**—utilisez toujours des guillemets simples inversés ('). Prenons l'exemple du caractère d dans la fonction stripchar('d',"drugA"). La seule exception concerne l'utilisation d'un entier pour désigner un caractère spécifique d'une chaîne. Prenons l'exemple du caractère 5 dans la fonction lowertoupper("drugA"(5)) → "A". *Remarque* : sur les claviers britanniques et américains standard, la touche du guillemet simple inversé (accent grave, unicode 0060) se trouve juste à côté de la touche Echap.
- **Champs**—Les champs utilisés dans les expressions CLEM ne sont généralement pas entourés de guillemets (subscr(2,arrayID)) → CHAR). Vous pouvez utiliser des guillemets simples si le champ comprend un espace ou un caractère spécial ('Order Number'). les champs placés entre guillemets mais non définis dans l'ensemble de données seront lus comme des chaînes.
- **Paramètres**—utilisez toujours des guillemets simples ('\$P-threshold').

Expressions et conditions

Les expressions CLEM utilisées pour le calcul de nouvelles valeurs peuvent renvoyer un résultat. Par exemple :

```
Weight * 2.2
Age + 1
sqrt(Signal-Echo)
```

Elles peuvent également indiquer un résultat *vrai* ou *faux* quand elles sont utilisées pour sélectionner une condition. Par exemple :

```
Drug = "drugA"
Age < 16
not(PowerFlux) and Power > 2000
```

Vous pouvez combiner des opérateurs et des fonctions de façon arbitraire dans les expressions CLEM. Par exemple :

```
sqrt(abs(Signal)) * max(T1, T2) + Baseline
```

Les parenthèses et l'ordre de priorité des opérateurs déterminent l'ordre dans lequel l'expression est évaluée. Dans cet exemple, l'ordre d'évaluation est le suivant :

- abs(Signal) est évalué et sqrt est appliqué à son résultat.
- max(T1, T2) est évalué.
- Les deux résultats sont multipliés : x est prioritaire par rapport à +.
- Enfin, Baseline est ajoutée au résultat.

L'ordre décroissant de priorité (à savoir, des opérations exécutées en premier aux opérations exécutées en dernier) est le suivant :

- arguments de fonctions
- appels de fonctions
- xx
- x / mod div rem
- + -
- > < >= <= /= == !=

Si vous voulez ignorer l'ordre de priorité, ou si vous doutez de l'ordre de l'évaluation, vous pouvez utiliser des parenthèses pour rendre ce dernier explicite. Par exemple :

```
sqrt(abs(Signal)) * (max(T1, T2) + Baseline)
```

Paramètres de flux, de session et de super noeud

Vous pouvez définir des paramètres afin de les utiliser dans les expressions CLEM et pour la génération de scripts. Il s'agit, en effet, de variables définies par l'utilisateur qui sont enregistrées et conservées dans le flux, la session ou le super noeud en cours ; elles sont accessibles à partir de l'interface utilisateur, ainsi que via la fonction de génération de scripts. Si vous enregistrez un flux, par exemple, les paramètres définis pour ce dernier sont également enregistrés. (Cela permet de les différencier des variables de script locales qui ne peuvent être utilisées que dans le script dans lequel elles sont déclarées.) Dans le cadre de la génération de scripts, les paramètres sont souvent intégrés dans une expression CLEM et la valeur du paramètre est spécifiée dans le script.

La portée d'un paramètre dépend de l'endroit où ce dernier est défini :

- Les paramètres de flux peuvent être définis dans un script de flux ou dans la boîte de dialogue des propriétés du flux ; ils sont disponibles pour tous les noeuds du flux. Ils sont affichés sur la liste des paramètres dans le Générateur de formules.
- Les paramètres de session peuvent être définis dans un script autonome ou dans la boîte de dialogue des paramètres de session. Ils sont disponibles pour tous les flux utilisés dans la session actuelle (tous les flux répertoriés dans l'onglet Flux du panneau des gestionnaires).

Il est également possible de définir des paramètres pour les super noeuds ; dans ce cas, les paramètres ne sont visibles que par les noeuds encapsulés au sein de ce super noeud. [Pour plus d'informations, reportez-vous à la section Définitions des paramètres de super noeud dans le chapitre 9 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Utilisation de paramètres dans les expressions CLEM

Les paramètres sont représentés dans les expressions CLEM par \$P-pname, où pname est le nom du paramètre. Lorsqu'ils sont utilisés dans des expressions CLEM, les paramètres doivent être placés entre guillemets simples—par exemple, '\$P-scale'.

Vous pouvez afficher les paramètres disponibles en utilisant le Générateur de formules. Pour afficher les paramètres actuels :

- ▶ Dans toute boîte de dialogue acceptant des expressions CLEM, cliquez sur le bouton Générateur de formules.
- ▶ Dans la liste Champs, sélectionnez Paramètres.

Vous pouvez sélectionner dans la liste les paramètres à insérer dans l'expression CLEM. [Pour plus d'informations, reportez-vous à la section Sélection de Champs, Paramètres et Variables globales sur p. 116.](#)

Utilisation de chaînes

Il existe un grand nombre d'opérations disponibles à utiliser pour les chaînes, notamment :

- Conversion d'une chaîne en majuscules ou en minuscules : `—uppertolower(CHAR)`.
- Suppression des caractères spécifiés, comme ``ID_`` ou ``$``, d'une variable de chaîne—`stripchar(CHAR,STRING)`.
- Détermination de la longueur (nombre de caractères) d'une variable de chaîne—`length(STRING)`.
- Vérification de l'ordre alphabétique des valeurs de chaîne—`alphabefore(STRING1, STRING2)`.
- Suppression des espaces blancs situés au début ou à la fin des valeurs—`trim(STRING)`, `trim_start(STRING)`, ou `trimend(STRING)`.
- Extraire les premiers ou les derniers caractères *n* d'une chaîne—`startstring(LENGTH, STRING)` ou `endstring(LENGTH, STRING)`. Par exemple, imaginons un champ *élément* qui associe un nom de produit à un code d'identification de 4 caractères (ACME CAMERA-D109). Pour créer un nouveau champ qui contient uniquement le code à 4 caractères, spécifiez la formule suivante dans un noeud Calculer.

```
endstring(4, item)
```

- Correspondance d'un motif spécifique—`STRING matches PATTERN`. Par exemple, pour sélectionner des personnes avec "marché" dans leur intitulé de poste, vous pouvez spécifier l'élément suivant dans un noeud Sélectionner :

```
job_title matches "**market**"
```

- Remplacement de toutes les instances de sous-chaîne par une chaîne—`replace(SUBSTRING, NEWSUBSTRING, STRING)`. Par exemple, pour remplacer toutes les instances d'un caractère non pris en charge, comme une barre verticale (|) par un point virgule, avant l'exploration du texte, utilisez la fonction `replace` dans un noeud Remplacer. Dans Renseigner les champs :,

sélectionnez tous les champs où le caractère peut être présent. Pour la condition Remplacer : sélectionnez Toujours, et spécifiez la condition suivante dans Remplacer par :

```
replace(';',',',@FIELD)
```

- Calcul d'un champ booléen basé sur la présence d'une sous-chaîne spécifique. Par exemple, vous pouvez utiliser un noeud Calculer afin de générer un champ booléen distinct pour chaque réponse ; entrez pour cela une expression semblable à ce qui suit :

```
hassubstring(museums,"museum_of_design")
```

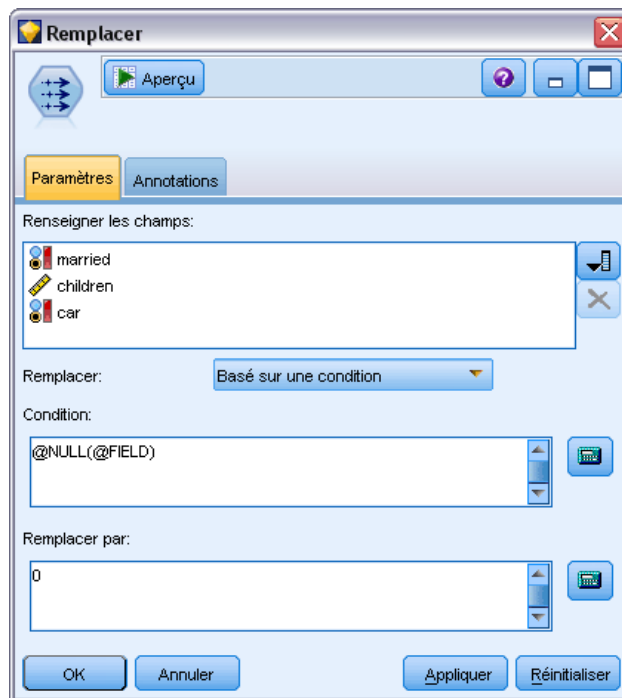
Pour plus d'informations, reportez-vous à la section [Fonctions sur chaînes dans le chapitre 8](#) sur p. 139.

Traitement des valeurs non renseignées et des valeurs manquantes

Le remplacement des valeurs non renseignées ou des valeurs manquantes est une tâche de préparation de données habituelle pour les data miners. CLEM vous propose plusieurs outils permettant d'automatiser le traitement des valeurs non renseignées. Les valeurs non renseignées sont généralement traitées dans le noeud Remplacer ; cependant, les fonctions suivantes peuvent être utilisées dans n'importe quel noeud acceptant les expressions CLEM :

- `@BLANK(FIELD)` peut être utilisée pour déterminer les enregistrements dont les valeurs sont vides pour un champ particulier tel que *Age*.
- `@NULL(FIELD)` peut être utilisée pour déterminer les enregistrements dont les valeurs sont absentes du système pour les champs concernés. Dans IBM® SPSS® Modeler, les valeurs système manquantes sont affichées en tant que valeurs \$null\$.

Figure 7-4
Noeud Remplacer remplaçant les valeurs manquantes du système par 0



Pour plus d'informations, reportez-vous à la section [Traitement des valeurs nulles et non renseignées](#) dans le chapitre 8 sur p. 156.

Utilisation de nombres

De nombreuses opérations standard sur des valeurs numériques sont disponibles dans IBM® SPSS® Modeler, notamment :

- Calcul du sinus de l'angle spécifié.— $\sin(\text{NUM})$
- Calcul du logarithme naturel de champs numériques— $\log(\text{NUM})$
- Calcul de la somme de deux nombres— $\text{NUM1} + \text{NUM2}$

Pour plus d'informations, reportez-vous à la section [Fonctions numériques](#) dans le chapitre 8 sur p. 135.

Utilisation de l'heure et de la date

Les formats d'heure et de date peuvent varier en fonction de votre source de données et de vos paramètres régionaux. Les formats de date et d'heure utilisés sont propres à chaque flux et sont définis dans la boîte de dialogue Propriétés du flux. Les exemples suivants sont des fonctions couramment utilisées avec les champs de date et d'heure.

Calcul de la durée écoulée

Vous pouvez facilement calculer la durée écoulée à partir d'une date de référence en utilisant une famille de fonctions identique à la suivante. Cette fonction indique le temps en mois en utilisant d'abord la date de référence, puis celle représentée par la chaîne de date DATE, sous forme de nombre réel. Ce chiffre est approximatif et utilise une base de 30 jours.

```
date_in_months(Date)
```

Comparaison de valeurs Date/heure

Les valeurs des champs Date/heure peuvent être comparées dans les enregistrements en utilisant des fonctions identiques à la suivante. Cette fonction renvoie une valeur vraie (*true*) si la chaîne de date DATE1 représente une date antérieure à celle représentée par la chaîne de date DATE2. Sinon, cette fonction renvoie la valeur 0.

```
date_before(Date1, Date2)
```

Calcul des différences

Vous pouvez également calculer la différence entre deux heures et deux dates en utilisant des fonctions du type :

```
date_weeks_difference(Date1, Date2)
```

Cette fonction indique le temps en semaines en partant de la date représentée par la chaîne de date DATE1 jusqu'à la date représentée par la chaîne de date DATE2, sous forme de nombre réel. La semaine de base utilisée comprend 7,0 jours. Si DATE2 est antérieure à DATE1, cette fonction renvoie un nombre négatif.

Date du jour

La date du jour peut être ajoutée à l'ensemble des données en utilisant la fonction @TODAY. La date du jour est ajoutée en tant que chaîne au champ spécifié ou au nouveau champ en utilisant le format de date sélectionné dans la boîte de dialogue des propriétés du flux. [Pour plus d'informations, reportez-vous à la section Fonctions date et heure dans le chapitre 8 sur p. 145.](#)

Récapitulation de plusieurs champs

Le langage CLEM comporte un certain nombre de fonctions renvoyant des statistiques récapitulatives sur plusieurs champs. Ces fonctions s'avèrent particulièrement utiles lors de l'analyse des données d'enquête, où les diverses réponses à une question peuvent être stockées dans plusieurs champs. [Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples sur p. 112.](#)

Fonctions de comparaison

Vous pouvez comparer les valeurs de plusieurs champs en utilisant les fonctions `min_n` et `max_n`, par exemple :

```
max_n(['card1fee' 'card2fee' 'card3fee' 'card4fee'])
```

Vous pouvez également utiliser diverses fonctions de comptage pour dénombrer les valeurs qui répondent à des critères spécifiques, même quand ces valeurs sont stockées dans plusieurs champs. Par exemple, pour compter le nombre de cartes conservées depuis plus de cinq ans :

```
count_greater_than(5, ['cardtenure' 'card2tenure' 'card3tenure'])
```

Pour compter les valeurs nulles sur le même ensemble de champs :

```
count_nulls(['cardtenure' 'card2tenure' 'card3tenure'])
```

Remarque : Cet exemple compte le nombre de cartes conservées, pas le nombre de personnes qui les possèdent. [Pour plus d'informations, reportez-vous à la section Fonctions de comparaison dans le chapitre 8 sur p. 132.](#)

Fonctions numériques

Vous pouvez obtenir des statistiques pour plusieurs champs en utilisant les fonctions `sum_n`, `mean_n` et `sdev_n` par exemple :

```
sum_n(['card1bal' 'card2bal' 'card3bal'])
```

```
mean_n(['card1bal' 'card2bal' 'card3bal'])
```

[Pour plus d'informations, reportez-vous à la section Fonctions numériques dans le chapitre 8 sur p. 135.](#)

Génération de listes de champs

Lors de l'utilisation des fonctions acceptant une liste de champs comme entrée, vous pouvez utiliser les fonctions spéciales `@FIELDS_BETWEEN(start, end)` et `@FIELDS_MATCHING(pattern)` comme entrées. Par exemple, en supposant que l'ordre des champs correspond à celui illustré dans l'exemple `sum_n` précédent, la formule suivante reviendrait au même :

```
sum_n(@FIELDS_BETWEEN(card1bal, card3bal))
```

Sinon, pour compter le nombre de valeurs nulles dans tous les champs commençant par « *card* » :

```
count_nulls(@FIELDS_MATCHING('card*'))
```

[Pour plus d'informations, reportez-vous à la section Champs spéciaux dans le chapitre 8 sur p. 157.](#)

Utilisation des données à réponses multiples

Plusieurs fonctions de comparaison permettent d'analyser des données à réponses multiples, y compris :

- `value_at`
- `first_index / last_index`
- `first_non_null / last_non_null`
- `first_non_null_index / last_non_null_index`
- `min_index / max_index`

Par exemple, imaginons une question à réponses multiples posée comme première, deuxième et troisième plus importantes raisons pour décider d'un achat précis (par exemple, le prix, la recommandation personnelle, revue, fournisseur local, etc.) Dans ce cas, vous pouvez déterminer l'importance du prix en calculant l'index du champ dans lequel il a été inclus pour la première fois :

```
first_index("price", [Reason1 Reason2 Reason3])
```

De même, supposez que vous avez demandé à vos clients de classer trois voitures par ordre de préférence et que vous avez codé les réponses en trois champs séparés, comme suit :

id client	voiture1	voiture2	voiture3
101	1	3	2
102	3	2	1
103	2	3	1

Dans ce cas, vous pouvez déterminer l'index du champ de la voiture préférée (classée n°1 ou le rang le plus bas) à l'aide de la fonction `min_index` :

```
min_index(['car1' 'car2' 'car3'])
```

Pour plus d'informations, reportez-vous à la section Fonctions de comparaison dans le chapitre 8 sur p. 132.

Référencement des ensembles à réponses multiples

La fonction spéciale `@MULTI_RESPONSE_SET` peut être utilisée pour référencer tous les champs dans un ensemble à réponses multiples. Par exemple, si les trois champs *voiture* de l'exemple précédent sont compris dans un ensemble à réponses multiples appelé *car_rankings*, le même résultat serait renvoyé :

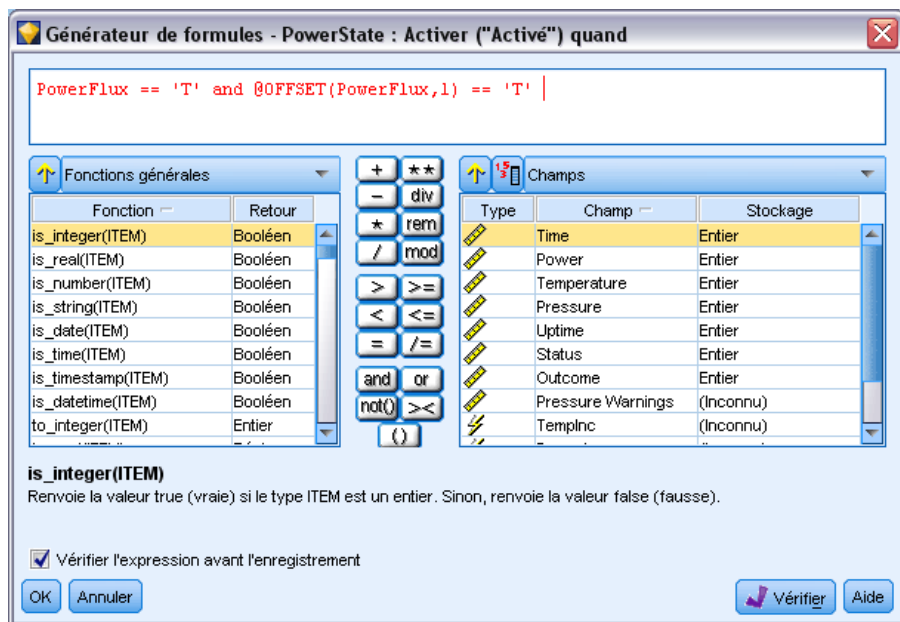
```
max_index(@MULTI_RESPONSE_SET("car_rankings"))
```

Pour plus d'informations, reportez-vous à la section Modification des ensembles de réponses multiples dans le chapitre 4 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.

Générateur de formules

Vous pouvez saisir des expressions CLEM manuellement ou utiliser le générateur de formules qui affiche une liste complète de fonctions et d'opérateurs CLEM ainsi que des champs de données du flux actuel, vous permettant de construire rapidement des expressions sans mémoriser les noms exacts des champs ou des fonctions. Par ailleurs, les commandes du Générateur ajoutent automatiquement les guillemets appropriés autour des champs et des valeurs, ce qui facilite la création d'expressions à la syntaxe correcte.

Figure 7-5
Boîte de dialogue Générateur de formules

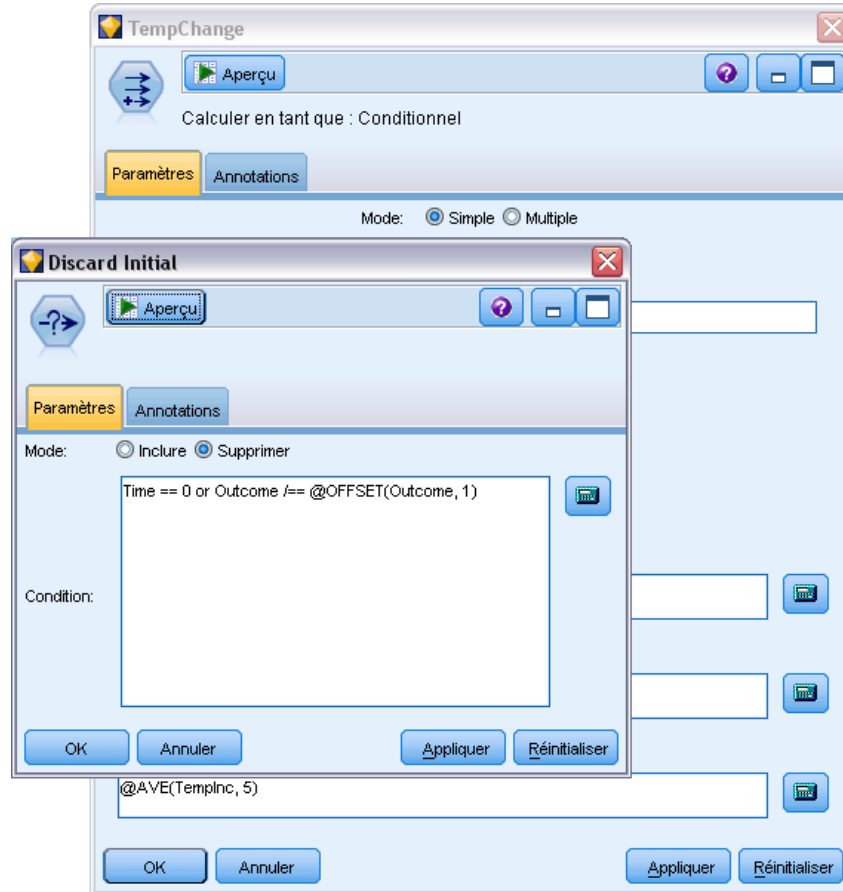


Remarque : il n'est pas possible d'utiliser la boîte de dialogue Générateur de formules pour générer des scripts ou définir des paramètres.

Accès au Générateur de formules

Le Générateur de formules est disponible dans tous les noeuds où des expressions CLEM sont utilisées, y compris les noeuds Sélectionner, Equilibrer, Calculer, Remplacer, Nuage, Courbes, analyses, rapports et tables. Vous pouvez l'ouvrir en cliquant sur le bouton représentant une calculatrice situé juste à droite du champ de formule.

Figure 7-6
Plusieurs noeuds contenant le bouton Générateur de formules



Création d'expressions

Non seulement la boîte de dialogue Générateur de formules contient des listes complètes de champs, de fonctions et d'opérateurs, mais elle vous permet également d'accéder à des valeurs de données si vos données sont instanciées.

Pour créer une formule à l'aide du générateur de formules

- ▶ Saisissez votre expression dans le champ en vous servant des listes de fonctions et de champs comme références.

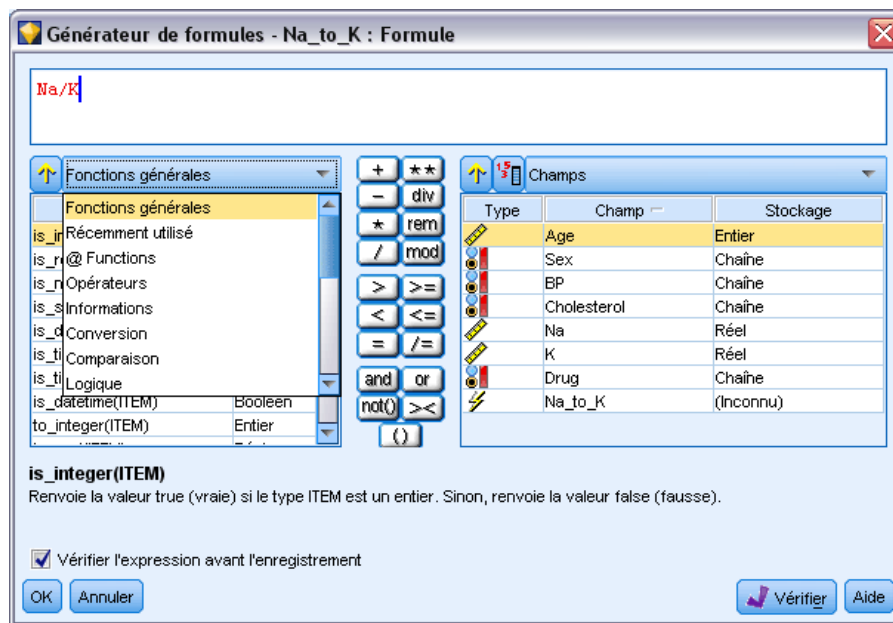
ou

- ▶ Sélectionnez les champs et fonctions requis dans les listes déroulantes.
- ▶ Pour ajouter le champ ou la fonction au champ des formules, cliquez sur le bouton en forme de flèche jaune ou double-cliquez directement sur la fonction ou le champ.
- ▶ Utilisez les boutons des opérandes situés au milieu de la boîte de dialogue pour insérer les opérandes dans l'expression.

Sélection de fonctions

La liste des fonctions affiche toutes les fonctions et tous les opérateurs CLEM disponibles. Faites défiler la liste pour sélectionner une fonction ou utilisez la liste déroulante pour afficher un sous-ensemble des fonctions ou des opérateurs. Les fonctions disponibles sont regroupées en catégories pour faciliter vos recherches.

Figure 7-7
Liste déroulante des fonctions



Deux catégories peuvent être particulièrement utiles :

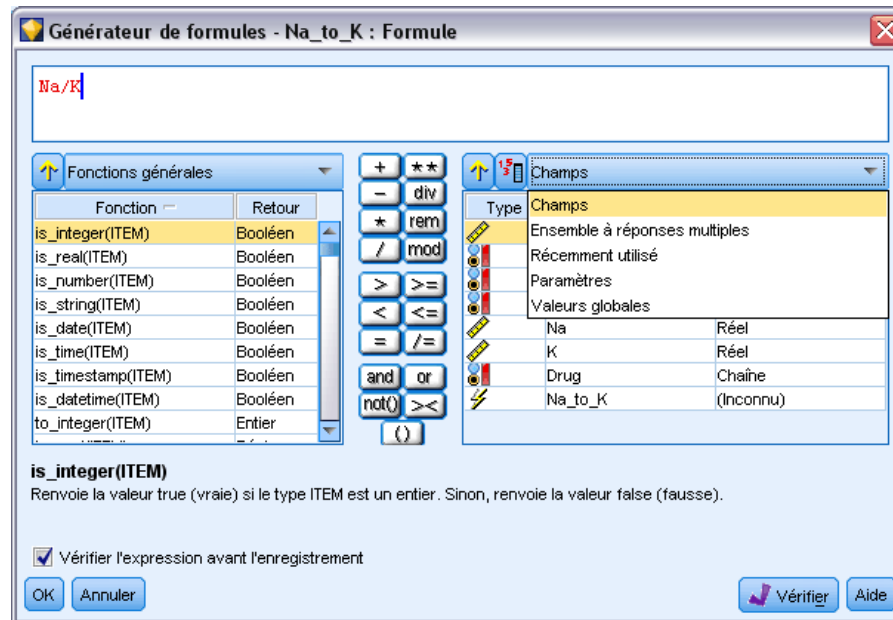
- **Toutes les fonctions** contient la liste complète des fonctions CLEM disponibles.
- **Récemment utilisé** contient la liste des fonctions CLEM utilisées dans la session en cours.

Une fois que vous avez sélectionné un groupe de fonctions, double-cliquez dessus pour les insérer dans le champ des formules à l'endroit indiqué par la position du curseur.

Sélection de Champs, Paramètres et Variables globales

La liste des champs affiche tous les champs actuellement disponibles dans le flux de données. Sélectionnez un champ dans la liste. Pour ajouter un champ à l'expression, cliquez sur le bouton en forme de flèche jaune ou double-cliquez directement sur le champ.

Figure 7-8
Générateur de formules : Liste Champs



Pour plus d'informations, reportez-vous à la section Paramètres de flux, de session et de super noeud sur p. 106.

En plus des champs, vous pouvez également choisir parmi les éléments suivants :

Ensembles à réponses multiples. Pour plus d'informations, consultez le guide *Noeuds IBM SPSS Modeler Source, d'exécution et de sortie*.

Récemment utilisés contient une liste des champs, des ensembles à réponses multiples, des paramètres et des valeurs globales utilisés dans la session actuelle.

Paramètres : Pour plus d'informations, reportez-vous à la section Paramètres de flux, de session et de super noeud sur p. 106.

Valeurs globales. Pour plus d'informations, consultez le guide *Noeuds IBM SPSS Modeler Source, d'exécution et de sortie*.

Affichage ou sélection de valeurs

Les valeurs de champs peuvent être affichées à partir de plusieurs endroits dans le système, y compris le générateur de formules, les rapports d'audits de données et lors de la modification de valeurs futures dans un noeud Intervalles de temps. Veuillez noter que pour utiliser cette fonction, les données doivent être entièrement instanciées dans une source ou un noeud Typier afin

que le stockage, les types et les valeurs soient connus. Pour plus d'informations, reportez-vous à la section [Utilisation de la boîte de dialogue Valeurs](#) dans le chapitre 4 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.

Figure 7-9

Liste des champs avec leurs valeurs indiquées pour le champ sélectionné



- Pour afficher les valeurs d'un champ à partir du générateur de formules ou d'un noeud Intervalles de temps, sélectionnez le champ requis et cliquez sur le bouton de sélection des valeurs pour ouvrir une boîte de dialogue des valeurs du champ sélectionné. Vous pouvez ensuite sélectionner une valeur et cliquer sur Insérer pour coller la valeur dans la formule ou la liste actuelle.

Figure 7-10

Bouton de sélection des valeurs

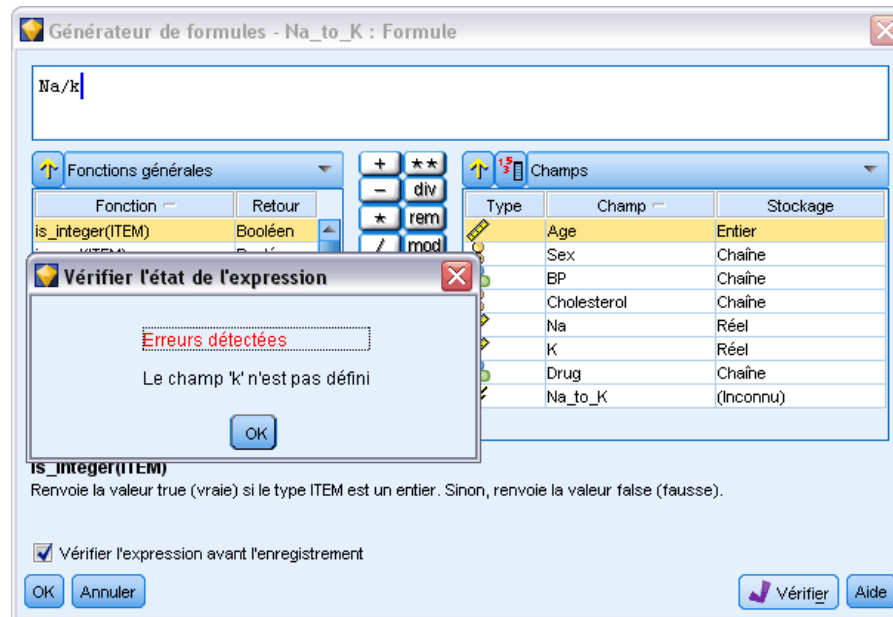


Pour les champs booléens et les champs nominaux, toutes les valeurs définies sont présentées. Pour les champs d'intervalles continus (intervalle numérique), les valeurs minimales et maximales sont affichées.

Vérification des expressions CLEM

Cliquez sur Vérifier dans le Générateur de formules (angle inférieur droit) pour valider l'expression. Expressions qui n'ont pas encore été vérifiées et affichées en rouge. En cas d'erreurs, un message indiquant la cause s'affiche.

Figure 7-11
Expression CLEM invalide



La vérification porte sur les éléments suivants :

- Utilisation correcte des guillemets dans les noms de valeurs et de champs
- Utilisation correcte des paramètres et des variables globales
- Utilisation correcte des opérateurs
- Existence des champs référencés
- Existence et définition des valeurs globales référencées

Si votre syntaxe comporte des erreurs, essayez de recréer l'expression en utilisant les listes et les boutons des opérateurs au lieu de taper l'expression manuellement. Cette méthode ajoute automatiquement les guillemets corrects pour les champs et les valeurs.

Rechercher et remplacer

La boîte de dialogue Rechercher/Remplacer est disponible aux emplacements où vous modifier un script ou le texte d'une expression, par exemple l'éditeur de script ou le générateur d'expression CLEM, ou encore lors de la définition d'un modèle dans le noeud Rapport. Lorsque vous éditez du texte dans l'un de ces champs, appuyez sur Ctrl-F pour accéder à la boîte de dialogue, en vous assurant que le curseur est centré sur une zone de texte. Si vous travaillez dans un noeud

Remplacer, par exemple, vous pouvez accéder à la boîte de dialogue depuis toute zone de texte de l'onglet Paramètres, ou depuis le champ de texte du générateur d'expression.

Figure 7-12
Boîte de dialogue Rechercher/Remplacer



- ▶ Lorsque le curseur se trouve sur une zone de texte, appuyez sur Ctrl+F pour accéder à la boîte de dialogue Rechercher/Remplacer.
- ▶ Entrez le texte que vous souhaitez rechercher ou faites un choix dans la liste déroulante des éléments récemment consultés.
- ▶ Saisissez éventuellement un texte de remplacement.
- ▶ Cliquez sur Suivant pour lancer la recherche.
- ▶ Cliquez sur Remplacer pour remplacer la sélection courante, ou Remplacer tout pour mettre à jour certaines instances ou les instances sélectionnées.
- ▶ La boîte de dialogue se ferme après chaque opération. Appuyez sur F3 depuis toute zone de texte pour répéter la dernière opération de recherche ou sur Ctrl+F pour accéder de nouveau à la boîte de dialogue.

Options de recherche

Respecter la casse. Spécifie si l'opération de recherche est sensible à la casse ; par exemple, si *myvar* correspond à *myVar*. Le texte de remplacement est toujours inséré exactement tel qu'il a été saisi, quel que soit le réglage de ce paramètre.

Mot entier. Spécifie si l'opération de recherche doit porter sur le texte inséré dans des mots. Si cette option est sélectionnée, par exemple, une recherche portant sur *spider* ne produira pas la réponse *spiderman* ou *spider-man*.

Caractères génériques. Spécifie si la syntaxe des caractères génériques est utilisée (voir la section suivante). Lorsqu'elle est sélectionnée, l'option Mot entier est désactivée et sa valeur est ignorée.

Texte sélectionné. Contrôle la portée de la recherche lorsque vous utilisez l'option Remplacer tout.

Syntaxe des caractères génériques

Les caractères génériques vous permettent de rechercher des caractères spéciaux tels que les tabulations ou les sauts de ligne, des classes ou des intervalles de caractères telles que *a* à *d*, toute valeur numérique ou non et les limites telles que le début ou la fin d'une ligne. Les types d'expression suivants sont pris en charge :

Correspondances de caractères

Caractères	Correspondances
x	Le caractère x
\\	Le caractère barre oblique inversée
\On	Le caractère présentant la valeur octale On (0 <= n <= 7)
\Onn	Le caractère présentant la valeur octale Onn (0 <= n <= 7)
\Omnn	Le caractère présentant la valeur octale Omnn (0 <= m <= 3, 0 <= n <= 7)
\xhh	Le caractère présentant la valeur hexadécimale Oxhh
\uhhhh	Le caractère présentant la valeur hexadécimale Oxhhhh
\t	Le caractère tabulation ('\\u0009')
\n	Le caractère saut de ligne (retour à la ligne) ('\\u000A')
\r	Le caractère retour chariot ('\\000D')
\f	Le caractère alimentation de formulaire ('\\u000C')
\a	Le caractère alerte (sonnerie) ('\\u0007')
\e	Le caractère d'échappement ('\\u001B')
\cx	Le caractère de contrôle correspondant à x

Classes de caractères correspondantes

Classes de caractères	Correspondances
[abc]	a, b ou c (classe simple)
[^abc]	Tout caractère excepté a, b ou c (soustraction)
[a-zA-Z]	Caractères de a à z ou de A à Z compris (intervalle)
[a-d[m-p]]	Caractères de a à d ou de m à p (union). Cette option peut aussi être spécifiée comme [a-dm-p]
[a-z&&[def]]	Caractères de a à z ainsi que d, e ou f (intersection)
[a-z&&[^bc]]	Caractères de a à z à l'exception de b et c (soustraction). Cette option peut aussi être spécifiée comme [ad-z]
[a-z&&[^m-p]]	Caractères de a à z, mais pas de m à p (soustraction). Cette option peut aussi être spécifiée comme [a-lq-z]

Classes de caractères prédéfinies

Classes de caractères prédéfinies	Correspondances
.	Tout caractère (peut correspondre ou non aux terminaisons de ligne)
\d	Tout chiffre : [0-9]
\D	Un caractère non numérique : [^0-9]
\s	Un espace blanc : [\t\n\x0B\f\r]

Classes de caractères prédéfinies	Correspondances
\S	Un espace non blanc : [^\s]
\w	Un mot : [a-zA-Z_0-9]
\W	Un caractère autre qu'un mot : [^\w]

Correspondances de limites

Correspondances de limites	Correspondances
^	Le début d'une ligne
\$	La fin d'une ligne
\b	Une limite de mot
\B	Une limite autre que celle d'un mot
\A	Le début de la saisie
\Z	La fin de la saisie à l'exception de la terminaison finale éventuelle
\z	La fin de la saisie

Référence du langage CLEM

Présentation de la référence CLEM

Cette section décrit le Langage de contrôle (CLEM) pour la Manipulation d'Expressions, puissant outil permettant d'analyser et de manipuler les données utilisées dans les flux IBM® SPSS® Modeler. Vous pouvez utiliser le langage CLEM au sein des noeuds pour exécuter entre autres les tâches suivantes : évaluation des conditions, calcul de valeurs, insertion de données dans les rapports. [Pour plus d'informations, reportez-vous à la section A propos de CLEM dans le chapitre 7 sur p. 99.](#)

Un sous-ensemble du langage CLEM peut également être utilisé lorsque vous générez des scripts dans l'interface utilisateur. Ceci vous permet d'effectuer la plupart des opérations de manipulation des données traitées de façon automatisée. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.*](#)

Les expressions CLEM se composent de valeurs, de noms de champs, d'opérateurs et de fonctions. En utilisant une syntaxe correcte, vous pouvez créer des opérations de données nombreuses et puissantes. [Pour plus d'informations, reportez-vous à la section CLEM Exemples dans le chapitre 7 sur p. 102.](#)

Types de données CLEM

Les types de données CLEM peuvent se composer des éléments suivants :

- Entiers
- Réels
- Caractères
- Chaînes
- Listes
- Champs
- Date/Heure

Règles des guillemets

Même si IBM® SPSS® Modeler est flexible lorsque vous déterminez des champs, des valeurs, des chaînes et des paramètres utilisés dans une expression CLEM, les règles générales suivantes fournissent la liste des conseils à suivre lorsque vous créez des expressions :

- Chaînes : utilisez toujours des guillemets doubles lorsque vous écrivez des chaînes, comme "Type 2". Vous pouvez également utiliser des guillemets simples mais cela risque de créer une confusion avec les champs placés entre guillemets.

- Champs : utilisez des guillemets simples, uniquement lorsque cela est rendu nécessaire par la présence d'espaces ou d'autres caractères spéciaux, comme 'Order Number'. les champs placés entre guillemets mais non définis dans l'ensemble de données seront lus comme des chaînes.
- Paramètres : utilisez toujours des guillemets simples lorsque vous employez des paramètres, comme '\$P-threshold'.
- Caractères : utilisez toujours des guillemets simples inversés (`), comme dans stripchar(`d`, "drugA").

Pour plus d'informations, reportez-vous à la section Valeurs et types de données dans le chapitre 7 sur p. 104. En outre, ces règles sont détaillées dans les rubriques suivantes.

Entiers

Les entiers sont représentés par une séquence de chiffres décimaux. (Facultatif) Vous pouvez placer un signe moins (-) devant un entier afin d'obtenir un nombre négatif—par exemple, 1234, 999, -77.

Le langage CLEM permet de manipuler les entiers de précision arbitraire. La taille maximale de l'entier varie selon votre plate-forme. Si les valeurs sont trop élevées pour être affichées dans un champ de nombre entier, changez le type de champ en Real pour restaurer la valeur.

Réels

Un nombre *Réel* est un nombre à virgule flottante. Les réels sont représentés par un ou plusieurs chiffres, suivis d'un séparateur décimal, suivi d'un ou de plusieurs autres chiffres. Les réels CLEM sont maintenus en double précision.

(Facultatif) Vous pouvez placer un signe moins (-) devant le réel afin d'obtenir un nombre négatif—par exemple, 1.234, 0.999, -77.001. Utilisez la syntaxe <nombre> e <exponent> pour exprimer un nombre réel en notation exponentielle, —par exemple, 1234.0e5, 1.7e-2. Lorsque l'application IBM® SPSS® Modeler lit des chaînes numériques en provenance de fichiers et les convertit automatiquement en nombres, les nombres sans chiffre avant ou après le séparateur décimal sont acceptés—par exemple, 999. ou.11. Cependant, ces formes sont incorrectes dans les expressions CLEM.

Remarque : Lors du référencement de nombres réels dans CLEM, un point doit être utilisé comme séparateur décimal, quels que soient les paramètres pour le flux en cours ou les paramètres régionaux. Par exemple précisez

Na > 0.6

plutôt que :

Na > 0,6

Cela s'applique même lorsque le symbole décimal choisi dans la boîte de dialogue Propriétés du flux est la virgule. En outre, cela est conforme à l'instruction générale selon laquelle la syntaxe de code doit être indépendante des conventions et paramètres régionaux.

Caractères

Des caractères (habituellement désignés par CHAR) sont généralement employés avec une expression CLEM pour réaliser des tests sur les chaînes. Ainsi, vous pouvez utiliser la fonction `isuppercode` pour déterminer si le premier caractère d'une chaîne est une majuscule. L'expression CLEM suivante se sert d'un caractère pour indiquer que le test doit être effectué sur le premier caractère de la chaîne :

```
isuppercode(subscrs(1, "MyString"))
```

Pour exprimer le code (et non l'emplacement) d'un caractère particulier dans une expression CLEM, utilisez des guillemets simples inversés : ``<caractère>``— par exemple, ``A``, ``Z``.

Remarque : Il n'existe pas de type de stockage CHAR pour les champs. Par conséquent, si un champ est calculé ou rempli avec une expression générant un caractère CHAR, ce résultat est converti en chaîne.

Chaînes

En général, vous devez placer les chaînes entre guillemets doubles. Les exemples de chaînes sont "c35product2" et "referrerID". Pour signaler des caractères spéciaux dans une chaîne, utilisez une barre oblique inverse—, par exemple, "\\$65443". (Pour indiquer une barre oblique inversée, utilisez deux barres obliques inversées, \\.). Vous pouvez placer une chaîne entre guillemets simples mais elle risque alors d'être confondue avec un champ placé entre guillemets ('referrerID'). [Pour plus d'informations, reportez-vous à la section Fonctions sur chaînes sur p. 139.](#)

Listes

Une liste est une séquence d'éléments triés, qui peuvent être de nature différente. Les listes sont placées entre crochets ([]), Les exemples de listes sont [1 2 4 16] et ["abc" "def"]. Les listes ne sont pas utilisées pour les valeurs des champs IBM® SPSS® Modeler. Elles sont utilisées pour fournir des arguments aux fonctions, comme `member` et `oneof`.

Champs

Dans les expressions CLEM, les noms autres que les noms de fonction sont considérés comme des noms de champ. Vous pouvez les écrire aussi simplement que `Power`, `val27`, `state_flag`, etc., mais si le nom commence par un chiffre ou contient des caractères non alphabétiques, comme des espaces (à l'exception du tiret bas), placez le nom entre guillemets simples—par exemple, `'Power Increase'`, `'2nd answer'`, `'#101'`, `'$P-NextField'`.

Remarque : les champs placés entre guillemets mais non définis dans l'ensemble de données seront lus comme des chaînes.

Dates

Les calculs de date reposent sur la date de référence spécifiée dans la boîte de dialogue des propriétés du flux. La date de référence par défaut est le 1er janvier 1900. [Pour plus d'informations, reportez-vous à la section Définition des options de flux dans le chapitre 5 sur p. 51.](#)

Le langage CLEM prend en charge les formats de date suivants.

Format	Exemples
DDMMYY	150163
MMDDYY	011563
YYMMDD	630115
YYYYMMDD	19630115
YYYYDDD	Année à quatre chiffres, suivie d'une valeur à trois chiffres représentant le jour de l'année— (par exemple, 2000032 correspond au 32e jour de l'année 2000, soit le 1er février 2000.
DAY	Jour de la semaine utilisé dans les paramètres régionaux actuels— (par exemple, Monday, Tuesday, etc. en anglais)
MONTH	Mois utilisé dans les paramètres régionaux actuels—par exemple, January, February,
DD/MM/YY	15/01/63
DD/MM/YYYY	15/01/1963
MM/DD/YY	01/15/63
MM/DD/YYYY	01/15/1963
DD-MM-YY	15-01-63
DD-MM-YYYY	15-01-1963
MM-DD-YY	01-15-63
MM-DD-YYYY	01-15-1963
DD.MM.YY	15.01.63
DD.MM.YYYY	15.01.1963
MM.DD.YY	01.15.63
MM.DD.YYYY	01.15.1963
DD-MON-YY	15-JAN-63, 15-jan-63, 15-Jan-63
DD/MON/YY	15/JAN/63, 15/jan/63, 15/Jan/63
DD.MON.YY	15.JAN.63, 15.jan.63, 15.Jan.63
DD-MON-YYYY	15-JAN-1963, 15-jan-1963, 15-Jan-1963
DD/MON/YYYY	15/JAN/1963, 15/jan/1963, 15/Jan/1963
DD.MON.YYYY	15.JAN.1963, 15.jan.1963, 15.Jan.1963
MON YYYY	Jan 2004

Format	Exemples
q Q YYYY	Date représentée par un chiffre (1–4) correspondant au trimestre, suivi de la lettre <i>Q</i> et d'une année à quatre chiffres— (par exemple, le 25 décembre 2004 est représenté par 4 Q 2004).
ww WK YYYY	Valeur à deux chiffres représentant la semaine de l'année, suivie des lettres <i>SM</i> , puis d'une année à quatre chiffres. La semaine de l'année est calculée en supposant que le premier jour de la semaine est un lundi et que la première semaine comprend au moins un jour.

Heure

Le langage CLEM prend en charge les formats d'heure suivants.

Format	Exemples
HHMMSS	120112, 010101, 221212
HHMM	1223, 0745, 2207
MMSS	5558, 0100
HH:MM:SS	12:01:12, 01:01:01, 22:12:12
HH:MM	12:23, 07:45, 22:07
MM:SS	55:58, 01:00
(H)H:(M)M:(S)S	12:1:12, 1:1:1, 22:12:12
(H)H:(M)M	12:23, 7:45, 22:7
(M)M:(S)S	55:58, 1:0
HH.MM.SS	12.01.12, 01.01.01, 22.12.12
HH.MM	12.23, 07.45, 22.07
MM.SS	55.58, 01.00
(H)H.(M)M.(S)S	12.1.12, 1.1.1, 22.12.12
(H)H.(M)M	12.23, 7.45, 22.7
(M)M.(S)S	55.58, 1.0

Opérateurs CLEM

Les opérateurs suivants sont disponibles.

Opération	Commentaires	Priorité (voir section suivante)
or	Utilisé entre deux expressions CLEM. Renvoie la valeur true (vrai) si l'une des deux, ou les deux expressions présentent la valeur true (vrai).	10
and	Utilisé entre deux expressions CLEM. Renvoie la valeur true (vrai) si les deux expressions présentent la valeur true (vrai).	9

Opération	Commentaires	Priorité (voir section suivante)
=	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 est égal à ITEM2.	7
==	Identique à =.	7
/=	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 <i>n'est pas</i> égal à ITEM2.	7
/==	Identique à /=.	7
>	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 est supérieur à ITEM2.	6
>=	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 est supérieur ou égal à ITEM2.	6
<	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 est inférieur à ITEM2.	6
<=	Utilisé entre deux éléments comparables. Renvoie la valeur true (vrai) si ITEM1 est inférieur ou égal à ITEM2.	6
&&=_0	Utilisé entre deux entiers. Equivaut à l'expression booléenne INT1 && INT2 = 0.	6
&&/=_0	Utilisé entre deux entiers. Equivaut à l'expression booléenne INT1 && INT2 /= 0.	6
+	Additionne deux nombres : NUM1 + NUM2.	5
><	Concatène deux chaînes, par exemple, STRING1 >< STRING2.	5
-	Soustrait un nombre à un autre : NUM1 - NUM2. Peut être également utilisé devant un nombre : - NUM.	5
*	Utilisé pour multiplier deux chiffres : NUM1 * NUM2.	4
&&	Utilisé entre deux entiers. Le résultat de l'opération est l'opérateur sur les bits « ET » des entiers INT1 et INT2.	4
&&~	Utilisé entre deux entiers. Le résultat de l'opération est l'opérateur sur les bits « ET » de l'entier INT1 et le complément sur les bits de l'entier INT2.	4
	Utilisé entre deux entiers. Le résultat de l'opération est l'opérateur sur les bits 'OU inclusif' des entiers INT1 et INT2.	4

Opération	Commentaires	Priorité (voir section suivante)
~~	Utilisé devant un entier. Produit le complément sur les bits de INT.	4
&	Utilisé entre deux entiers. Le résultat de l'opération est l'opérateur sur les bits 'OU exclusif' des entiers INT1 et INT2.	4
INT1 << N	Utilisé entre deux entiers. Produit le motif binaire de l'entier INT décalé vers la gauche de N positions.	4
INT1 >> N	Utilisé entre deux entiers. Produit le motif binaire de l'entier INT décalé vers la droite de N positions.	4
/	Utilisé pour diviser un nombre par un autre : NUM1 / NUM2.	4
**	Utilisé entre deux nombres : BASE ** POWER. Renvoie la valeur de base BASE élevée à la puissance POWER.	3
rem	Utilisé entre deux entiers : INT1 rem INT2. Renvoie le reste, INT1 - (INT1 div INT2) * INT2.	2
div	Utilisé entre deux entiers : INT1 div INT2. Effectue la division des entiers.	2

Priorité des opérateurs

La priorité des opérateurs détermine l'ordre dans lequel les opérations des expressions complexes sont traitées, notamment dans les expressions sans parenthèses avec plusieurs opérateurs infixes. Par exemple,

$3 + 4 * 5$

est analysé en tant que $3 + (4 * 5)$ plutôt que $(3 + 4) * 5$ parce que les priorités relatives indiquent que * doit être analysé avant +. A chaque opérateur du langage CLEM est affectée une valeur de priorité. Plus cette valeur est faible, plus l'opérateur est prioritaire. Il sera donc traité avant ceux ayant un niveau de priorité plus élevé.

Informations de référence sur les fonctions

Les fonctions CLEM suivantes sont disponibles lorsque vous travaillez avec des données dans IBM® SPSS® Modeler. Vous entrez ces fonctions sous forme de code dans un grand nombre de boîtes de dialogue, comme celles des noeuds Calculer et Binariser. Vous pouvez également utiliser

le Générateur de formules pour créer des expressions CLEM valides sans avoir à mémoriser des listes de fonctions ou des noms de champs.

Type de fonction	Description
Informations	Servent à obtenir un aperçu des valeurs de champ. Par exemple, la fonction <code>is_string</code> renvoie une valeur true (vrai) pour tous les enregistrements de type chaîne.
Conversion	Servent à créer des champs ou à convertir un type de stockage. Par exemple, la fonction <code>to_timestamp</code> convertit le champ sélectionné en un horodatage.
Comparaison	Servent à comparer des valeurs de champ, entre elles ou par rapport à la chaîne spécifiée. Par exemple, la fonction <code><=</code> permet de savoir si la valeur d'un champ est inférieure ou égale.
Logique	Servent à exécuter des opérations logiques, comme <code>if</code> , <code>then</code> , <code>else</code> .
Numérique	Servent à effectuer des calculs numériques, comme le calcul du logarithme naturel de valeurs de champ.
Trigonométrique	Servent à effectuer des calculs trigonométriques, comme le calcul du cosinus inverse d'un angle.
Probabilité	Renvoie des probabilités basées sur différentes distributions, comme la probabilité qu'une valeur de la distribution <i>t</i> de Student soit inférieure à une valeur spécifique.
Sur les bits	Servent à manipuler des nombres entiers comme éléments binaires.
Aléatoire	Servent à sélectionner des éléments ou à générer des nombres .
Chaîne	Servent à exécuter un grand nombre d'opérations sur des chaînes, comme <code>stripchar</code> , qui permet de supprimer le caractère spécifié.
SoundEx	Sert à trouver des chaînes lorsque l'orthographe exacte est inconnue, sur la base d'hypothèses phonétiques (modalités de prononciation de certaines lettres).
La date et l'heure	Servent à exécuter un grand nombre d'opérations sur les champs de date, d'heure et d'horodatage.
Séquence	Servent à obtenir un aperçu de la séquence d'enregistrements d'un ensemble de données ou à exécuter des opérations basées sur cette séquence.
Global	Servent à accéder aux valeurs globales créées par un noeud V. globales (Valeurs globales). Par exemple, <code>@MEAN</code> sert à faire référence à la moyenne de toutes les valeurs d'un champ dans tout l'ensemble de données.
Nul et non renseigné	Servent à atteindre, à marquer et souvent à remplir des blancs définis par l'utilisateur ou des valeurs absentes du système. Par exemple, <code>@BLANK(FIELD)</code> permet d'afficher une valeur true (vrai) pour les enregistrements qui comportent des blancs.
Champs spéciaux	Servent à signaler les champs spéciaux en cours d'examen. Par exemple, <code>@FIELD</code> est utilisé lors du calcul de plusieurs champs.

Conventions pour la description des fonctions

Les conventions suivantes sont utilisées dans ce guide pour faire référence aux éléments d'une fonction.

Convention	Description
<i>BOOL</i>	Un booléen, comme true ou false (vrai, faux).
<i>NUM, NUM1, NUM2</i>	Tout nombre.
<i>REAL, REAL1, REAL2</i>	Nombre réel, comme 1.234 ou -77.01.
<i>INT, INT1, INT2</i>	Un entier, comme 1 ou -77.
<i>CHAR</i>	Code de caractère, comme 'A'.
<i>STRING</i>	Chaîne, comme "referrerID".
<i>LIST</i>	Liste d'éléments, comme ["abc" "def"].
<i>ITEM</i>	Un champ, comme Customer ou extract_concept.
<i>DATE</i>	Champ de date, comme start_date, dont les valeurs sont au format DD-MON-YYYY.
<i>TIME</i>	Champ d'heure, comme power_flux, dont les valeurs sont au format HHMMSS.

Les fonctions couvertes dans ce guide sont répertoriées dans un tableau à trois colonnes : la première colonne indique le nom de la fonction, la deuxième le type de résultat (entier, chaîne, etc.) et la troisième, une description (si elle est disponible). Par exemple, la description suivante est la description d'une fonction rem.

Fonction	Résultat	Description
INT1 rem INT2	<i>Nombre</i>	Renvoie le reste de la division de l'entier <i>INT1</i> par l'entier <i>INT2</i> . Par exemple, $INT1 - (INT1 \text{ div } INT2) * INT2$

Les informations relatives aux conventions d'utilisation, comme la manière de répertorier des éléments ou de spécifier les caractères d'une fonction, sont décrites dans une autre rubrique. [Pour plus d'informations, reportez-vous à la section Types de données CLEM sur p. 122.](#)

Fonctions d'informations

Les fonctions d'informations sont utilisées pour obtenir un aperçu des valeurs d'un champ particulier. Elles sont généralement utilisées pour calculer des champs booléens. Par exemple, vous pouvez utiliser la fonction @BLANK pour créer un champ booléen indiquant les enregistrements dont les valeurs sont vides pour le champ sélectionné. De même, vous pouvez

vérifier le type de stockage d'un champ à l'aide de l'une des fonctions de type de stockage, comme `is_string`.

Fonction	Résultat	Description
@BLANK(FIELD)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements dont les valeurs sont non renseignées, d'après les règles de traitement des blancs définies dans un noeud Typers ou un noeud source en amont (onglet Types). Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
@NULL(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements dont les valeurs ne sont pas définies. Les valeurs non définies sont des valeurs système nulles, affichées dans IBM® SPSS® Modeler sous la forme \$null\$. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
is_date(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type date.
is_datetime(ITEM)	Booléen	Renvoie la valeur vraie (True) pour tous les enregistrements de type date, heure ou horodatage.
is_integer(ITEM)	Booléen	Renvoie la valeur vraie (True) pour tous les enregistrements de type entier.
is_number(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type nombre.
is_real(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type réel.
is_string(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type chaîne.
is_time(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type heure.
is_timestamp(ITEM)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements de type horodatage.

Fonctions de conversion

Les fonctions de conversion permettent de créer des champs et de convertir le type de stockage des champs existants. Par exemple, vous pouvez former de nouvelles chaînes en joignant ou en séparant des chaînes existantes. Pour joindre des chaînes, utilisez l'opérateur `><`. Par exemple, si le champ `Site` a la valeur "BRAMLEY", alors `"xx" >< Site` renvoie "xxBRAMLEY". Le résultat de `><` est toujours une chaîne, même si les arguments n'en sont pas. Par conséquent, si le champ `V1` est 3 et le champ `V2` est 5, alors `V1 >< V2` renvoie "35" (une chaîne pas un chiffre).

Les fonctions de conversion (et toutes les autres fonctions qui nécessitent un type spécifique d'entrée, par exemple une valeur de date ou d'heure) dépendent des formats actuels indiqués dans la boîte de dialogue des options de flux. Par exemple, si vous souhaitez convertir un champ de type chaîne avec des valeurs *Jan 2003*, *Fév 2003*, etc., sélectionnez `MOIS AAAA` comme format de

date par défaut pour le flux. [Pour plus d'informations, reportez-vous à la section Définition des options de flux dans le chapitre 5 sur p. 51.](#)

Fonction	Résultat	Description
ITEM1 << ITEM2	Chaîne	Concatène les valeurs pour deux champs et renvoie la chaîne résultante sous la forme <i>ITEM1ITEM2</i> .
to_integer(ITEM)	Entier	Convertit le stockage du champ spécifié en un entier.
to_real(ITEM)	Réel	Convertit le stockage du champ spécifié en un réel.
to_number(ITEM)	Nombre	Convertit le stockage du champ spécifié en un nombre.
to_string(ITEM)	Chaîne	Convertit le stockage du champ spécifié en une chaîne.
to_time(ITEM)	Heure	Convertit le stockage du champ spécifié en une heure.
to_date(ITEM)	Date	Convertit le stockage du champ spécifié en une date.
to_timestamp(ITEM)	Horodatage	Convertit le stockage du champ spécifié en un horodatage.
to_datetime(ITEM)	Valeurdate	Convertit le stockage du champ spécifié en une valeur d'heure, de date ou d'horodatage.
datetime_date(ITEM)	Date	Renvoie la valeur de date pour un <i>nombre</i> , une <i>chaîne</i> , ou un <i>horodatage</i> . Veuillez noter que cette fonction est la seule fonction vous permettant de reconverter un nombre (en secondes) en date. Si ITEM est une chaîne, crée une date en analysant une chaîne au format de date sélectionné. Pour que cette fonction réussisse, vous devez spécifier un format de date correct dans la boîte de dialogue des propriétés du flux. Si ITEM est un nombre, il est interprété comme un nombre de secondes écoulées depuis la date de départ (ou période). Les fractions de jour sont supprimées. Si ITEM est un horodatage, la partie date de cet horodatage est renvoyée. Si ITEM est une date, il est renvoyé sans modification.

Fonctions de comparaison

Les fonctions de comparaison sont utilisées pour comparer des valeurs de champ, entre elles ou par rapport à la chaîne spécifiée. Par exemple, vous pouvez vérifier l'égalité de deux chaînes en utilisant =. Voici un exemple de vérification de l'égalité des chaînes : `Class = "class 1"`.

Lors d'une comparaison de nombres, *supérieur à* signifie que le nombre est plus proche de l'infini positif et *inférieur à* signifie qu'il se rapproche de l'infini négatif. Autrement dit, tous les nombres négatifs sont inférieurs à tous les nombres positifs.

Fonction	Résultat	Description
count_equal(ITEM1, LIST)	Entier	Renvoie le nombre de valeurs, issues d'une liste de champs, égales à <i>ITEM1</i> ou une valeur nulle si <i>ITEM1</i> est nul. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
count_greater_than(ITEM1, LIST)	Entier	Renvoie le nombre de valeurs, issues d'une liste de champs, supérieures à <i>ITEM1</i> ou une valeur nulle si <i>ITEM1</i> est nul.
count_less_than(ITEM1, LIST)	Entier	Renvoie le nombre de valeurs, issues d'une liste de champs, inférieures à <i>ITEM1</i> ou une valeur nulle si <i>ITEM1</i> est nul.
count_not_equal(ITEM1, LIST)	Entier	Renvoie le nombre de valeurs, issues d'une liste de champs, différentes de <i>ITEM1</i> ou une valeur nulle si <i>ITEM1</i> est nul.

Fonction	Résultat	Description
count_nulls(LIST)	Entier	Renvoie le nombre de valeurs nulles issues d'une liste de champs.
count_non_nulls(LIST)	Entier	Renvoie le nombre de valeurs non nulles issues d'une liste de champs.
date_before(DATE1, DATE2)	Booléen	Utilisée pour vérifier l'ordre des valeurs de date. Renvoie une valeur true (vrai) si DATE1 est antérieure à DATE2.
first_index(ITEM, LIST)	Entier	Renvoie l'index du premier champ contenant ITEM depuis une LISTE de champs ou 0 si aucune valeur n'est trouvée. Pris en charge uniquement pour les types de chaînes, d'entiers et de réels. Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples dans le chapitre 7 sur p. 112.
first_non_null(LIST)	Tout	Renvoie la première valeur non nulle de la liste de champs fournie. Tous les types de stockage sont pris en charge.
first_non_null_index(LIST)	Entier	Renvoie l'index du premier champ dans la LISTE spécifiée contenant une valeur non nulle ou 0 si toutes les valeurs sont nulles. Tous les types de stockage sont pris en charge.
ITEM1 = ITEM2	Booléen	Renvoie la valeur true (vrai) pour les enregistrements où ITEM1 est égal à ITEM2.
ITEM1 /= ITEM2	Booléen	Renvoie la valeur true (vrai) si les deux chaînes sont différentes ou renvoie 0 si elles sont identiques.
ITEM1 < ITEM2	Booléen	Renvoie la valeur true (vrai) pour les enregistrements où ITEM1 est inférieur à ITEM2.
ITEM1 <= ITEM2	Booléen	Renvoie la valeur true (vrai) pour les enregistrements où ITEM1 est inférieur ou égal à ITEM2.
ITEM1 > ITEM2	Booléen	Renvoie la valeur true (vrai) pour les enregistrements où ITEM1 est supérieur à ITEM2.
ITEM1 >= ITEM2	Booléen	Renvoie la valeur true (vrai) pour les enregistrements où ITEM1 est supérieur ou égal à ITEM2.
last_index(ITEM, LIST)	Entier	Renvoie l'index du dernier champ contenant ITEM depuis une LISTE de champs ou 0 si aucune valeur n'est trouvée. Pris en charge uniquement pour les types de chaînes, d'entiers et de réels. Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples dans le chapitre 7 sur p. 112.
last_non_null(LIST)	Tout	Renvoie la dernière valeur non nulle de la liste de champs fournie. Tous les types de stockage sont pris en charge.
last_non_null_index(LIST)	Entier	Renvoie l'index du dernier champ dans la LISTE spécifiée contenant une valeur non nulle ou 0 si toutes les valeurs sont nulles. Tous les types de stockage sont pris en charge.
max(ITEM1, ITEM2)	Tout	Renvoie le plus grand des deux éléments ITEM1 et ITEM2.
max_index(LIST)	Entier	Renvoie l'index du champ contenant la valeur maximale d'une liste de champs numériques ou 0 si toutes les valeurs de champ sont nulles. Par exemple, si le troisième champ de la liste contient la valeur maximale, la valeur 3 de l'index est renvoyée. Si plusieurs champs contiennent la valeur maximale, le premier de la liste (le plus à gauche) est renvoyé. Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples dans le chapitre 7 sur p. 112.

Fonction	Résultat	Description
max_n(LIST)	Nombre	Renvoie la valeur maximale d'une liste de champs numériques ou une valeur nulle si toutes les valeurs de champ sont nulles. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
member(ITEM, LIST)	Booléen	Renvoie la valeur true (vrai) si <i>ITEM</i> est un membre de la liste (<i>LIST</i>) spécifiée. Sinon, renvoie une valeur fausse (False). Il est également possible d'indiquer une liste de noms de champ. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
min(ITEM1, ITEM2)	Tout	Renvoie le plus petit des deux éléments <i>ITEM1</i> et <i>ITEM2</i> .
min_index(LIST)	Entier	Renvoie l'index du champ contenant la valeur minimale d'une liste de champs numériques ou 0 si toutes les valeurs de champ sont nulles. Par exemple, si le troisième champ de la liste contient la valeur minimale, la valeur 3 de l'index est renvoyée. Si plusieurs champs contiennent la valeur minimale, le premier de la liste (le plus à gauche) est renvoyé. Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples dans le chapitre 7 sur p. 112.
min_n(LIST)	Nombre	Renvoie la valeur minimale d'une liste de champs numériques ou une valeur nulle si toutes les valeurs de champ sont nulles.
time_before(TIME1, TIME2)	Booléen	Utilisée pour vérifier l'ordre des valeurs d'heure. Renvoie une valeur true (vrai) si <i>TIME1</i> est antérieure à <i>TIME2</i> .
value_at(INT, LIST)		Renvoie la valeur de décalage INT ou NULL de chaque champ de la liste si le décalage est en-dehors de l'intervalle de valeurs valides (c'est-à-dire, inférieur à 1 ou supérieur au nombre de champs de la liste). Tous les types de stockage sont pris en charge.

Fonctions logiques

Vous pouvez utiliser des expressions CLEM pour effectuer des opérations logiques.

Fonction	Résultat	Description
COND1 and COND2	Booléen	Cette opération est une conjonction logique et renvoie une valeur true (vrai) si <i>COND1</i> et <i>COND2</i> sont toutes les deux vraies. Si <i>COND1</i> est fausse, <i>COND2</i> n'est pas évaluée ; il est alors possible d'avoir des conjonctions où <i>COND1</i> teste d'abord si une opération dans <i>COND2</i> est autorisée. Par exemple, <code>length(Label) >=6</code> et <code>Label(6) = 'x'</code> .
COND1 or COND2	Booléen	Cette opération est une disjonction logique (inclusive) et renvoie une valeur true (vrai) si l'une des deux conditions <i>COND1</i> or <i>COND2</i> est vraie ou si es deux le sont. Si <i>COND1</i> est vraie, <i>COND2</i> n'est pas évaluée.
not(COND)	Booléen	Cette opération est une négation logique et renvoie une valeur true (vrai) si <i>COND</i> est fausse. Sinon, cette opération renvoie la valeur 0.

Fonction	Résultat	Description
if COND then EXPR1 else EXPR2 endif	Tout	Cette opération est une évaluation conditionnelle. Si <i>COND</i> est vraie, cette opération renvoie le résultat de l'évaluation de l'expression <i>EXPR1</i> . Sinon, elle renvoie le résultat de l'évaluation de l'expression <i>EXPR2</i> .
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Tout	Cette opération est une évaluation conditionnelle à plusieurs branches. Si <i>COND1</i> est vraie, cette opération renvoie le résultat de l'évaluation de l'expression <i>EXPR1</i> . En revanche, si <i>COND2</i> est vraie, l'opération renvoie le résultat de l'évaluation de l'expression <i>EXPR2</i> . Sinon, le résultat de l'évaluation de l'expression <i>EXPR_N</i> est renvoyé.

Fonctions numériques

CLEM contient un certain nombre de fonctions numériques courantes.

Fonction	Résultat	Description
-NUM	Nombre	Utilisée pour inverser <i>NUM</i> . Renvoie le même nombre, avec le signe opposé .
NUM1 + NUM2	Nombre	Renvoie la somme de <i>NUM1</i> et <i>NUM2</i> .
code -NUM2	Nombre	Renvoie la valeur <i>NUM2</i> , soustraite de <i>NUM1</i> .
NUM1 * NUM2	Nombre	Renvoie la valeur <i>NUM1</i> multipliée par <i>NUM2</i> .
NUM1 / NUM2	Nombre	Renvoie la valeur <i>NUM1</i> divisée par <i>NUM2</i> .
INT1 div INT2	Nombre	Utilisée pour la division en nombres entiers. Renvoie la valeur de l'entier <i>INT1</i> divisé par l'entier <i>INT2</i> .
INT1 rem INT2	Nombre	Renvoie le reste de la division de l'entier <i>INT1</i> par l'entier <i>INT2</i> . Par exemple, $INT1 - (INT1 \text{ div } INT2) * INT2$
INT1 mod INT2	Nombre	Cette fonction a été remplacée. Utilisez la fonction <i>rem</i> à la place.
BASE ** POWER	Nombre	Renvoie la valeur <i>BASE</i> élevée à la puissance <i>POWER</i> . Les deux valeurs peuvent être tout nombre (en revanche, <i>BASE</i> ne doit pas avoir la valeur zéro si <i>POWER</i> est elle-même égale à une valeur nulle autre que l'entier 0). Si <i>POWER</i> est un entier, le calcul est effectué en multipliant successivement les puissances de <i>BASE</i> . Ainsi, si <i>BASE</i> est un entier, le résultat l'est également. Si <i>POWER</i> est l'entier 0, le résultat est toujours un 1 du même type que <i>BASE</i> . Sinon, si <i>POWER</i> n'est pas un entier, le résultat est calculé selon la formule $\exp(POWER * \log(BASE))$.
abs(NUM)	Nombre	Renvoie la valeur absolue de <i>NUM</i> , qui est toujours un nombre du même type.
exp(NUM)	Réel	Renvoie <i>e</i> élevé à la puissance <i>NUM</i> , <i>e</i> étant la base des logarithmes naturels.
fracof(NUM)	Réel	Renvoie la partie fractionnelle de <i>NUM</i> , définie selon la formule $NUM - \text{intof}(NUM)$.
intof(NUM)	Entier	Tronque son argument pour obtenir un entier. Il renvoie l'entier du même signe que <i>NUM</i> ayant la valeur absolue la plus élevée, comme $\text{abs}(INT) \leq \text{abs}(NUM)$.
log(NUM)	Réel	Renvoie le logarithme naturel (base <i>e</i>) de <i>NUM</i> , qui ne doit en aucun cas avoir la valeur zéro.

Fonction	Résultat	Description
log10(NUM)	Réel	Renvoie le logarithme décimal de <i>NUM</i> qui ne doit en aucun cas avoir la valeur zéro. Cette fonction est définie selon la formule $\log(\text{NUM}) / \log(10)$.
negate(NUM)	Nombre	Utilisée pour inverser <i>NUM</i> . Renvoie le même nombre, avec le signe opposé.
round(NUM)	Entier	Utilisé pour arrondir <i>NUM</i> à un entier en prenant $\text{intof}(\text{NUM}+0.5)$ si <i>NUM</i> est positif $\text{intof}(\text{NUM}-0.5)$ si <i>NUM</i> est négatif.
sign(NUM)	Nombre	Permet de déterminer le signe de <i>NUM</i> . Cette opération renvoie -1, 0, ou 1 si <i>NUM</i> est un entier. Si <i>NUM</i> est un nombre réel, elle renvoie -1,0, 0,0 ou 1,0 selon que <i>NUM</i> est négatif, positif ou égal à zéro.
sqrt(NUM)	Réel	Renvoie la racine carrée de <i>NUM</i> . <i>NUM</i> doit être positif.
sum_n(LIST)	Nombre	Renvoie la somme des valeurs d'une liste de champs numériques ou une valeur nulle si toutes les valeurs de champ sont nulles. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
mean_n(LIST)	Nombre	Renvoie la valeur moyenne d'une liste de champs numériques ou une valeur nulle si toutes les valeurs de champ sont nulles.
sdev_n(LIST)	Nombre	Renvoie l'écart-type d'une liste de champs numériques ou une valeur nulle si toutes les valeurs de champ sont nulles.

Fonctions trigonométriques

Toutes les fonctions de cette section utilisent un angle comme argument ou en renvoient un comme résultat. Dans les deux cas, les unités d'angle (radians ou degrés) sont contrôlées par les paramètres de l'option de flux appropriée.

Fonction	Résultat	Description
arccos(NUM)	Réel	Calcule le cosinus inverse de l'angle spécifié.
arccosh(NUM)	Réel	Calcule le sinus inverse hyperbolique de l'angle spécifié.
arcsin(NUM)	Réel	Calcule le sinus inverse de l'angle spécifié.
arcsinh(NUM)	Réel	Calcule le sinus inverse hyperbolique de l'angle spécifié.
arctan(NUM)	Réel	Calcule la tangente inverse de l'angle spécifié.
arctan2(NUM_Y, NUM_X)	Réel	Calcule la tangente inverse de $\text{NUM_Y} / \text{NUM_X}$ et utilise les signes des deux nombres pour calculer des informations du cadran. Le résultat est un réel dans l'intervalle $-\pi < \text{ANGLE} \leq \pi$ (radians) – $-180 < \text{ANGLE} \leq 180$ (degrees)
arctanh(NUM)	Réel	Calcule la tangente inverse hyperbolique de l'angle spécifié.
cos(NUM)	Réel	Calcule le cosinus de l'angle spécifié.
cosh(NUM)	Réel	Calcule le cosinus hyperbolique de l'angle spécifié.
pi	Réel	Cette constante est la meilleure approximation réelle de π .
sin(NUM)	Réel	Calcule le sinus de l'angle spécifié.
sinh(NUM)	Réel	Calcule le sinus hyperbolique de l'angle spécifié.
tan(NUM)	Réel	Calcule la tangente de l'angle spécifié.
tanh(NUM)	Réel	Calcule la tangente hyperbolique de l'angle spécifié.

Fonctions de probabilité

Les fonctions de probabilité renvoient des probabilités basées sur différentes distributions, comme la probabilité qu'une valeur de la distribution t de Student soit inférieure à une valeur spécifique.

Fonction	Résultat	Description
<code>cdf_chisq(NUM, DF)</code>	Réel	Renvoie la probabilité qu'une valeur de la distribution du Chi-deux, avec les degrés de liberté indiqués, soit inférieure au nombre spécifié.
<code>cdf_f(NUM, DF1, DF2)</code>	Réel	Renvoie la probabilité qu'une valeur de la distribution F , avec les degrés de liberté $DF1$ et $DF2$ soit inférieure au nombre spécifié.
<code>cdf_normal(NUM, MEAN, STDDEV)</code>	Réel	Renvoie la probabilité qu'une valeur de la distribution normale, avec la moyenne et l'écart-type indiqués, soit inférieure au nombre spécifié.
<code>cdf_t(NUM, DF)</code>	Réel	Renvoie la probabilité qu'une valeur de la distribution t de Student, avec les degrés de liberté indiqués, soit inférieure au nombre spécifié.

Opérations d'entiers sur les bits

Ces fonctions permettent aux entiers d'être manipulés en tant que profils binaires représentant des valeurs de complément à 2, où la position binaire N a la pondération 2^{*N} . Les bits sont numérotés en commençant par 0. Ces opérations se comportent comme si le bit de signe d'un entier était décalé vers la gauche à l'infini. Ainsi, partout au-delà du bit le plus significatif, un entier positif a le bit 0 et un entier négatif a le bit 1.

Remarque : Il n'est pas possible d'appeler des fonctions sur les bits à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans *Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2*.](#)

Fonction	Résultat	Description
<code>~~ INT1</code>	Entier	Produit le complément sur les bits de l'entier $INT1$. C'est-à-dire qu'il existe un 1 dans le résultat pour chaque position binaire pour laquelle $INT1$ contient un 0. $~~ INT = -(INT + 1)$ est toujours vraie. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans <i>Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2</i>.
<code>INT1 INT2</code>	Entier	Le résultat de cette opération est l'opérateur sur les bits "inclusif ou" des entiers $INT1$ et $INT2$. Autrement dit, un 1 est inséré dans le résultat pour chaque position binaire où existe un 1 dans $INT1$, dans $INT2$ ou dans les deux entiers.
<code>INT1 /& INT2</code>	Entier	Le résultat de cette opération est l'opérateur sur les bits "exclusif ou" des entiers $INT1$ et $INT2$. Autrement dit, un 1 est inséré dans le résultat pour chaque position binaire où existe un 1 soit dans $INT1$, soit dans $INT2$, mais pas dans les deux.

Fonction	Résultat	Description
<code>INT1 && INT2</code>	<i>Entier</i>	Produit l'opération booléenne « ET » sur les bits des entiers <i>INT1</i> et <i>INT2</i> . Cela signifie qu'un 1 est présent dans le résultat pour chaque position binaire où il existe un 1 dans les deux entiers <i>INT1</i> et <i>INT2</i> .
<code>INT1 &&~ INT2</code>	<i>Entier</i>	Produit le booléen « ET » de l'entier <i>INT1</i> et le complément sur les bits de l'entier <i>INT2</i> . Cela signifie qu'un 1 est présent dans le résultat pour chaque position binaire où il existe un 1 dans <i>INT1</i> et un 0 dans <i>INT2</i> . Cette opération est identique à <code>INT1 && (~INT2)</code> . Elle s'avère utile pour l'élimination des bits d' <i>INT1</i> défini dans <i>INT2</i> .
<code>INT << N</code>	<i>Entier</i>	Produit le motif binaire de l'entier <i>INT1</i> , décalé vers la gauche de <i>N</i> positions. Utilisez une valeur <i>N</i> négative pour produire un décalage vers la droite.
<code>INT >> N</code>	<i>Entier</i>	Produit le motif binaire de l'entier <i>INT1</i> , décalé vers la droite de <i>N</i> positions. Utilisez une valeur <i>N</i> négative pour produire un décalage vers la gauche.
<code>INT1 &&=_0 INT2</code>	<i>Booléen</i>	Equivalut à l'expression booléenne <code>INT1 && INT2 != 0</code> , mais est plus efficace.
<code>INT1 &&/=_0 INT2</code>	<i>Booléen</i>	Equivalut à l'expression booléenne <code>INT1 && INT2 == 0</code> , mais est plus efficace.
<code>integer_bitcount(INT)</code>	<i>Entier</i>	Compte le nombre des bits 1 ou 0 dans la représentation de complément à 2 de l'entier <i>INT</i> . Si <i>INT</i> n'est pas négatif, <i>N</i> est le nombre de bits 1. Si <i>INT</i> est négatif, <i>N</i> correspond au nombre de bits 0. En raison de l'extension de signe, il existe un nombre infini de bits 0 dans un entier positif ou de bits 1 dans un entier négatif. C'est toujours le cas que <code>integer_bitcount(INT) = integer_bitcount(-(INT+1))</code> .
<code>integer_leastbit(INT)</code>	<i>Entier</i>	Renvoie la position binaire <i>N</i> du bit le moins significatif défini dans l'entier <i>INT</i> . <i>N</i> est la puissance de 2 la plus élevée par laquelle il est possible de diviser <i>INT</i> sans reste.
<code>integer_length(INT)</code>	<i>Entier</i>	Renvoie la longueur en bits de l'entier <i>INT</i> , sous forme d'entier de complément à 2. C'est-à-dire, <i>N</i> est le plus petit entier comme <code>INT < (1 << N)</code> if <code>INT >= 0</code> <code>INT >= (-1 << N)</code> if <code>INT < 0</code> . Si <i>INT</i> n'est pas négatif, la représentation d' <i>INT</i> sous forme d'entier non signé nécessite un champ d'au moins <i>N</i> bits. Sinon, un minimum de <i>N+1</i> bits est nécessaire pour représenter <i>INT</i> en tant qu'entier signé, quel que soit son signe.
<code>testbit(INT, N)</code>	<i>Booléen</i>	Teste le bit de position <i>N</i> dans l'entier <i>INT</i> et renvoie l'état du bit <i>N</i> sous forme de valeur booléenne, true (vrai) s'il s'agit de 1 et false (faux) pour 0.

Fonctions aléatoires

Les fonctions suivantes sont utilisées pour sélectionner des éléments ou générer des nombres.

Fonction	Résultat	Description
oneof(LIST)	<i>Tout</i>	Renvoie un élément choisi au hasard dans <i>LIST</i> . Vous devez entrer les éléments de liste au format [ITEM1,ITEM2,...,ITEM_N]. Il est également possible d'indiquer une liste de noms de champ. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
random(NUM)	<i>Nombre</i>	Renvoie un nombre aléatoire distribué uniformément et du même type (<i>INT</i> ou <i>REAL</i>), de 1 à <i>NUM</i> . Si vous utilisez un entier, seuls des entiers sont renvoyés. Si vous utilisez un nombre réel (décimal), des nombres réels sont renvoyés (précision décimale déterminée par les options de flux). Le plus grand nombre aléatoire qui peut être renvoyé par la fonction est égal à <i>NUM</i> .
random0(NUM)	<i>Nombre</i>	Possède les mêmes propriétés que <i>random(NUM)</i> , mais commence à 0. Le plus grand nombre aléatoire renvoyé par la fonction sera toujours différent de <i>X</i> .

Fonctions sur chaînes

Dans le langage CLEM, vous pouvez effectuer les opérations suivantes avec des chaînes :

- Comparer des chaînes
- Créer des chaînes
- Accéder à des caractères

Dans une expression CLEM, une chaîne est une suite de caractères insérée entre guillemets doubles ("string quotes"). Tout caractère alphanumérique unique peut être un caractère (CHAR). Ils sont déclarés dans les expressions CLEM à l'aide de guillemets simples sous la forme de `*caractère*`, comme `z`, `A`, ou `2`. Les caractères hors limites ou les indices négatifs d'une chaîne génèrent un comportement indéfini.

Remarque. Les comparaisons entre les chaînes qui utilisent ou non les répercussions SQL peuvent produire des résultats différents lorsqu'il existe des espaces en fin de chaîne.

Fonction	Résultat	Description
allbutfirst(N, STRING)	<i>Chaîne</i>	Renvoie la chaîne <i>STRING</i> où les <i>N</i> premiers caractères ont été supprimés.
allbutlast(N, STRING)	<i>Chaîne</i>	Renvoie la chaîne <i>STRING</i> où les derniers caractères ont été supprimés.
alphabefore(STRING1, STRING2)	<i>Booléen</i>	Utilisée pour vérifier l'ordre alphabétique des chaînes. Renvoie une valeur true (vrai) si <i>STRING1</i> précède <i>STRING2</i> .
endstring(LENGTH, STRING)	<i>Chaîne</i>	Extrait les <i>N</i> derniers caractères de la chaîne indiquée. Si la longueur de la chaîne est inférieure ou égale à la valeur spécifiée, aucune modification n'a lieu.
hasendstring(STRING, SUBSTRING)	<i>Entier</i>	Cette fonction est la même que <i>isendstring(SUBSTRING, STRING)</i> .

Fonction	Résultat	Description
hasmidstring(<i>STRING</i> , <i>SUBSTRING</i>)	Entier	Cette fonction est la même que <code>ismidstring(<i>SUBSTRING</i>, <i>STRING</i>)</code> (sous-chaîne imbriquée).
hasstartstring(<i>STRING</i> , <i>SUBSTRING</i>)	Entier	Cette fonction est la même que <code>isstartstring(<i>SUBSTRING</i>, <i>STRING</i>)</code> .
hassubstring(<i>STRING</i> , <i>N</i> , <i>SUBSTRING</i>)	Entier	Cette fonction est la même que <code>issubstring(<i>SUBSTRING</i>, <i>N</i>, <i>STRING</i>)</code> , où la valeur par défaut de <i>N</i> est 1.
count_substring(<i>STRING</i> , <i>SUBSTRING</i>)	Entier	Renvoie le nombre de fois où la sous-chaîne indiquée apparaît dans la chaîne. Par exemple, <code>count_substring("foooo.txt", "oo")</code> renvoie 3.
hassubstring(<i>STRING</i> , <i>SUBSTRING</i>)	Entier	Cette fonction est la même que <code>issubstring(<i>SUBSTRING</i>, 1, <i>STRING</i>)</code> , où la valeur par défaut de <i>N</i> est 1.
isalphacode(<i>CHAR</i>)	Booléen	Renvoie la valeur true (vrai) si <i>CHAR</i> est un caractère dont le code de caractère est une lettre dans la chaîne indiquée (souvent un nom de champ). Sinon, cette fonction renvoie la valeur 0. Par exemple, <code>isalphacode(produce_num(1))</code> .
isendstring(<i>SUBSTRING</i> , <i>STRING</i>)	Entier	Si la chaîne <i>STRING</i> se termine par la sous-chaîne <i>SUBSTRING</i> , cette fonction renvoie l'indice d'entier de <i>SUBSTRING</i> dans <i>STRING</i> . Sinon, cette fonction renvoie la valeur 0.
islowercode(<i>CHAR</i>)	Booléen	Renvoie la valeur true (vrai) si <i>CHAR</i> est un caractère en minuscule dans la chaîne indiquée (souvent un nom de champ). Sinon, cette fonction renvoie la valeur 0. Par exemple, <code>islowercode('')</code> et <code>islowercode(country_name(2))</code> sont toutes deux des expressions valides.
ismidstring(<i>SUBSTRING</i> , <i>STRING</i>)	Entier	Si <i>SUBSTRING</i> est une sous-chaîne de la chaîne <i>STRING</i> , mais qu'elle ne commence pas par le premier caractère de <i>STRING</i> et ne se termine pas par le dernier, cette fonction renvoie l'indice auquel la sous-chaîne commence. Sinon, cette fonction renvoie la valeur 0.
isnumbercode(<i>CHAR</i>)	Booléen	Renvoie la valeur true (vrai) si <i>CHAR</i> est un caractère dont le code de caractère est un chiffre dans la chaîne indiquée (souvent un nom de champ). Sinon, cette fonction renvoie la valeur 0. Par exemple, <code>isnumbercode(product_id(2))</code> .
isstartstring(<i>SUBSTRING</i> , <i>STRING</i>)	Entier	Si la chaîne <i>STRING</i> commence par la sous-chaîne <i>SUBSTRING</i> , cette fonction renvoie l'indice 1. Sinon, elle renvoie une valeur 0.

Fonction	Résultat	Description
issubstring(SUBSTRING, N, STRING)	Entier	Recherche dans la chaîne <i>STRING</i> , à partir du caractère numéro <i>N</i> , la sous-chaîne <i>SUBSTRING</i> . Si cette sous-chaîne est trouvée, la fonction renvoie l'indice d'entier auquel la sous-chaîne correspondante commence. Sinon cette fonction renvoie une valeur de 0. Si la valeur de <i>N</i> n'est pas fournie, cette fonction prend par défaut la valeur 1.
issubstring(SUBSTRING, STRING)	Entier	Recherche dans la chaîne <i>STRING</i> , à partir du caractère numéro <i>N</i> , la sous-chaîne <i>SUBSTRING</i> . Si cette sous-chaîne est trouvée, la fonction renvoie l'indice d'entier auquel la sous-chaîne correspondante commence. Sinon cette fonction renvoie une valeur de 0. Si la valeur de <i>N</i> n'est pas fournie, cette fonction prend par défaut la valeur 1.
issubstring_count(SUBSTRING, N, STRING):	Entier	Renvoie l'index de la <i>N</i> ème occurrence de la sous-chaîne <i>SUBSTRING</i> au sein de la chaîne <i>STRING</i> indiquée. Si le nombre d'occurrences de la sous-chaîne <i>SUBSTRING</i> est inférieur à <i>N</i> occurrences, 0 est renvoyé.
issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)	Entier	Cette fonction est identique à <i>issubstring</i> , mais la correspondance doit commencer à l'indice <i>STARTLIM</i> ou avant, et doit se terminer à l'indice <i>ENDLIM</i> ou avant. Les contraintes <i>STARTLIM</i> ou <i>ENDLIM</i> peuvent être désactivées en donnant une valeur fausse à l'un des arguments—par exemple, <i>issubstring_lim(SUBSTRING, N, false, false, STRING)</i> est égale à <i>issubstring</i> .
isuppercode(CHAR)	Booléen	Renvoie une valeur true (vrai) si <i>CHAR</i> est un caractère majuscule. Sinon, cette fonction renvoie la valeur 0. Par exemple, <i>isuppercode('')</i> et <i>isuppercode(country_name(2))</i> sont toutes deux des expressions valides.
last(CHAR)	Chaîne	Renvoie le dernier caractère <i>CHAR</i> de la chaîne <i>STRING</i> (qui doit contenir au moins un caractère).
length(STRING)	Entier	Renvoie la longueur de la chaîne <i>STRING</i> , en nombre de caractères.

Fonction	Résultat	Description
locchar(Char, N, STRING)	Entier	Sert à identifier l'emplacement de caractères dans des champs symboliques. Cette fonction recherche le caractère <i>CHAR</i> dans la chaîne <i>STRING</i> , en commençant par le caractère numéro <i>N</i> de <i>STRING</i> . Cette fonction renvoie une valeur indiquant l'emplacement (à partir de <i>N</i>) où se trouve le caractère. Si le caractère est introuvable, la fonction renvoie la valeur 0. Si la fonction a un décalage (<i>N</i>) incorrect (par exemple, décalage dépassant la longueur de la chaîne), elle renvoie \$null\$. Par exemple, locchar(`n`, 2, web_page) recherche le caractère `n` dans le champ <i>web_page</i> en commençant par le deuxième caractère de la valeur du champ. <i>Remarque</i> : Veuillez à encadrer le caractère spécifié de guillemets simples inversés.
locchar_back(Char, N, STRING)	Entier	Semblable à locchar, cette fonction effectue la recherche vers l'arrière, en commençant par le caractère numéro <i>N</i> . Par exemple, locchar_back(`n`, 9, web_page) effectue une recherche dans le champ <i>web_page</i> en commençant par le neuvième caractère, puis en repartant vers le début de la chaîne. Si cette fonction a un décalage incorrect (par exemple, décalage dépassant la longueur de la chaîne), elle renvoie \$null\$. Pour plus d'efficacité, employez locchar_back en association avec la fonction length(<field>) afin d'utiliser la longueur de la valeur actuelle du champ de manière dynamique. Par exemple, locchar_back(`n`, (length(web_page)), web_page)
lowertoupper(Char) lowertoupper(STRING)	CHAR ou Chaîne	L'entrée peut être soit une chaîne soit un caractère, qui est utilisé dans cette fonction pour renvoyer un nouvel élément de même type, avec les caractères minuscules convertis dans leurs équivalents majuscules. Par exemple, lowertoupper(`a`), lowertoupper("My string"), et lowertoupper(field_name(2)) sont toutes des expressions valides.
matches	Booléen	Renvoie la valeur True (vrai) si une chaîne correspond à un patron spécifié. Le patron doit être un littéral de chaîne ; il ne peut pas s'agir d'un nom de champ contenant un patron. Il est possible d'inclure un point d'interrogation (?) dans le patron pour obtenir une correspondance exacte avec un caractère ; un astérisque (*) implique une correspondance avec zéro ou plusieurs caractères. Pour obtenir une correspondance avec un point d'interrogation ou un astérisque de littéral (plutôt que de les utiliser comme caractères génériques), il est

Fonction	Résultat	Description
		possible d'utiliser une barre oblique inverse (\) en tant que caractère d'échappement.
replace(SUBSTRING, NEWSUBSTRING, STRING)	Chaîne	Dans la chaîne <i>STRING</i> indiquée, remplace toutes les instances de la sous-chaîne <i>SUBSTRING</i> par la nouvelle sous-chaîne <i>NEWSUBSTRING</i> .
replicate(COUNT, STRING)	Chaîne	Renvoie une chaîne qui se compose de la chaîne d'origine copiée le nombre de fois indiqué.
stripchar(Char,STRING)	Chaîne	Cette fonction permet de supprimer des caractères d'une chaîne ou d'un champ. Vous pouvez l'utiliser, par exemple, pour supprimer de vos données des symboles de devises afin d'obtenir un simple nombre ou un nom. Par exemple, la syntaxe stripchar('\$', 'Cost') renvoie un nouveau champ où toutes les valeurs apparaissent sans le symbole dollar. <i>Remarque</i> : Veillez à encadrer le caractère spécifié de guillemets simples inversés.
skipchar(Char, N, STRING)	Entier	Recherche tout autre caractère que <i>CHAR</i> dans la chaîne <i>STRING</i> , en commençant par le caractère numéro <i>N</i> . Cette fonction renvoie un indice d'entier indiquant l'endroit auquel le caractère a été trouvé. Si tous les caractères qui suivent le caractère numéro <i>N</i> sont des caractères <i>CHAR</i> , la fonction renvoie 0. Si cette fonction a un décalage incorrect (par exemple, décalage dépassant la longueur de la chaîne), elle renvoie \$null\$. locchar est souvent utilisé en association avec les fonctions skipchar afin de déterminer la valeur de <i>N</i> (point de la chaîne à partir duquel commence la recherche) Par exemple, skipchar('s', (locchar('s', 1, "MyString")), "MyString")
skipchar_back(Char, N, STRING)	Entier	Semblable à skipchar, cette fonction effectue la recherche vers l'arrière , en commençant par le caractère numéro <i>N</i> .
startstring(LENGTH, STRING)	Chaîne	Extrait les <i>N</i> premiers caractères de la chaîne indiquée. Si la longueur de la chaîne est inférieure ou égale à la valeur spécifiée, aucune modification n'a lieu.
strmember(Char, STRING)	Entier	Semblable à locchar(Char, 1, STRING). Renvoie un indice entier qui indique l'endroit où <i>CHAR</i> s'est produit pour la première fois, ou 0. Si la fonction a un décalage non valide (par exemple, un décalage au-delà de la longueur de la chaîne), cette fonction renvoie \$null\$.
subscrs(N, STRING)	CHAR	Renvoie le caractère (<i>CHAR</i>) numéro <i>N</i> de la chaîne de saisie <i>STRING</i> . Cette fonction peut également être écrite en sténographie, STRING(N). Par exemple, lowertoupper("name"(1)) est une expression valide.

Fonction	Résultat	Description
substring(N, LEN, STRING)	Chaîne	Renvoie la sous-chaîne <i>SUBSTRING</i> , constituée de <i>LEN</i> caractères de la chaîne <i>STRING</i> , à partir du caractère situé à l'indice <i>N</i> .
substring_between(N1, N2, STRING)	Chaîne	Renvoie la sous-chaîne de la chaîne <i>STRING</i> qui commence à l'indice <i>N1</i> et se termine à l'indice <i>N2</i> .
trim(STRING)	Chaîne	Supprime les espaces blancs situés au début et à la fin de la chaîne indiquée.
trim_start(STRING)	Chaîne	Supprime les espaces blancs situés au début de la chaîne indiquée.
trimend(STRING)	Chaîne	Supprime les espaces blancs situés à la fin de la chaîne indiquée.
unicode_char(NUM)	CHAR	Renvoie le caractère avec la valeur Unicode <i>NUM</i> .
unicode_value(CHAR)	NUM	Renvoie la valeur Unicode de <i>CHAR</i> .
uppertolower(CHAR) uppertolower (STRING)	CHAR ou Chaîne	L'entrée peut être soit une chaîne soit un caractère, qui est utilisé dans cette fonction pour renvoyer un nouvel élément de même type avec les caractères en majuscule convertis dans leurs équivalents en minuscule. <i>Remarque</i> : n'oubliez pas d'encadrer les chaînes spécifiées de guillemets doubles et les caractères de guillemets simples inversés. Les noms de champ simples ne doivent pas contenir de guillemets.

Fonctions SoundEx

SoundEx est une méthode qui permet de trouver des chaînes lorsque le son est connu mais pas l'orthographe exacte. Développée en 1918, elle recherche des mots présentant un son similaire, sur la base d'hypothèses phonétiques (modalités de prononciation de certaines lettres). Il est possible, par exemple, de l'utiliser pour rechercher des noms dans une base de données où l'orthographe et la prononciation de noms similaires peuvent varier. L'algorithme SoundEx de base est décrit dans diverses sources et, malgré un certain nombre de limitations connues (par exemple, les combinaisons de lettres de début telles que ph et f n'entraînent pas de correspondance même si le son est identique), il est pris en charge d'une certaine manière par la plupart des bases de données.

Fonction	Résultat	Description
soundex(STRING)	Entier	Renvoie le code SoundEx à quatre caractères pour la chaîne <i>STRING</i> indiquée.
soundex_difference(STRING1, STRING2)	Entier	Renvoie un entier entre 0 et 4 qui indique le nombre de caractères identiques dans le codage SoundEx pour les deux chaînes (0 = absence de similarité, 4 = similarité élevée ou chaînes identiques).

Fonctions date et heure

CLEM comprend une série de fonctions permettant de manipuler des champs avec un stockage datetime contenant des variables de chaîne représentant des dates et des heures. Les formats de date et d'heure utilisés sont propres à chaque flux et sont définis dans la boîte de dialogue Propriétés du flux. Les fonctions de date et d'heure analysent les chaînes de date et d'heure en fonction du format sélectionné.

Lorsque vous indiquez dans une date une année sur deux chiffres seulement (c'est-à-dire sans préciser le siècle), IBM® SPSS® Modeler utilise le siècle par défaut défini dans la boîte de dialogue des propriétés du flux.

Remarque : Il n'est pas possible d'appeler des fonctions de date et d'heure à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Fonction	Résultat	Description
@TODAY	Chaîne	Si vous sélectionnez Passer jours/minutes dans la boîte de dialogue des propriétés du flux, cette fonction renvoie la date actuelle en tant que chaîne sous le format de date actuel. Si vous utilisez un format de date sur deux chiffres sans sélectionner l'option Passer jours/minutes, la fonction renvoie la valeur \$null\$ sur le serveur actuel. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
to_time(ITEM)	Heure	Convertit le stockage du champ spécifié en une heure.
to_date(ITEM)	Date	Convertit le stockage du champ spécifié en une date.
to_timestamp(ITEM)	Horodatage	Convertit le stockage du champ spécifié en un horodatage.
to_datetime(ITEM)	Valeurdate	Convertit le stockage du champ spécifié en une valeur d'heure, de date ou d'horodatage.
datetime_date(ITEM)	Date	Renvoie la valeur de date pour un <i>nombre</i> , une <i>chaîne</i> , ou un <i>horodatage</i> . Veuillez noter que cette fonction est la seule fonction vous permettant de reconverter un nombre (en secondes) en date. Si ITEM est une chaîne, crée une date en analysant une chaîne au format de date sélectionné. Pour que cette fonction réussisse, vous devez spécifier un format de date correct dans la boîte de dialogue des propriétés du flux. Si ITEM est un nombre, il est interprété comme un nombre de secondes écoulées depuis la date de départ (ou période). Les fractions de jour sont supprimées. Si ITEM est un horodatage, la partie date de cet horodatage est renvoyée. Si ITEM est une date, il est renvoyé sans modification.
date_before(DATE1, DATE2)	Booléen	Renvoie une valeur true (vrai) si DATE1 représente une date ou un horodatage antérieur à DATE2. Sinon, cette fonction renvoie la valeur 0.
date_days_difference(DATE1, DATE2)	Entier	Renvoie la durée en jours depuis la date ou l'horodatage qui sépare DATE1 de DATE2, sous forme d'entier. Si DATE2 est antérieure à DATE1, cette fonction renvoie un nombre négatif.

Fonction	Résultat	Description
<code>date_in_days(</code> <i>DATE</i> <code>)</code>	<i>Entier</i>	Renvoie la durée en jours qui sépare la date ou l'horodatage de référence de l'élément <i>DATE</i> indiqué, sous la forme d'un entier. Si <i>DATE</i> est antérieure à la date de référence, cette fonction renvoie un nombre négatif. Vous devez inclure une date réelle pour que le calcul fonctionne correctement. Par exemple, vous ne devez pas entrer la date : 29 février 2001. En effet, 2001 n'est pas une année bissextile et cette date n'existe pas.
<code>date_in_months(</code> <i>DATE</i> <code>)</code>	<i>Réel</i>	Renvoie la durée en mois qui sépare la date ou l'horodatage de référence de l'élément <i>DATE</i> indiqué, sous la forme d'un nombre réel. Ce chiffre est approximatif et utilise une base de 30.4375 jours par mois. Si <i>DATE</i> est antérieure à la date de référence, cette fonction renvoie un nombre. Vous devez inclure une date réelle pour que le calcul fonctionne correctement. Par exemple, vous ne devez pas entrer la date : 29 février 2001. En effet, 2001 n'est pas une année bissextile et cette date n'existe pas.
<code>date_in_weeks(</code> <i>DATE</i> <code>)</code>	<i>Réel</i>	Renvoie la durée en semaines qui sépare la date ou l'horodatage de référence de l'élément <i>DATE</i> indiqué, sous la forme d'un nombre réel. La semaine de base utilisée comprend 7,0 jours. Si <i>DATE</i> est antérieure à la date de référence, cette fonction renvoie un nombre négatif. Vous devez inclure une date réelle pour que le calcul fonctionne correctement. Par exemple, vous ne devez pas entrer la date : 29 février 2001. En effet, 2001 n'est pas une année bissextile et cette date n'existe pas.
<code>date_in_years(</code> <i>DATE</i> <code>)</code>	<i>Réel</i>	Renvoie la durée en années qui sépare la date ou l'horodatage de référence de l'élément <i>DATE</i> indiqué, sous la forme d'un nombre réel. Ce chiffre est approximatif et utilise une base de 365.25 jours. Si <i>DATE</i> est antérieure à la date de référence, cette fonction renvoie un nombre négatif. Vous devez inclure une date réelle pour que le calcul fonctionne correctement. Par exemple, vous ne devez pas entrer la date : 29 février 2001. En effet, 2001 n'est pas une année bissextile et cette date n'existe pas.
<code>date_months_difference(</code> <i>DATE1</i> , <i>DATE2</i> <code>)</code>	<i>Réel</i>	Renvoie la durée en mois depuis la date ou l'horodatage qui sépare <i>DATE1</i> de <i>DATE2</i> , sous forme d'un nombre réel. Ce chiffre est approximatif et utilise une base de 30.4375 jours par mois. Si <i>DATE2</i> est antérieure à <i>DATE1</i> , cette fonction renvoie un nombre négatif.
<code>datetime_date(</code> <i>YEAR</i> , <i>MONTH</i> , <i>DAY</i> <code>)</code>	<i>Date</i>	Renvoie la valeur de date pour les éléments <i>YEAR</i> , <i>MONTH</i> et <i>DAY</i> (année, mois, jour) indiqués. Ces arguments doivent être des entiers.
<code>datetime_day(</code> <i>DATE</i> <code>)</code>	<i>Entier</i>	Renvoie le jour du mois à partir de la valeur <i>DATE</i> ou d'un horodatage. Le résultat est un entier compris entre 1 et 31.
<code>datetime_day_name(</code> <i>DAY</i> <code>)</code>	<i>Chaîne</i>	Renvoie le nom complet de l'élément <i>DAY</i> (jour) indiqué. L'argument doit être un entier compris entre 1 (dimanche) et 7 (samedi).

Fonction	Résultat	Description
datetime_hour(TIME)	Entier	Renvoie l'heure à partir de la valeur <i>TIME</i> ou d'un horodatage. Le résultat est un entier compris entre 0 et 23.
datetime_in_seconds(TIME)	Réel	Renvoie la partie des secondes stockée dans <i>TIME</i> .
datetime_in_seconds(DATE), datetime_in_seconds(DATE-TIME)	Réel	Renvoie le chiffre cumulé, converti en secondes, de la différence entre la date <i>DATE</i> ou <i>DATE</i> <i>TIME</i> actuelle et la date de référence (1900-01-01).
datetime_minute(TIME)	Entier	Renvoie le nombre de minutes, à partir de <i>TIME</i> ou d'un horodatage. Le résultat est un entier compris entre 0 et 59.
datetime_month(DATE)	Entier	Renvoie le mois, à partir de <i>DATE</i> ou d'un horodatage. Le résultat est un entier compris entre 1 et 12.
datetime_month_name(MONTH)	Chaîne	Renvoie le nom complet de l'élément <i>MONTH</i> (mois) indiqué. L'argument doit être un entier compris entre 1 et 12.
datetime_now	Horodatage	Renvoie l'heure actuelle sous la forme d'un horodatage.
datetime_second(TIME)	Entier	Renvoie le nombre de secondes, à partir de <i>TIME</i> ou d'un horodatage. Le résultat est un entier compris entre 0 et 59.
datetime_day_short_name(DAY)	Chaîne	Renvoie le nom abrégé de l'élément <i>DAY</i> (jour) indiqué. L'argument doit être un entier compris entre 1 (dimanche) et 7 (samedi).
datetime_month_short_name(MONTH)	Chaîne	Renvoie le nom abrégé de l'élément <i>MONTH</i> (mois) indiqué. L'argument doit être un entier compris entre 1 et 12.
datetime_time(HOUR, MINUTE, SECOND)	Heure	Renvoie la valeur d'heure pour les éléments <i>HOUR</i> , <i>MINUTE</i> et <i>SECOND</i> (heures, minutes, secondes) indiqués. Ces arguments doivent être des entiers.
datetime_time(ITEM)	Heure	Renvoie la valeur d'heure de l'élément <i>ITEM</i> indiqué.
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	Horodatage	Renvoie la valeur d'horodatage pour les éléments <i>YEAR</i> , <i>MONTH</i> , <i>DAY</i> , <i>HOUR</i> , <i>MINUTE</i> et <i>SECOND</i> (année, mois, jour, heures, minutes, secondes) indiqués.
datetime_timestamp(DATE, TIME)	Horodatage	Renvoie la valeur d'horodatage des éléments <i>DATE</i> et <i>TIME</i> (date, heure) indiqués.
datetime_timestamp (NUMBER)	Horodatage	Renvoie la valeur d'horodatage du nombre de secondes indiqué.
datetime_weekday(DATE)	Entier	Renvoie le jour de la semaine à partir de la valeur <i>DATE</i> ou d'un horodatage.
datetime_year(DATE)	Entier	Renvoie l'année, à partir de <i>DATE</i> ou d'un horodatage. Le résultat est un entier (comme 2002).
date_weeks_difference (DATE1, DATE2)	Réel	Renvoie la durée en semaines depuis la date ou l'horodatage qui sépare <i>DATE1</i> de <i>DATE2</i> , sous forme d'un nombre réel. La semaine de base utilisée comprend 7,0 jours. Si <i>DATE2</i> est antérieure à <i>DATE1</i> , cette fonction renvoie un nombre négatif.
date_years_difference (DATE1, DATE2)	Réel	Renvoie la durée en années depuis la date ou l'horodatage qui sépare <i>DATE1</i> de <i>DATE2</i> , sous forme d'un nombre réel. Ce chiffre est approximatif et utilise une base de 365.25 jours. Si <i>DATE2</i> est antérieure à <i>DATE1</i> , cette fonction renvoie un nombre négatif.

Fonction	Résultat	Description
time_before(TIME1, TIME2)	Booléen	Renvoie une valeur true (vrai) si <i>TIME1</i> représente une heure ou un horodatage antérieur à <i>TIME2</i> . Sinon, cette fonction renvoie la valeur 0.
time_hours_difference (TIME1, TIME2)	Réel	Renvoie la durée en heures qui sépare les heures ou les horodatages représentés par <i>TIME1</i> et <i>TIME2</i> , sous la forme d'un nombre réel. Si vous sélectionnez l'option Passer jours/minutes dans la boîte de dialogue des propriétés du flux, une valeur plus élevée de <i>TIME1</i> fait référence au jour précédent. Si vous ne sélectionnez pas cette option, une valeur plus élevée de <i>TIME1</i> entraîne une valeur négative.
time_in_hours(TIME)	Réel	Renvoie le nombre d'heures représenté par l'heure <i>TIME</i> , sous la forme d'un nombre réel. Par exemple, pour le format d'heure HHMM, l'expression time_in_hours('0130') renvoie la valeur 1,5. <i>TIME</i> peut représenter une heure ou un horodatage.
time_in_mins(TIME)	Réel	Renvoie le nombre de minutes représenté par l'heure <i>TIME</i> , sous la forme d'un nombre réel. <i>TIME</i> peut représenter une heure ou un horodatage.
time_in_secs(TIME)	Entier	Renvoie le nombre de secondes représenté par l'heure <i>TIME</i> , sous la forme d'un entier. <i>TIME</i> peut représenter une heure ou un horodatage.
time_mins_difference(TIME1, TIME2)	Réel	Renvoie la durée en minutes qui sépare les heures ou les horodatages représentés par <i>TIME1</i> et <i>TIME2</i> , sous la forme d'un nombre réel. Si vous sélectionnez l'option Passer jours/minutes dans la boîte de dialogue des propriétés du flux, une valeur plus élevée de <i>TIME1</i> est utilisée pour faire référence au jour précédent (ou à l'heure précédente, si les minutes et les secondes sont uniquement spécifiées dans le format actuel). Si vous ne sélectionnez pas cette option, une valeur plus élevée de <i>TIME1</i> entraîne une valeur négative.
time_secs_difference(TIME1, TIME2)	Entier	Renvoie la durée en secondes qui sépare l'heure ou l'horodatage représenté par <i>TIME1</i> de celle représentée par <i>TIME2</i> , sous la forme d'un entier. Si vous sélectionnez l'option Passer jours/minutes dans la boîte de dialogue des propriétés du flux, une valeur plus élevée de <i>TIME1</i> est utilisée pour faire référence au jour précédent (ou à l'heure précédente, si les minutes et les secondes sont uniquement spécifiées dans le format actuel). Si vous ne sélectionnez pas cette option, une valeur plus élevée de <i>TIME1</i> entraîne une valeur négative.

Conversion des valeurs date et heure.

Notez que les fonctions de conversion (et toutes les autres fonctions qui nécessitent un type spécifique d'entrée, par exemple une valeur de date ou d'heure) dépendent des formats actuels indiqués dans la boîte de dialogue des options de flux. Par exemple, si vous avez un champ *DATE* de type chaîne avec les valeurs *Jan 2003*, *Fev 2003*, etc., vous pouvez le convertir en stockage de date sous le format suivant :

to_date(DATE)

Pour que cette conversion fonctionne, sélectionnez le format de date correspondant MON YYYY comme format de date par défaut pour le flux. [Pour plus d'informations, reportez-vous à la section Définition des options de flux dans le chapitre 5 sur p. 51.](#)

Pour un exemple de conversion de valeurs de chaînes en dates à l'aide d'un noeud Remplacer, consultez le flux *broadband_create_models.str*, installé dans le dossier *\Demos* du sous-dossier *streams*. [Pour plus d'informations, reportez-vous à la section Préviation avec le noeud Séries temporelles dans le chapitre 15 dans Guide des applications de IBM SPSS Modeler 14.2.](#)

Dates stockées sous forme de nombres. Veuillez noter que *DATE* dans l'exemple précédent est le nom d'un champ, alors que *to_date* est une fonction CLEM. Si vous avez des dates stockées sous la forme de nombres, vous pouvez les convertir en utilisant la fonction *datetime_date* où le nombre est interprété comme un nombre de secondes écoulées depuis la date de départ (ou période).

```
datetime_date(DATE)
```

En convertissant une date en un nombre de secondes (et vice-versa), vous pouvez effectuer des calculs tels que le calcul de la date actuelle plus ou moins un certain nombre de jours, par exemple :

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

Fonctions séquentielles

Dans le cadre de certaines opérations, la séquence des événements est importante. Cette application vous permet de travailler avec les séquences d'enregistrements suivantes :

- Séquences et séries temporelles
- Fonctions séquentielles
- Indexation des enregistrements
- Moyenne, somme et valeurs comparatives
- Contrôle des changements - Différenciation
- @SINCE
- Valeurs de décalage
- Outils séquentiels supplémentaires

Pour nombre d'applications, chaque enregistrement circulant dans un flux peut être considéré comme un cas particulier, indépendant de tous les autres. Auxquels cas, l'ordre des enregistrements importe généralement peu.

Toutefois, pour certaines catégories de problèmes, la séquence des enregistrements est importante. Il s'agit en général de situations de séries temporelles dans lesquelles la séquence des enregistrements représente une séquence ordonnée d'événements ou d'occurrences. Chaque enregistrement représente un instantané d'un moment particulier. Cependant, une grande partie des informations les plus complexes peut être contenue non pas dans des valeurs instantanées, mais dans la façon dont de telles valeurs évoluent et se comportent dans le temps.

Bien entendu, le paramètre pertinent peut être autre que le temps. Par exemple, les enregistrements peuvent représenter des analyses effectuées à distance sur une ligne, mais les mêmes principes s'appliquent.

Les fonctions séquentielles et les fonctions spéciales sont reconnaissables aux caractéristiques suivantes.

- Elles sont toutes précédées du préfixe @.
- Leur nom apparaît en majuscules.

Les fonctions séquentielles font référence à l'enregistrement actuellement traité par un noeud, aux enregistrements qui sont déjà passés dans un noeud et même, dans un cas, aux enregistrements qui ne sont pas encore passés dans un noeud. Les fonctions séquentielles peuvent être librement mélangées avec d'autres composants d'expressions CLEM, bien que certains aient des restrictions quant à leurs arguments.

Exemples

Vous pouvez avoir besoin de connaître le temps écoulé depuis un certain événement ou une condition vraie. Pour cela, utilisez la fonction @SINCE. Par exemple :

```
@SINCE(Income > Outgoings)
```

Cette fonction renvoie le décalage du dernier enregistrement où cette condition était vraie, c'est-à-dire le nombre d'enregistrements précédant celui où la condition était vraie. Si la condition n'a jamais été vraie, @SINCE renvoie @INDEX + 1.

Vous pouvez également devoir faire référence à une valeur de l'enregistrement en cours dans l'expression utilisée par @SINCE. Pour cela, utilisez la fonction @THIS, qui spécifie qu'un nom de champ s'applique toujours à l'enregistrement en cours. Pour déterminer le décalage du dernier enregistrement dont la valeur de champ Concentration est plus de deux fois supérieure à celle de l'enregistrement en cours, vous pouvez utiliser :

```
@SINCE(Concentration > 2 * @THIS(Concentration))
```

Dans certains cas, la condition attribuée à @SINCE est par définition vraie pour l'enregistrement en cours. Par exemple :

```
@SINCE(ID == @THIS(ID))
```

Pour cette raison, @SINCE n'évalue pas la condition de l'enregistrement en cours. Utilisez une fonction identique, @SINCE0, pour évaluer la condition de l'enregistrement en cours et des précédents. Si la condition est vraie dans l'enregistrement en cours, @SINCE0 renvoie 0.

Remarque : Il n'est pas possible d'appeler des fonctions @ à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Fonction	Résultat	Description
MEAN(FIELD)	Réel	Renvoie la moyenne des valeurs du ou des champs (<i>FIELD</i> ou <i>FIELDS</i>) spécifiés.
@MEAN(FIELD, EXPR)	Réel	Renvoie la moyenne des valeurs du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, y compris l'enregistrement actuel. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la moyenne des enregistrements reçus est renvoyée. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
@MEAN(FIELD, EXPR, INT)	Réel	Renvoie la moyenne des valeurs du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, y compris l'enregistrement actuel. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la moyenne des enregistrements reçus est renvoyée. <i>INT</i> indique le nombre maximal de valeurs à prendre en compte. Cette méthode est beaucoup plus efficace que si vous limitez à deux arguments.
@DIFF1(FIELD)	Réel	Renvoie le premier différentiel du champ <i>FIELD1</i> . La forme à un seul argument renvoie simplement la différence entre la valeur de champ actuelle et la valeur de champ précédente. Renvoie 0 si aucun enregistrement précédent pertinent n'existe.
@DIFF1(FIELD1, FIELD2)	Réel	La forme à deux arguments renvoie le premier différentiel du champ <i>FIELD1</i> par rapport au champ <i>FIELD2</i> . Renvoie 0 si aucun enregistrement précédent pertinent n'existe.
@DIFF2(FIELD)	Réel	Renvoie le second différentiel du champ <i>FIELD1</i> . La forme à un seul argument renvoie simplement la différence entre la valeur de champ actuelle et la valeur de champ précédente. Renvoie 0 si aucun enregistrement précédent pertinent n'existe.
@DIFF2(FIELD1, FIELD2)	Réel	La forme à deux arguments renvoie le premier différentiel du champ <i>FIELD1</i> par rapport au champ <i>FIELD2</i> . Renvoie 0 si aucun enregistrement précédent pertinent n'existe.

Fonction	Résultat	Description
@INDEX	Entier	Renvoie l'index de l'enregistrement actuel. Les indices sont affectés à chaque enregistrement lors de son arrivée sur le noeud actuel. Le premier enregistrement reçoit l'index 1 et la valeur d'index augmente de 1 à chacun des enregistrements suivants.
@LAST_NON_BLANK(FIELD)	Tout	Renvoie la dernière valeur du champ <i>FIELD</i> qui était renseigné, tel que défini dans un noeud <i>Type</i> ou source en amont. S'il n'existe aucune valeur renseignée pour le champ <i>FIELD</i> dans les enregistrements lus jusqu'à présent, la fonction renvoie \$null\$. Les valeurs non renseignées, également appelées valeurs manquantes utilisateur, peuvent être définies séparément pour chaque champ.
@MAX(FIELD)	Nombre	Renvoie la valeur maximale du champ <i>FIELD</i> spécifié.
@MAX(FIELD, EXPR)	Nombre	Renvoie la valeur maximale du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus jusqu'à présent, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression dont le résultat est un entier supérieur à 0.
@MAX(FIELD, EXPR, INT)	Nombre	Renvoie la valeur maximale du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus jusqu'à présent, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la valeur minimale des enregistrements reçus est renvoyée. <i>INT</i> indique le nombre maximal de valeurs à prendre en compte. Cette méthode est beaucoup plus efficace que si vous vous limitez à deux arguments.
@MIN(FIELD)	Nombre	Renvoie la valeur minimale du champ <i>FIELD</i> spécifié.
@MIN(FIELD, EXPR)	Nombre	Renvoie la valeur minimale du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus jusqu'à présent, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression dont le résultat est un entier supérieur à 0.
@MIN(FIELD, EXPR, INT)	Nombre	Renvoie la valeur minimale du champ <i>FIELD</i> correspondant aux <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus jusqu'à présent, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la valeur minimale des enregistrements reçus est renvoyée. <i>INT</i> indique le nombre maximal de valeurs à prendre en compte. Cette méthode est beaucoup plus efficace que si vous vous limitez à deux arguments.

Fonction	Résultat	Description
@OFFSET(FIELD, EXPR)	Tout	Renvoie la valeur de <i>FIELD</i> pour l'enregistrement décalé de la valeur <i>EXPR</i> par rapport à l'enregistrement actuel. Un décalage positif permet d'accéder à un enregistrement qui a déjà eu lieu, alors qu'un décalage négatif indique un enregistrement anticipé. Par exemple, @OFFSET(Status, 1) renvoie la valeur du champ Status de l'enregistrement précédent, alors que @OFFSET(Status, -4) « anticipe » quatre enregistrements dans la séquence (enregistrements qui ne sont pas encore passés par ce noeud) pour obtenir la valeur. <i>Veillez noter qu'un décalage négatif (fonction d'anticipation) doit être spécifié comme constante.</i> Pour les décalages positifs uniquement, <i>EXPR</i> peut également être une expression CLEM arbitraire, évaluée pour que l'enregistrement actuel calcule le décalage. Dans ce cas, l'utilisation de la version à trois arguments de cette fonction devrait améliorer les performances (consultez la fonction suivante). Si le résultat de cette expression n'est pas un entier non négatif, une erreur est générée. Il est en effet interdit de calculer des décalages anticipés. <i>Remarque</i> : Une fonction @OFFSET autoréférentielle ne peut pas utiliser une anticipation littérale. Par exemple, dans un noeud Remplacer, vous ne pouvez pas remplacer la valeur de field1 par une expression comme @OFFSET(field1,-2).
@OFFSET(FIELD, EXPR, INT)	Tout	Effectue la même opération que la fonction @OFFSET, avec un troisième argument, <i>INT</i> , qui indique le nombre maximal de valeurs à prendre en compte. Si le décalage est calculé à partir d'une expression, ce troisième argument devrait améliorer les performances. Par exemple, dans une formule telle que @OFFSET(Foo, Month, 12), le système sait comment conserver uniquement les douze dernières valeurs de Foo. Sinon, il doit stocker chaque valeur à tout hasard. Pour les décalages dont la valeur est constante—y compris les décalages « anticipés » négatifs qui doivent être constants—le troisième argument n'est d'aucune utilité. Il est recommandé d'utiliser la version à deux arguments de cette fonction. Consultez également la note sur les fonctions autoréférentielles dans la version à deux arguments décrite plus tôt.
@SDEV(FIELD)	Réel	Renvoie l'écart-type des valeurs du ou des champs <i>FIELD</i> ou <i>FIELDS</i> spécifiques.
@SDEV(FIELD, EXPR)	Réel	Renvoie l'écart-type des valeurs du champ <i>FIELD</i> pour les <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, y compris l'enregistrement actuel. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, l'écart-type des enregistrements reçus est renvoyé.

Fonction	Résultat	Description
@SDEV(FIELD, EXPR, INT)	Réel	Renvoie l'écart-type des valeurs du champ <i>FIELD</i> pour les <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, y compris l'enregistrement actuel. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, l'écart-type des enregistrements reçus est renvoyé. <i>INT</i> indique le nombre maximal de valeurs à prendre en compte. Cette méthode est beaucoup plus efficace que si vous vous limitez à deux arguments.
@SINCE(EXPR)	Tout	Renvoie le nombre d'enregistrements qui ont été transmis depuis que l'expression <i>CLEM</i> arbitraire <i>EXPR</i> est devenue vraie.
@SINCE(EXPR, INT)	Tout	Lorsque vous ajoutez le second argument, <i>INT</i> , vous indiquez le nombre maximal d'enregistrements à prendre en compte. Si <i>EXPR</i> n'a jamais été vraie, <i>INT</i> est équivalent à @INDEX+1.
@SINCE0(EXPR)	Tout	Considère l'enregistrement en cours, contrairement à @SINCE. La fonction @SINCE0 renvoie la valeur 0 si l'expression <i>EXPR</i> est vraie pour l'enregistrement actuel.
@SINCE0(EXPR, INT)	Tout	Lorsque vous ajoutez le second argument, <i>INT</i> indique le nombre maximal d'enregistrements à prendre en compte.
@SUM(FIELD)	Nombre	Renvoie la somme des valeurs du ou des champs (<i>FIELD</i> ou <i>FIELDS</i>) spécifiés.
@SUM(FIELD, EXPR)	Nombre	Renvoie la somme des valeurs du champ <i>FIELD</i> pour les <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la somme des enregistrements reçus est renvoyée.
@SUM(FIELD, EXPR, INT)	Nombre	Renvoie la somme des valeurs du champ <i>FIELD</i> pour les <i>EXPR</i> (nombre représenté par une expression) derniers enregistrements reçus par le noeud actuel, enregistrement actuel compris. <i>FIELD</i> doit être le nom d'un champ numérique. <i>EXPR</i> peut être toute expression donnant comme résultat un entier supérieur à 0. Si l'expression <i>EXPR</i> est omise ou donne un résultat supérieur au nombre d'enregistrements reçus jusqu'à présent, la somme des enregistrements reçus est renvoyée. <i>INT</i> indique le nombre maximal de valeurs à prendre en compte. Cette méthode est beaucoup plus efficace que si vous vous limitez à deux arguments.
@THIS(FIELD)	Tout	Renvoie la valeur du champ <i>FIELD</i> dans l'enregistrement actuel. Utilisée uniquement dans les expressions @SINCE.

Fonctions globales

Les fonctions @MEAN, @SUM, @MIN, @MAX, et @SDEV fonctionnent avec, au plus, tous les enregistrements lus jusqu'à l'enregistrement en cours compris. Toutefois, dans certains cas, il est utile de comparer les valeurs de l'enregistrement actuel à celles de tout l'ensemble de données. Lorsque vous utilisez un noeud V. globales (Valeurs globales) pour générer des valeurs dans tout l'ensemble de données, vous pouvez accéder à ces valeurs dans une expression CLEM en utilisant les fonctions globales.

Par exemple,

@GLOBAL_MAX(Age)

renvoie la valeur maximale du champ Age dans l'ensemble de données, tandis que l'expression

$(\text{Value} - @GLOBAL_MEAN(\text{Value})) / @GLOBAL_SDEV(\text{Value})$

exprime la différence entre la valeur du champ Value dans l'enregistrement actuel et la moyenne globale, sous la forme d'un écart-type. Vous ne pouvez utiliser les valeurs globales qu'après leur calcul par un noeud V. globales (Valeurs globales). Vous pouvez annuler toutes les valeurs globales actuelles en cliquant sur le bouton Effacer les valeurs globales, dans l'onglet Valeurs globales de la boîte de dialogue des propriétés du flux.

Remarque : Il n'est pas possible d'appeler des fonctions @ à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Fonction	Résultat	Description
@GLOBAL_MAX(FIELD)	Nombre	Renvoie la valeur maximale du champ <i>FIELD</i> sur tout l'ensemble de données, tel qu'il a été précédemment généré par un noeud V. globales (Valeurs globales). <i>FIELD</i> doit être le nom d'un champ numérique. Une erreur se produit si la valeur globale correspondante n'a pas été définie. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
@GLOBAL_MIN(FIELD)	Nombre	Renvoie la valeur minimale du champ <i>FIELD</i> sur tout l'ensemble de données, tel qu'il a été précédemment généré par un noeud V. globales (Valeurs globales). <i>FIELD</i> doit être le nom d'un champ numérique. Une erreur se produit si la valeur globale correspondante n'a pas été définie.
@GLOBAL_SDEV(FIELD)	Nombre	Renvoie l'écart-type des valeurs du champ <i>FIELD</i> sur tout l'ensemble de données, tel qu'il a été précédemment généré par un noeud V. globales (Valeurs globales). <i>FIELD</i> doit être le nom d'un champ numérique. Une erreur se produit si la valeur globale correspondante n'a pas été définie.

Fonction	Résultat	Description
@GLOBAL_MEAN(FIELD)	Nombre	Renvoie la moyenne des valeurs du champ <i>FIELD</i> sur tout l'ensemble de données, tel qu'il a été précédemment généré par un noeud V. globales (Valeurs globales). <i>FIELD</i> doit être le nom d'un champ numérique. Une erreur se produit si la valeur globale correspondante n'a pas été définie.
@GLOBAL_SUM(FIELD)	Nombre	Renvoie la somme des valeurs du champ <i>FIELD</i> sur tout l'ensemble de données, tel qu'il a été précédemment généré par un noeud V. globales (Valeurs globales). <i>FIELD</i> doit être le nom d'un champ numérique. Une erreur se produit si la valeur globale correspondante n'a pas été définie.

Traitement des valeurs nulles et non renseignées

Le langage CLEM permet d'indiquer que certaines valeurs d'un champ doivent être considérées comme "blancs", ou manquantes. Vous pouvez utiliser les fonctions suivantes pour travailler avec des valeurs non renseignées.

Remarque : Il n'est pas possible d'appeler des fonctions @ à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Fonction	Résultat	Description
@BLANK(FIELD)	Booléen	Renvoie la valeur true (vrai) pour tous les enregistrements dont les valeurs sont vides, d'après les règles de traitement des blancs définies dans un noeud Typier ou un noeud source en amont (onglet Types). Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
@LAST_NON_BLANK(FIELD)	Tout	Renvoie la dernière valeur du champ <i>FIELD</i> qui était renseigné, tel que défini dans un noeud Typier ou source en amont. S'il n'existe aucune valeur renseignée pour le champ <i>FIELD</i> dans les enregistrements lus jusqu'à présent, la fonction renvoie \$null\$. Les valeurs non renseignées, également appelées valeurs manquantes utilisateur, peuvent être définies séparément pour chaque champ.
@NULL(FIELD)	Booléen	Renvoie la valeur true (vrai) si la valeur de <i>FIELD</i> est la valeur \$null\$, manquante dans le système. Renvoie la valeur false (faux) pour toutes les autres valeurs, y compris les valeurs non renseignées par les utilisateurs. Si vous souhaitez rechercher les deux valeurs, utilisez @BLANK(FIELD) et @NULL(FIELD).
undef	Tout	Cette fonction est généralement utilisée dans CLEM pour entrer une valeur \$null\$.—Par exemple, pour remplacer des valeurs non renseignées par des valeurs nulles dans le noeud Remplacer.

Les champs non renseignés peuvent être « remplis » avec le noeud Remplacer. Dans les noeuds Remplacer et Calculer (mode Multiple uniquement), la fonction spéciale CLEM@FIELD fait référence aux champs en cours.

Champs spéciaux

Les fonctions spéciales servent à signaler les champs spéciaux en cours d'analyse ou à générer une liste de champs en tant qu'entrée. Par exemple, lors du calcul simultané de plusieurs champs, utilisez @FIELD pour que le programme comprenne que « l'action de calcul doit porter sur les champs sélectionnés ». L'expression log(@FIELD) calcule un nouveau champ de logarithme pour chaque champ sélectionné.

Remarque : Il n'est pas possible d'appeler des fonctions @ à partir de scripts. [Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.](#)

Fonction	Résultat	Description
@FIELD	Tout	Effectue une action sur tous les champs spécifiés dans le contexte de l'expression. Il est impossible d'appeler cette fonction à partir d'un script. Pour plus d'informations, reportez-vous à la section Expressions CLEM dans les scripts dans le chapitre 3 dans Guide de génération de scripts et d'automatisation IBM SPSS Modeler 14.2.
@TARGET	Tout	Quand une expression CLEM est utilisée dans une fonction d'analyse définie par l'utilisateur, @TARGET représente le champ cible ou la « valeur correcte » de la paire cible/prédite analysée. Cette fonction est fréquemment utilisée dans les noeuds Analyse.
@PREDICTED	Tout	Quand une expression CLEM est utilisée dans une fonction d'analyse définie par l'utilisateur, @PREDICTED représente la valeur prédite de la paire cible/prédite analysée. Cette fonction est fréquemment utilisée dans les noeuds Analyse.
@PARTITION_FIELD	Tout	Remplace le nom du champ de partition actuel.
@TRAINING_PARTITION	Tout	Renvoie la valeur de la partition d'apprentissage actuelle. Par exemple, pour sélectionner des noeuds d'apprentissage utilisant un noeud Sélectionner, utilisez l'expression CLEM suivante : @PARTITION_FIELD = @TRAINING_PARTITION Ainsi, le noeud Sélectionner sera toujours fonctionnel, quelles que soient les valeurs utilisées pour représenter chaque partition des données.
@TESTING_PARTITION	Tout	Renvoie la valeur de la partition de test actuelle.
@VALIDATION_PARTITION	Tout	Renvoie la valeur de la partition de validation actuelle.

Fonction	Résultat	Description
@FIELDS_BETWEEN(start, end)	<i>Tout</i>	Renvoie la liste de noms des champs situés entre les champs de début et de fin spécifiés (ces champs étant inclus), en fonction de l'ordre naturel (c'est-à-dire l'ordre d'insertion) des champs dans les données. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
@FIELDS_MATCHING(pattern)	<i>Tout</i>	Renvoie la liste des noms de champ qui correspondent à un patron spécifié. Il est possible d'inclure un point d'interrogation (?) dans le patron pour obtenir une correspondance exacte avec un caractère ; un astérisque (*) implique une correspondance avec zéro ou plusieurs caractères. Pour obtenir une correspondance avec un point d'interrogation ou un astérisque de littéral (plutôt que de les utiliser comme caractères génériques), il est possible d'utiliser une barre oblique inverse (\) en tant que caractère d'échappement. Pour plus d'informations, reportez-vous à la section Récapitulation de plusieurs champs dans le chapitre 7 sur p. 110.
@MULTI_RESPONSE_SET	<i>Tout</i>	Renvoie la liste des champs dans l'ensemble à réponses multiples nommé. Pour plus d'informations, reportez-vous à la section Utilisation des données à réponses multiples dans le chapitre 7 sur p. 112.

Utilisation de IBM SPSS Modeler avec un référentiel

A propos du IBM SPSS Collaboration and Deployment Services Repository

IBM® SPSS® Modeler peut s'utiliser avec le référentiel IBM SPSS Collaboration and Deployment Services, ce qui vous permet de gérer le cycle de vie des modèles de Data Mining et des objets prédictifs associés qui pourront être utilisés par des applications, outils et solutions d'entreprise. Les objets SPSS Modeler qui peuvent être partagés de cette façon sont les flux, les noeuds, les sorties de flux, les scénarios, les projets et les modèles. Les objets sont stockés dans le référentiel central, dans lequel ils peuvent être partagés avec d'autres applications et suivis à l'aide de fonctions étendues de gestion des versions, des métadonnées et de recherche.

Remarque : Une licence distincte est requise pour accéder à un référentiel IBM® SPSS® Collaboration and Deployment Services. Pour plus d'informations, reportez-vous à <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

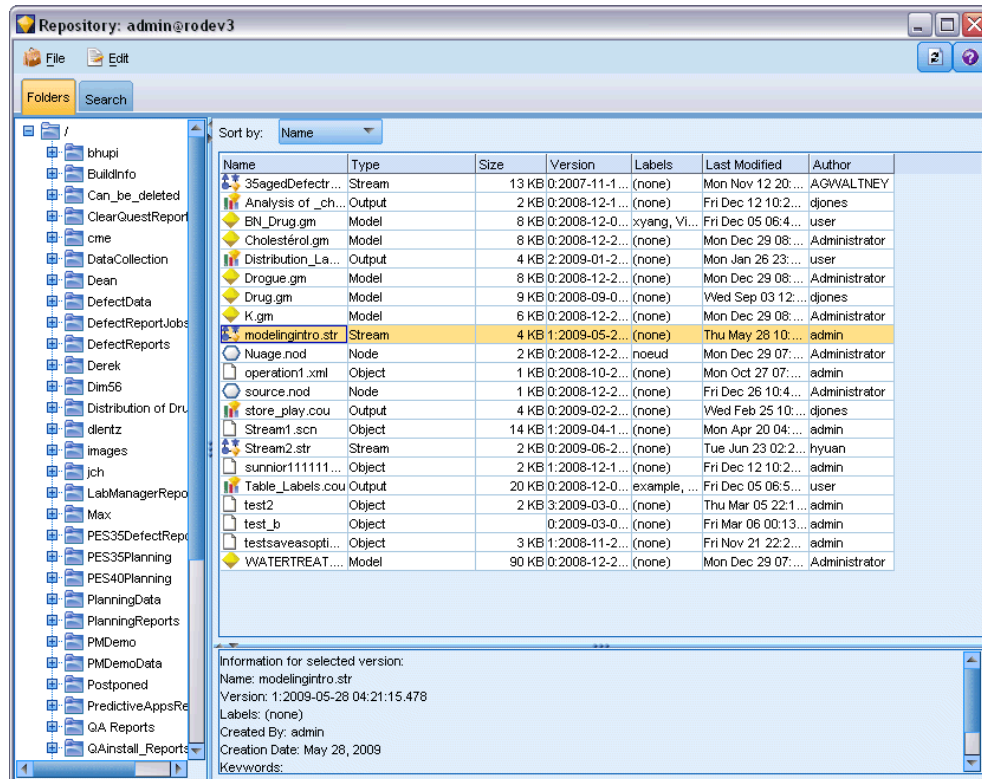
Avant que vous ne puissiez utiliser SPSS Modeler avec le référentiel, vous devez installer un adaptateur sur l'hôte du référentiel. Sans cet adaptateur, il est possible que le message suivant apparaisse lors d'une tentative d'accès aux objets du référentiel de certains noeuds ou modèles SPSS Modeler :

The repository may need updating to support new node, model and output types.

Pour les instructions concernant l'installation de l'adaptateur, reportez-vous au guide *Installation du déploiement de SPSS Modeler*, disponible dans le DVD du déploiement de SPSS Modeler. Vous pouvez obtenir des informations sur l'accès aux objets du référentiel de SPSS Modeler depuis IBM® SPSS® Collaboration and Deployment Services Deployment Manager dans le *guide de déploiement de SPSS Modeler*.

Les sections suivantes fournissent des informations sur l'accès au référentiel depuis SPSS Modeler.

Figure 9-1
Objets du IBM SPSS Collaboration and Deployment Services Repository



Prise en charge étendue de la gestion des versions et de la recherche

Le référentiel contient des versions d'objet et des fonctionnalités de recherche étendues. Par exemple, supposons que vous créez un flux et que vous le stockiez dans le référentiel (à partir duquel il peut être partagé avec les chercheurs d'autres services). Si vous mettez ensuite le flux à jour dans SPSS Modeler, vous pouvez ajouter la version mise à jour au référentiel sans remplacer la version précédente. Toutes les versions restent accessibles et peuvent faire l'objet d'une recherche sur la base du nom, de l'étiquette, et des champs utilisés ou d'autres attributs. Par exemple, vous pouvez rechercher toutes les versions de modèle utilisant les revenus nets comme entrée ou tous les modèles créés par un auteur en particulier. (Pour effectuer cette recherche dans un système de fichiers traditionnel, vous devriez enregistrer chaque version sous un nom de fichier différent et le logiciel ne reconnaîtrait pas les relations entre les versions.)

Connexion unique

La fonction de connexion unique permet à l'utilisateur de se connecter au IBM® SPSS® Collaboration and Deployment Services Repository sans devoir entrer son nom d'utilisateur et son mot de passe à chaque fois. Les informations de connexion du réseau local existant fournissent

l'authentification nécessaire à IBM SPSS Collaboration and Deployment Services. Cette fonction dépend des paramètres suivants :

- IBM SPSS Collaboration and Deployment Services doit être configuré pour utiliser un fournisseur de connexion unique.
- L'utilisateur doit être connecté à un hôte compatible avec le fournisseur.

Pour plus d'informations, reportez-vous à la section [Connexion au IBM SPSS Collaboration and Deployment Services Repository](#) sur p. 162.

Stockage et et déploiement des objets IBM SPSS Collaboration and Deployment Services Repository

Les **flux** créés dans IBM® SPSS® Modeler peuvent être stockés tels quels dans IBM® SPSS® Collaboration and Deployment Services Repository, tout comme les fichiers avec l'extension *.str*. De cette façon, plusieurs utilisateurs de la même entreprise peuvent accéder à un même flux. [Pour plus d'informations, reportez-vous à la section Stockage des objets dans le IBM SPSS Collaboration and Deployment Services Repository](#) sur p. 166.

Il est également possible de déployer un flux dans le IBM SPSS Collaboration and Deployment Services Repository. Un **flux déployé** est stocké comme fichier avec l'extension *.str*, mais avec des métadonnées supplémentaires pour activer le flux que l'application client léger IBM® SPSS® Modeler Advantage doit utiliser. [Pour plus d'informations, reportez-vous à la section Ouverture d'un flux dans IBM SPSS Modeler Advantage dans le chapitre 10](#) sur p. 201.

Cependant, pour tirer le meilleur parti des fonctionnalités professionnelles de IBM® SPSS® Collaboration and Deployment Services, le flux doit être déployé en tant que **scénario**. Il s'agit d'un fichier (avec l'extension *.scn*) contenant un flux SPSS Modeler avec des noeuds, des modèles et des propriétés supplémentaires spécifiques qui lui permettent d'être utilisé dans le IBM SPSS Collaboration and Deployment Services Repository à des fins de scoring et d'actualisation de modèles automatiques. Par exemple, un modèle d'auto-apprentissage (MRAA) peut être automatiquement mis à jour à intervalles réguliers lorsque de nouvelles données sont disponibles. Il est aussi possible de déployer un ensemble de flux pour l'analyse champion/challenger dans laquelle des scénarios sont comparés afin de déterminer celui qui contient le modèle prédictif le plus efficace.

Pour des informations supplémentaires sur le stockage des flux en tant que scénarios, consultez [Options de déploiement des flux](#) sur p. 188.

Configuration requise pour les flux déployés en tant que scénarios

- L'accès régulier aux données d'entreprise nécessite l'accès aux flux déployés en tant que scénarios via le composant IBM® SPSS® Collaboration and Deployment Services Enterprise View de IBM SPSS Collaboration and Deployment Services. Cela signifie que dans SPSS Modeler, chaque branche de scoring ou de modélisation dans le flux doit contenir au moins un noeud Enterprise View source. [Pour plus d'informations, reportez-vous à la section Noeud Enterprise View dans le chapitre 2 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

- Pour utiliser le noeud Enterprise View, IBM SPSS Collaboration and Deployment Services doit d'abord être installé, configuré sur votre site, avec Enterprise View, une Vue d'application et des Définitions de fournisseur de données (DPD) déjà définis. Pour plus d'informations, contactez votre administrateur local ou consultez le site Web d'entreprise à l'adresse <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- Un DPD est défini par rapport à une source de données ODBC particulière. Pour utiliser un DPD à partir de SPSS Modeler, vous devez avoir une source de données ODBC définie sur l'hôte du serveur SPSS Modeler qui porte le même nom et qui se connecte au même magasin de données que celui référencé dans le DPD.
- En outre, le IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver doit être installé sur chaque ordinateur utilisé pour modifier ou exécuter le flux. Pour Windows, installez simplement le lecteur sur l'ordinateur où IBM® SPSS® Modeler ou IBM® SPSS® Modeler Server est installé et où aucune autre configuration du lecteur n'est nécessaire. Sur UNIX, une référence au script `pev.sh` doit être ajoutée au script de démarrage. [Pour plus d'informations, reportez-vous à la section Configuration d'un pilote pour le noeud Enterprise View dans l'annexe B dans Administration et guide des performances de IBM SPSS Modeler Server 14.2.](#) Contactez votre administrateur local pour des détails sur l'installation du IBM SPSS Collaboration and Deployment Services Enterprise View Driver.

Autres options de déploiement

L'application IBM SPSS Collaboration and Deployment Services offre des fonctions très étendues pour la gestion du contenu d'entreprise, mais d'autres mécanismes sont également disponibles pour le déploiement ou l'exportation de flux :

- Utilisez l'assistant de Predictive Applications version 4.x pour exporter des flux à déployer dans cette version de Predictive Applications. [Pour plus d'informations, reportez-vous à la section Assistant des applications prédictives 4.x dans le chapitre 10 sur p. 201.](#)
- Exportez le flux et le modèle afin de pouvoir les utiliser ultérieurement avec IBM® SPSS® Modeler Solution Publisher Runtime. [Pour plus d'informations, reportez-vous à la section IBM SPSS Modeler Solution Publisher dans le chapitre 2 dans IBM SPSS Modeler 14.2 Solution Publisher.](#)
- Exportez un ou plusieurs modèles au format PMML. Ce format XML permet de coder les informations sur le modèle. [Pour plus d'informations, reportez-vous à la section Importation et exportation de modèles au format PMML dans le chapitre 10 sur p. 209.](#)

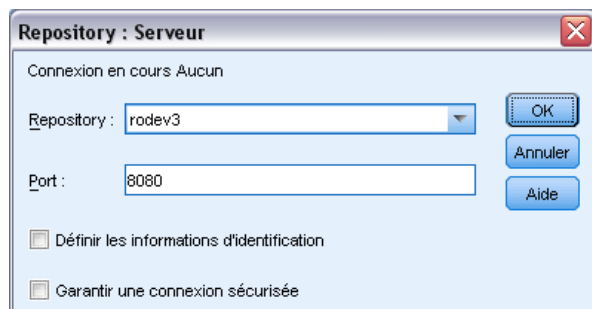
Connexion au IBM SPSS Collaboration and Deployment Services Repository

- ▶ Pour vous connecter au IBM® SPSS® Collaboration and Deployment Services Repository à partir du menu IBM® SPSS® Modeler principal, cliquez sur :
Outils > Repository > Options...
- ▶ Indiquez les options de connexion requises.

Les paramètres sont propres à chaque site ou installation. Pour les informations de port et de connexion spécifiques, contactez votre administrateur système local.

Remarque : Une licence distincte est requise pour accéder à un référentiel IBM® SPSS® Collaboration and Deployment Services. Pour plus d'informations, reportez-vous à <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Figure 9-2
Connexion au IBM SPSS Collaboration and Deployment Services Repository



Référentiel. L'installation du IBM SPSS Collaboration and Deployment Services Repository à laquelle vous souhaitez accéder. En règle générale, elle correspond au nom du serveur hôte sur lequel le référentiel est installé. Vous pouvez vous connecter à un seul référentiel à la fois.

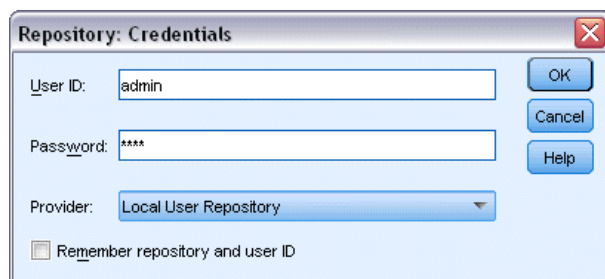
Port. Port utilisé pour établir la connexion (par défaut, le port 8080).

Définir les informations d'identification. Laissez cette case décochée pour permettre à la fonction de **connexion unique** de se connecter à l'ordinateur local à l'aide de vos informations de nom d'utilisateur et de mot de passe. Si la connexion unique n'est pas disponible, ou si vous cochez la case pour désactiver la connexion unique (par exemple pour vous connecter à un compte administrateur), un écran s'affiche dans lequel vous entrez vos informations d'identification.

Garantir une connexion sécurisée. Indique si une connexion SSL (Secure Socket Layer) doit être utilisée. Le protocole SSL est fréquemment utilisé pour la sécurisation des données sur un réseau. Pour pouvoir utiliser cette fonction, vous devez activer le protocole SSL sur le serveur hébergeant le IBM SPSS Collaboration and Deployment Services Repository. Si nécessaire, contactez votre administrateur local pour plus d'informations.

Saisie des informations d'identification pour IBM SPSS Collaboration and Deployment Services Repository

Figure 9-3
Saisie des informations d'identification de IBM SPSS Collaboration and Deployment Services Repository



ID utilisateur et mot de passe. Indiquez un nom d'utilisateur et un mot de passe valides lors de la connexion. Si nécessaire, contactez votre administrateur local pour plus d'informations.

Fournisseur. Choisir un fournisseur de sécurité pour l'authentification. Le IBM® SPSS® Collaboration and Deployment Services Repository peut être configuré pour utiliser différents fournisseurs de sécurité, si nécessaire, contactez votre administrateur local pour plus d'informations.

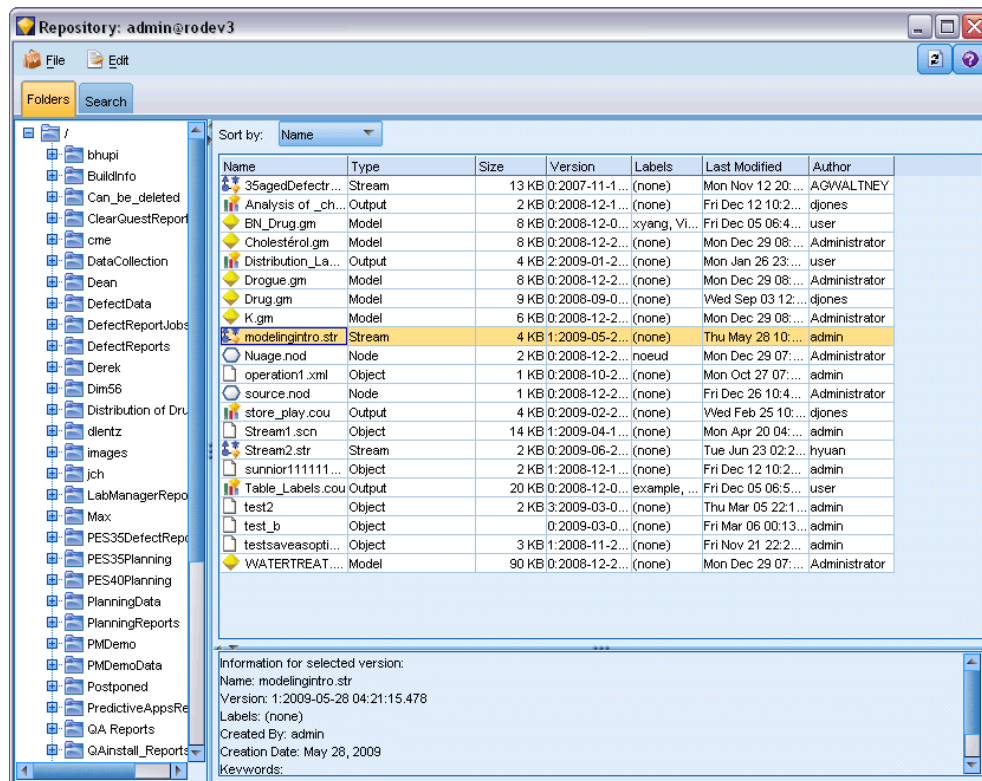
Mémoriser le référentiel et l'ID d'utilisateur. Enregistrez les paramètres actuels en tant que paramètres par défaut pour vous éviter de les entrer à nouveau à chaque connexion.

Accès au contenu du IBM SPSS Collaboration and Deployment Services Repository

Le IBM® SPSS® Collaboration and Deployment Services Repository vous permet de parcourir le contenu stocké de la même manière que l'Explorateur Windows ; vous pouvez également accéder aux différentes *versions* de chaque objet stocké.

- ▶ Pour ouvrir la fenêtre du IBM SPSS Collaboration and Deployment Services Repository à partir des menus IBM® SPSS® Modeler, cliquez sur :
Outils > Repository > Explorer
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.

Figure 9-4
Accès au contenu du IBM SPSS Collaboration and Deployment Services Repository

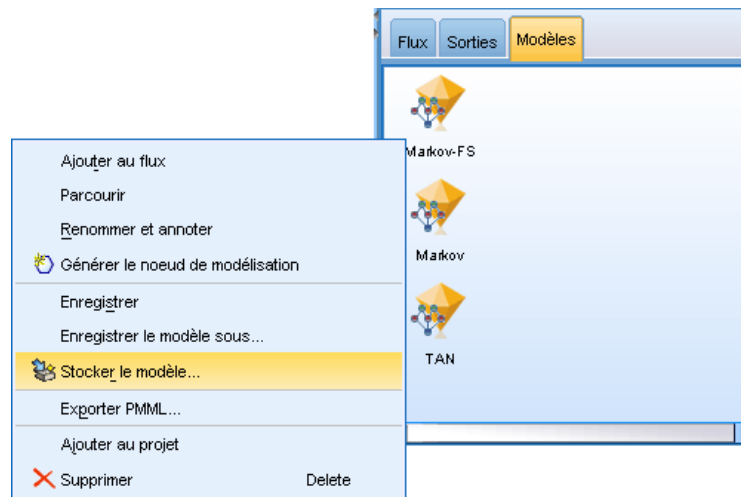


La fenêtre de l'explorateur affiche au début un arbre de la hiérarchie des dossiers. Cliquez sur le nom du dossier pour afficher son contenu.

Les objets correspondant à la sélection ou aux critères de recherche actuels sont répertoriés dans le panneau de droite, et les informations détaillées concernant la version sélectionnée apparaissent dans le panneau inférieur droit. Les attributs affichés s'appliquent à la version la plus récente.

Stockage des objets dans le IBM SPSS Collaboration and Deployment Services Repository

Figure 9-5
Stockage d'un modèle



Vous pouvez stocker les flux, les noeuds, les modèles, les palettes de modèles, les projets et les objets de sortie dans le référentiel, à partir duquel les applications et les autres utilisateurs peuvent accéder.

Remarque : Une licence distincte est requise pour accéder à un référentiel IBM® SPSS® Collaboration and Deployment Services. Pour plus d'informations, reportez-vous à <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Vous pouvez également publier la sortie d'un flux dans le IBM® SPSS® Collaboration and Deployment Services Repository dans un format permettant aux autres utilisateurs de la consulter sur Internet grâce à IBM® SPSS® Collaboration and Deployment Services Deployment Portal. Pour plus d'informations, reportez-vous à la section [Publication sur le Web](#) dans le chapitre 6 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.

Définition des propriétés d'un objet

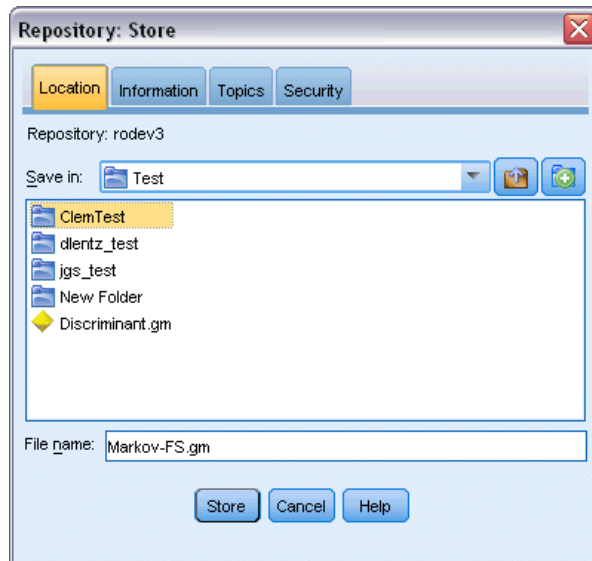
Lorsque vous stockez un objet, le référentiel : la boîte de dialogue Stocker apparaît, pour que vous y définissiez les valeurs de certaines propriétés de l'objet. Vous pouvez :

- cliquer sur le nom et le dossier du référentiel sous lequel l'objet doit être stocké
- Ajouter des informations sur l'objet tel que l'étiquette de version et d'autres propriétés pouvant être recherchées
- Affecter une ou plusieurs rubriques de classification à l'objet
- Définir les options de sécurité de l'objet

Les sections suivantes décrivent les propriétés que vous pouvez définir.

Choix de l'emplacement de stockage des objets

Figure 9-6
Choix de l'emplacement de stockage d'un objet



Enregistrer dans. Indique le dossier en cours : l'emplacement où l'objet sera stocké. Faites un double clic sur le nom du dossier dans la liste pour définir ce dossier comme le dossier en cours. Utilisez le bouton Dossier supérieur pour naviguer vers le dossier parent. Utilisez le bouton Nouveau dossier pour créer un dossier au niveau actuel.

Nom du fichier. Le nom sous lequel l'objet sera stocké.

Stocker. Stocke l'objet à l'emplacement actuel.

Ajout d'informations sur les objets stockés

Tous les champs de cet onglet sont optionnels.

Figure 9-7
Ajout d'informations sur l'objet

The screenshot shows a dialog box titled "Repository: Store" with a close button in the top right corner. It has four tabs: "Location", "Information" (which is selected and highlighted in yellow), "Topics", and "Security". Below the tabs, there are several input fields and controls:

- Author:** A text field containing the text "admin".
- Version Label:** A list box containing three items: "LATEST" (which is checked with a blue square), "All SLRM-TrainingData", and "Baseline". To the right of the list box is an "Add..." button.
- Description:** A large empty text area.
- Keywords:** A large empty text area.
- Expiration:** Two radio buttons are present. The first is labeled "None" and is selected. The second is labeled "Date:" and is followed by a text field containing "July 30, 2010" and a small calendar icon.

At the bottom of the dialog box, there are three buttons: "Store", "Cancel", and "Help".

Auteur. Le nom d'utilisateur de l'utilisateur créant l'objet dans le référentiel. Par défaut, cela affiche le nom d'utilisateur utilisé pour la connexion du référentiel mais vous pouvez modifier ce nom ici.

Étiquette de version. Sélectionnez une étiquette dans la liste pour indiquer la version de l'objet ou cliquez sur Ajouter pour créer une nouvelle étiquette. Évitez d'utiliser le caractère "[" dans l'étiquette. Vérifiez qu'aucune case n'est sélectionnée si vous ne souhaitez pas assigner une étiquette à la version de cet objet. [Pour plus d'informations, reportez-vous à la section Affichage et modification des propriétés des objets sur p. 184.](#)

Description : Une description de l'objet. Les utilisateurs peuvent rechercher les objets par description (reportez-vous à la remarque).

Mots-clés. Un ou plusieurs mots-clés associés à l'objet et qui peuvent être utilisés pour la recherche (reportez-vous à la remarque).

Expiration. Une date à partir de laquelle l'objet n'est plus visible aux utilisateurs réguliers, bien qu'elle puisse encore être vue par son propriétaire et par l'administrateur du référentiel. Pour définir une date d'expiration, sélectionnez l'option Date et saisissez la date ou choisissez-en une à l'aide du bouton du calendrier.

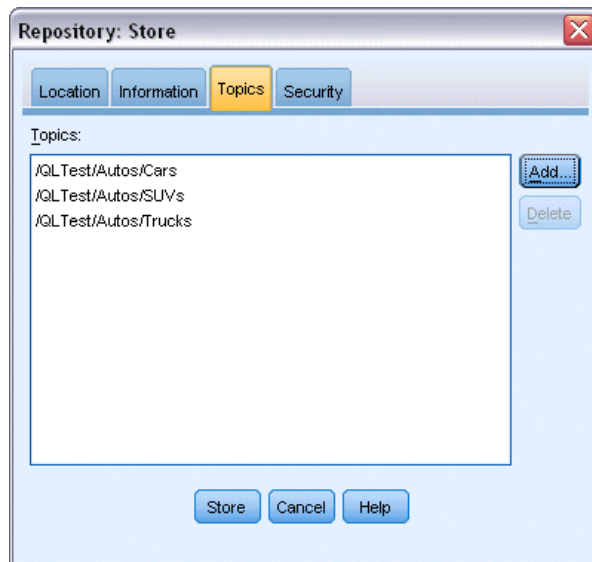
Stocker. Stocke l'objet à l'emplacement actuel.

Remarque : Les informations dans les champs Description et mots-clés sont traitées séparément de tout ce qui peut être saisi dans IBM® SPSS® Modeler dans l'onglet Annotations de l'objet. Une recherche dans le référentiel par description ou mot-clé ne renvoie pas d'informations de l'onglet Annotations. [Pour plus d'informations, reportez-vous à la section Recherche d'objets dans le IBM SPSS Collaboration and Deployment Services Repository sur p. 177.](#)

Affectation des rubriques à un objet stocké

Les rubriques sont un système de classification hiérarchique du contenu stocké dans le référentiel. Vous pouvez choisir parmi les rubriques disponibles lors du stockage d'objets et les utilisateurs peuvent également rechercher des objets par rubrique. La liste des rubriques disponibles est définie par des utilisateurs du référentiel avec les privilèges appropriés (pour plus d'informations, consultez le *Guide de l'utilisateur de Deployment Manager*).

Figure 9-8
Affectation des rubriques à un objet



Pour affecter une rubrique à l'objet :

- ▶ Cliquez sur le bouton Ajouter.
- ▶ Cliquez sur le nom d'une rubrique dans la liste des rubriques disponibles.
- ▶ Cliquez sur OK.

Pour supprimer l'affectation d'une rubrique :

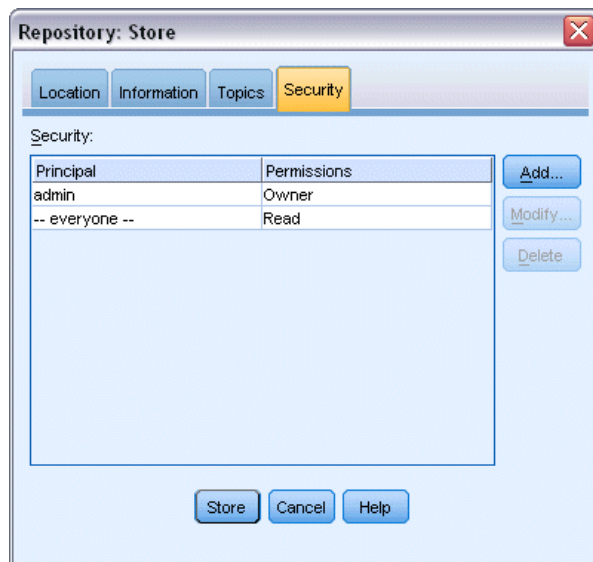
- ▶ Sélectionnez la rubrique dans la liste de rubriques affectées.
- ▶ Cliquez sur Supprimer.

Définition des options de sécurité des objets stockés

Vous pouvez définir ou modifier un certain nombre d'options de sécurité pour un objet stocké. Pour une ou plusieurs **entité(s)** (c'est-à-dire des utilisateurs ou des groupes d'utilisateurs), vous pouvez :

- affecter les droits d'accès à l'objet
- Modifier les droits d'accès à l'objet
- Supprimer les droits d'accès à l'objet

Figure 9-9
Définition des options de sécurité d'un objet



Entité. Le nom d'utilisateur du référentiel d'un utilisateur ou d'un groupe qui a les droits d'accès à cet objet.

Droits d'accès. Les droits d'accès que cet utilisateur ou ce groupe ont pour cet objet.

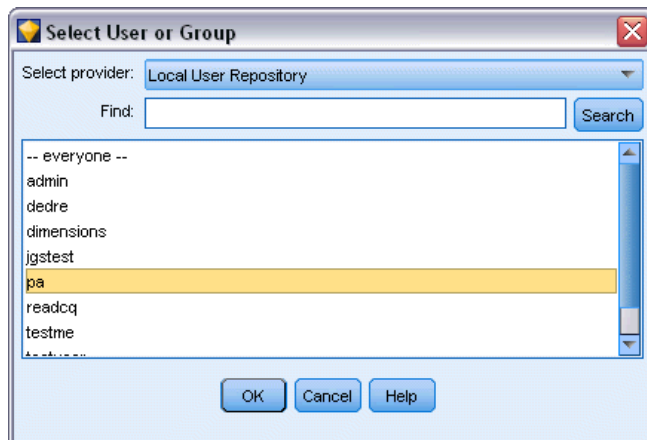
Ajouter. Permet d'ajouter un ou plusieurs utilisateurs ou groupes à la liste de ceux ayant des droits d'accès à cet objet. [Pour plus d'informations, reportez-vous à la section Ajout d'un utilisateur à la liste des droits d'accès sur p. 171.](#)

Modifier. Permet de modifier les droits d'accès de l'utilisateur ou du groupe à cet objet. L'accès en lecture est accordé par défaut. Cette option permet d'accorder des droits d'accès supplémentaires qui sont les droits de Propriétaire, d'Écriture, de Supprimer et de Modifier.

Supprimer. Supprime l'utilisateur ou le groupe sélectionné de la liste des droits d'accès à cet objet.

Ajout d'un utilisateur à la liste des droits d'accès

Figure 9-10
Ajout d'un utilisateur à la liste des droits d'accès à un objet



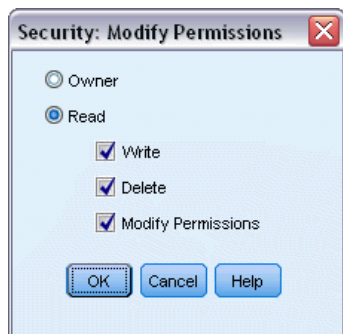
Sélectionner un fournisseur. Choisir un fournisseur de sécurité pour l'authentification. Le IBM® SPSS® Collaboration and Deployment Services Repository peut être configuré pour utiliser différents fournisseurs de sécurité, si nécessaire, contactez votre administrateur local pour plus d'informations.

Chercher. Saisissez le nom d'utilisateur du référentiel de l'utilisateur ou du groupe à ajouter puis cliquez sur Rechercher pour afficher le nom dans la liste d'utilisateurs. Pour ajouter plusieurs noms d'utilisateur en une fois, laissez ce champ vide et cliquez simplement sur Rechercher pour afficher une liste de tous les noms d'utilisateur du référentiel.

Liste des utilisateurs. Sélectionnez un ou plusieurs noms d'utilisateur et cliquez sur OK pour les ajouter à la liste des droits d'accès.

Modification des droits d'accès à un objet

Figure 9-11
Modification des droits d'accès à un objet



Propriétaire. Sélectionnez cette option pour accorder à cet utilisateur ou à ce groupe les droits d'accès à cet objet. Le propriétaire a le contrôle entier sur l'objet, y compris les droits d'accès Supprimer et Modifier.

Lecture. Par défaut, un utilisateur ou un groupe qui n'est pas le propriétaire de l'objet ne possède que les droits d'accès en lecture à l'objet. Sélectionnez les cases appropriées pour ajouter les droits d'accès Ecriture, Supprimer et Modifier pour cet utilisateur ou ce groupe.

Stockage des flux

Vous pouvez stocker un flux en tant que fichier *.str* dans le référentiel où il sera accessible aux autres utilisateurs.

Remarque : Pour des informations sur le déploiement d'un flux, pour bénéficier de fonctionnalités supplémentaires du référentiel, consultez Déploiement de flux sur p. 187.

Pour stocker le flux actuel :

- ▶ Dans le menu principal, cliquez sur :
Fichier > Stocker > Stocker comme flux...
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Stockage de projets

Vous pouvez stocker un projet IBM® SPSS® Modeler entier en tant que fichier *.cpj* dans le référentiel où il sera accessible aux autres utilisateurs.

Parce qu'un fichier de projet est un conteneur pour d'autres objets SPSS Modeler, vous devez informer SPSS Modeler de stocker les objets du projet dans le IBM® SPSS® Collaboration and Deployment Services Repository. Vous utilisez pour cela un paramètre de la boîte de dialogue Propriétés du projet. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un projet dans le chapitre 11 sur p. 220.](#)

Une fois que vous avez configuré un projet en vue de stocker des objets dans le référentiel, SPSS Modeler vous invite automatiquement à stocker les objets dès que vous les ajoutez au projet.

Lorsque la session SPSS Modeler est terminée, vous devez stocker une nouvelle version du fichier de projet de façon à ce que les ajouts soient mémorisés. Le fichier de projet inclut automatiquement (et extrait) les versions les plus récentes des objets qu'il contient. Si vous n'avez pas ajouté d'objets à un projet au cours d'une session SPSS Modeler, il est inutile de stocker de nouveau le fichier de projet. Vous devez toutefois stocker la nouvelle version des objets modifiés (flux, sortie, etc.).

Pour stocker un projet

- ▶ Sélectionnez le projet dans l'onglet CRISP-DM ou Classes dans le panneau des gestionnaires dans SPSS Modeler, et dans le menu principal, cliquez sur :
Fichier > Projet > Stocker le projet..
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Stockage des noeuds

Vous pouvez stocker une définition de noeud individuel du flux actuel en tant que fichier *.nod* dans le référentiel où il sera accessible aux autres utilisateurs.

Pour stocker un noeud :

- ▶ cliquez avec le bouton droit de la souris sur le noeud dans l'espace de travail de flux et cliquez sur Stocker le noeud.
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Stockage des objets de sortie

Vous pouvez stocker un objet de sortie du flux actuel en tant que fichier *.cou* dans le référentiel où il sera accessible aux autres utilisateurs.

Pour stocker un objet de sortie :

- ▶ Cliquez sur l'objet dans l'onglet Sorties dans le panneau des gestionnaires dans IBM® SPSS® Modeler, et dans le menu principal, cliquez sur :
Fichier > Sorties > Stocker la sortie...
- ▶ Vous pouvez également cliquer avec le bouton droit de la souris sur un objet de l'onglet Sorties et cliquer sur Stocker.

- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Stockage des modèles et des palettes de modèles

Vous pouvez stocker un modèle individuel en tant que fichier *.gm* dans le référentiel où il sera accessible aux autres utilisateurs. Vous pouvez également stocker le contenu complet de la palette des modèles en tant que fichier *.gen* dans le référentiel.

Stockage d'un modèle

- ▶ Cliquez sur l'objet de la palette Modèles dans IBM® SPSS® Modeler, et dans le menu principal, cliquez sur :
Fichier > Modèles > Stocker le modèle...
- ▶ Vous pouvez également cliquer avec le bouton droit de la souris sur un objet de la palette des modèles et cliquer sur Stocker le modèle.
- ▶ Continuez depuis la rubrique « Terminer la procédure de stockage ».

Stockage d'une palette de modèles

- ▶ Faites un clic droit sur l'arrière-plan de la palette des modèles.
- ▶ Dans le menu contextuel, cliquez sur Stocker la palette.
- ▶ Continuez depuis la rubrique « Terminer la procédure de stockage ».

Terminer la procédure de stockage

- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Extraction d'objets du IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez extraire des flux, des modèles, des palettes de modèles, des noeuds, des projets et des objets de sortie qui ont été stockés dans le IBM® SPSS® Collaboration and Deployment Services Repository.

Remarque : En plus d'utiliser les options de menu décrites ici, vous pouvez également extraire des flux, des objets de sortie, des modèles et des palettes de modèles en faisant un clic droit sur l'onglet approprié du panneau des gestionnaires en haut à droite de la fenêtre IBM® SPSS® Modeler.

- ▶ Pour extraire un flux, dans le menu SPSS Modeler principal, cliquez sur :
Fichier > Extraire le flux...

 - ▶ Pour extraire un modèle, une palette de modèles, un projet ou un objet de sortie, dans le menu SPSS Modeler principal, cliquez sur :
Fichier > Modèles > Extraire le modèle...

 - ou*
Fichier > Modèles > Extraire la palette des modèles...

 - ou*
Fichier > Projets > Extraire un projet...

 - ou*
Fichier > Sorties > Extraire la sortie...
- ▶ Vous pouvez également cliquer avec le bouton droit de la souris dans le panneau du projet ou des gestionnaires, puis cliquer sur Extraire dans le menu contextuel.

 - ▶ Pour extraire un noeud, dans le menu SPSS Modeler principal, cliquez sur :
Insertion > Noeud depuis le référentiel/Super noeud à partir du référentiel

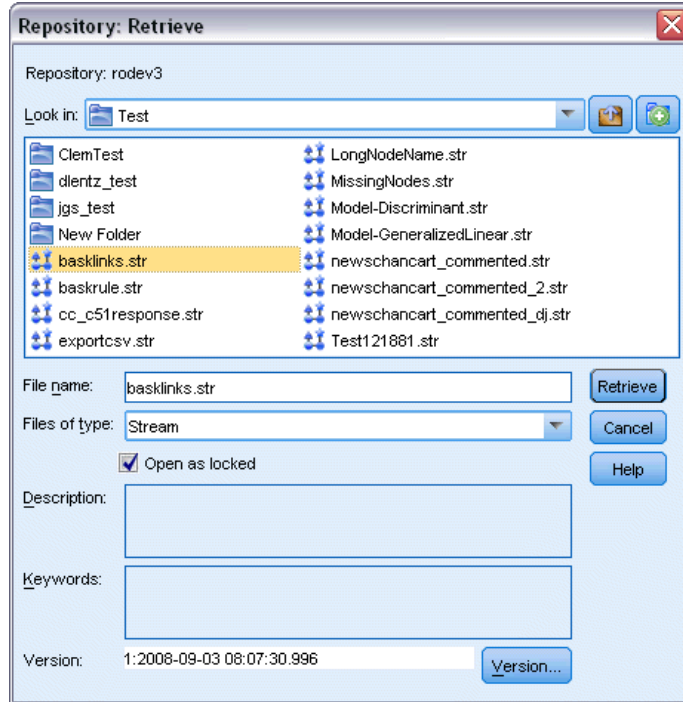
 - ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.

 - ▶ Dans le référentiel : Récupérez la boîte de dialogue, recherchez l'objet, sélectionnez-le et cliquez sur le bouton Extraire.

Choix d'un objet à extraire

Figure 9-12

Extraction d'un objet dans le IBM SPSS Collaboration and Deployment Services Repository



Chercher dans. Affiche la hiérarchie des dossiers pour le dossier actuel. Pour naviguer vers un autre dossier, sélectionnez-en un dans cette liste pour vous y rendre directement, ou naviguez à l'aide de la liste des objets sous ce champ.

Bouton Dossier supérieur. Passe à un niveau au-dessus du dossier actuel dans la hiérarchie.

Bouton Nouveau dossier. Crée un nouveau dossier au niveau actuel dans la hiérarchie.

Nom du fichier. Le nom de fichier du référentiel de l'objet sélectionné. Pour extraire cet objet, cliquez sur Extraire.

Fichiers de type. Le type d'objet que vous avez choisi d'extraire. Seuls les objets de ce type, en plus des dossiers, apparaissent dans la liste d'objets. Pour afficher les objets d'un autre type d'extraction, sélectionnez le type d'objet dans la liste.

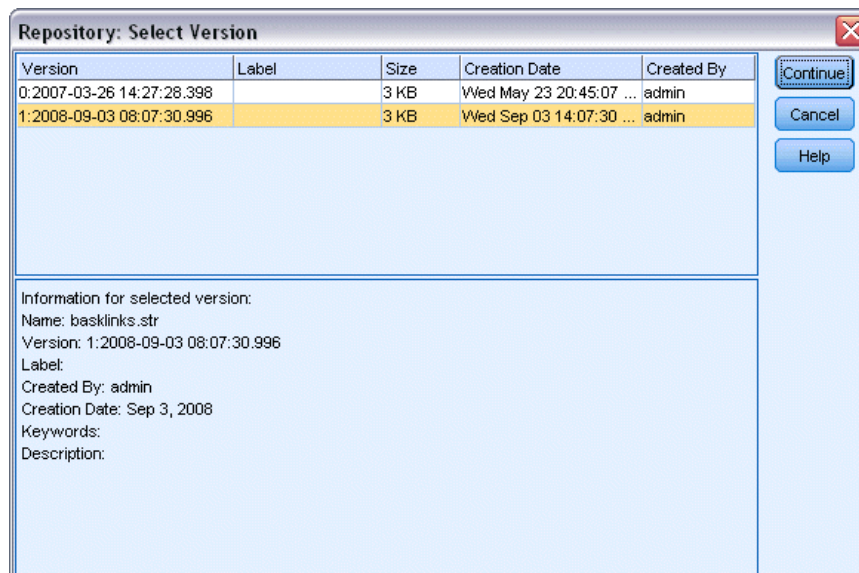
Ouvrir en tant qu'objet verrouillé. Par défaut, lorsqu'un objet est extrait, il est verrouillé dans le référentiel pour que les autres personnes ne puissent pas le mettre à jour. Si vous ne souhaitez pas que l'objet soit verrouillé lors de l'extraction, décochez cette case.

Description, mots-clés. Si des informations supplémentaires sur l'objet ont été définies lors de son stockage, ces informations apparaissent ici. [Pour plus d'informations, reportez-vous à la section Ajout d'informations sur les objets stockés sur p. 167.](#)

Version. Pour extraire une version de l'objet autre que la plus récente, cliquez sur le bouton. Les informations détaillées de toutes les versions qui apparaissent vous permettent de sélectionner la version de votre choix.

Sélection de la version d'objet

Figure 9-13
Sélection de la version d'objet



Pour sélectionner la version spécifique d'un objet du référentiel.

- ▶ (Facultatif) Pour trier la liste par version, étiquette, taille, date de création ou créateur, double-cliquez sur l'en-tête de la colonne appropriée.
- ▶ Sélectionnez la version d'objet à utiliser.
- ▶ Cliquez sur Poursuivre.

Recherche d'objets dans le IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez rechercher des objets en fonction du nom, du dossier, du type, de l'étiquette, de la date ou d'autres critères.

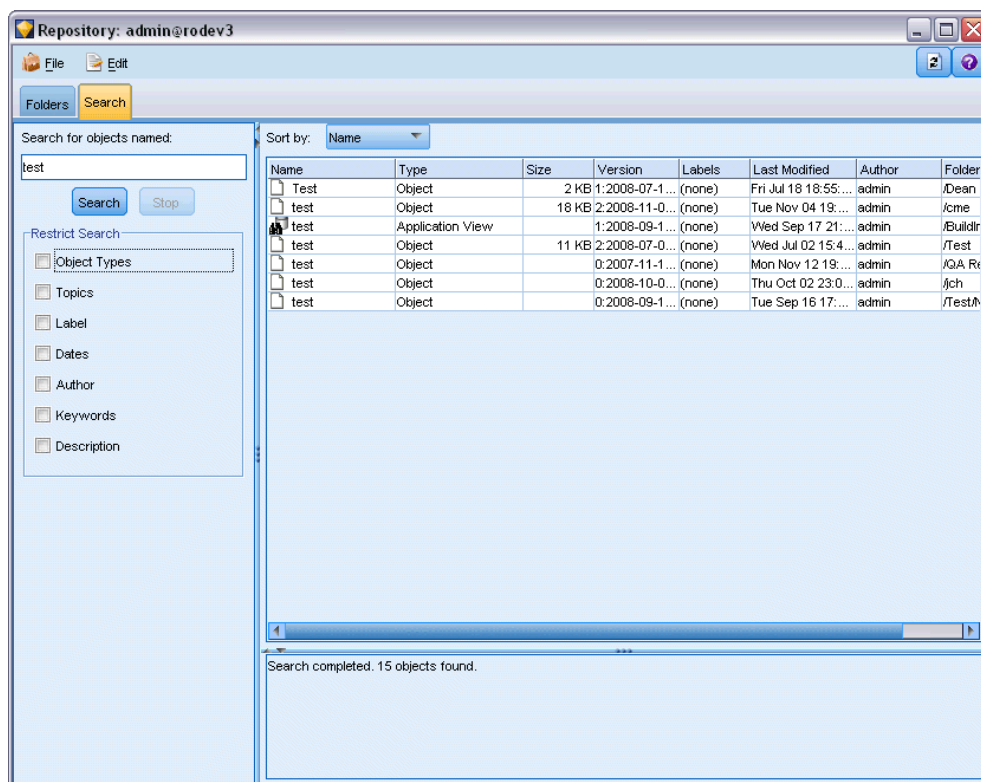
Recherche par nom

Pour rechercher les objets par nom :

- ▶ Dans le menu IBM® SPSS® Modeler principal, cliquez sur :
Outils > IBM® SPSS® Collaboration and Deployment Services Repository > Explorer

- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Cliquez sur l'onglet Rechercher.
- ▶ Dans le champ Rechercher des objets nommés, spécifiez le nom de l'objet à rechercher.

Figure 9-14
Recherche d'objets en fonction du nom



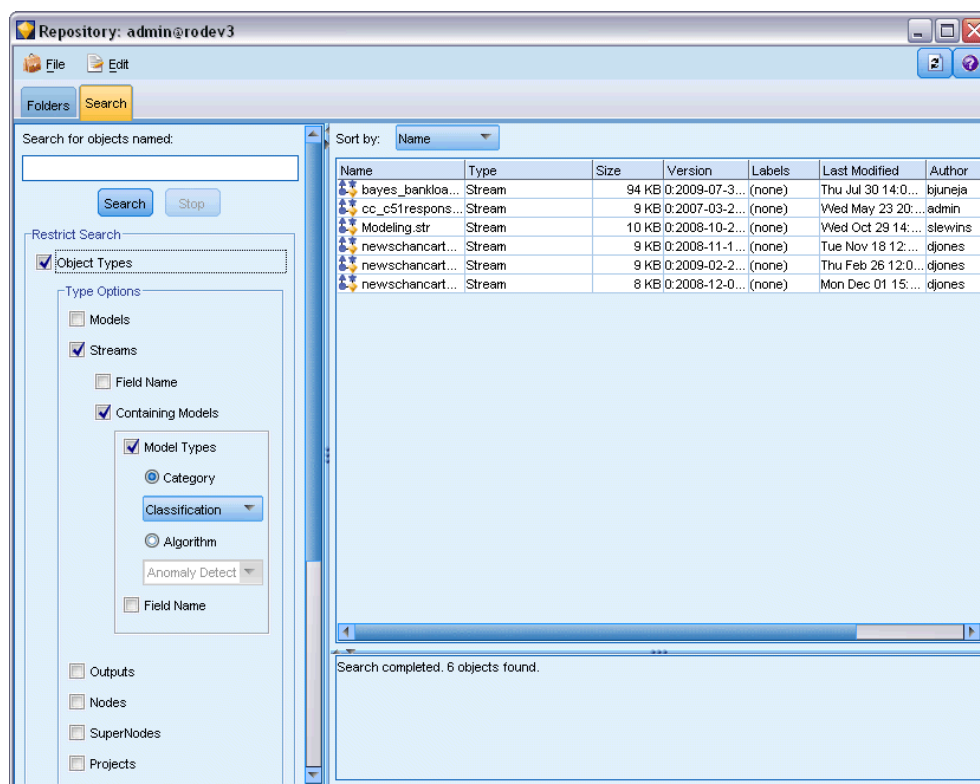
Lorsque vous recherchez des objets en fonction de leur nom, vous pouvez utiliser un astérisque (*) comme caractère générique remplaçant toute chaîne de caractères et un point d'interrogation (?) qui, lui, remplace un caractère unique. Par exemple, une recherche portant sur **classe** renvoie tous les objets dont le nom inclut la chaîne *classe*. La recherche de la chaîne *m0?_** renvoie *M01_classe.str* et *M02_classe.str* mais pas *M01a_classe.str*. Les recherches ne distinguent pas les majuscules des minuscules (la recherche de la chaîne *classe* renvoie aussi bien *Classe* que *CLASSE*).

Remarque : Si le nombre d'objets est élevé, les recherches peuvent prendre un certain temps.

Recherche en fonction d'autres critères

Vous pouvez effectuer une recherche en fonction du titre, de l'étiquette, de la date, de l'auteur, de mots-clés, d'un contenu indexé ou de la description d'un objet. Seuls les objets correspondant à tous les critères de recherche indiqués sont détectés. Par exemple, vous pouvez localiser tous les flux qui contiennent un ou plusieurs modèles de classification, présentant une étiquette spécifique et ayant été modifiés après une date définie.

Figure 9-15
Recherche de flux contenant un type spécifique de modèle



Types d'objet. Vous pouvez restreindre la recherche aux modèles, aux flux, aux sorties, aux noeuds, aux super noeuds, aux projets, aux palettes de modèles, aux scénarios ou à d'autres types d'objets.

- **Modèles :** Vous pouvez rechercher des modèles par catégorie (classification, approximation, classification non supervisée, etc.) ou par algorithme de modélisation particulier tel que Kohonen.

Vous pouvez également effectuer une recherche selon les champs utilisés, comme tous les modèles qui utilisent un champ *revenu* en tant que champ d'entrée ou de sortie (cible).

- **Flux.** Pour les flux, vous pouvez restreindre la recherche aux champs utilisés ou au type de modèle (catégorie ou algorithme) contenu dans le flux.

Rubriques. Vous pouvez rechercher les modèles associés à des rubriques particulières dans une liste définie par les utilisateurs du référentiel qui ont les droits d'accès appropriés (pour plus d'informations, consultez le *Guide des utilisateurs de Deployment Manager*). Pour obtenir la liste,

cochez cette case puis cliquez sur le bouton « Ajouter des rubriques » qui apparaît, sélectionnez une ou plusieurs rubriques dans la liste puis cliquez sur OK.

Étiquette. Limite la recherche à des étiquettes de version d'objet spécifiques.

Dates. Vous pouvez indiquer une date de création ou de modification, et rechercher les objets antérieurs, ultérieurs ou appartenant à l'intervalle de dates indiqué.

Auteur. Limite la recherche à des objets créés par un utilisateur particulier.

Mots-clés. Recherche reposant sur des mots-clés précis. Dans SPSS Modeler, les mots-clés sont indiqués dans l'onglet Annotation d'un flux, d'un modèle ou d'un objet de sortie.

Description : Recherche des termes précis figurant dans le champ de description. Dans SPSS Modeler, la description figure dans l'onglet Annotation d'un flux, d'un modèle ou d'un objet de sortie. Vous pouvez séparer les expressions recherchées par des points-virgules, comme l'illustre l'exemple suivant : revenu ; type de culture ; valeur de la demande. (Au sein d'une expression recherchée, les espaces ont une importance. Par exemple, l'occurrence type de culture dotée d'un espace et l'occurrence type de culture incluant deux espaces sont différentes.)

Modification des objets IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez modifier des objets existants dans le IBM® SPSS® Collaboration and Deployment Services Repository directement depuis IBM® SPSS® Modeler. Vous pouvez :

- créer, renommer ou supprimer des dossiers
- Verrouiller ou déverrouiller des objets
- Supprimer des objets

Création, changement de nom et suppression de dossiers

- ▶ Pour effectuer des opérations sur des dossiers dans le IBM® SPSS® Collaboration and Deployment Services Repository, dans le menu IBM® SPSS® Modeler principal, cliquez sur : Outils > IBM® SPSS® Collaboration and Deployment Services Repository > Explorer
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.
- ▶ Vérifiez que l'onglet Dossiers est actif.
- ▶ Pour créer un nouveau dossier, cliquez avec le bouton droit de la souris sur le dossier parent et cliquez sur Nouveau dossier.
- ▶ Pour renommer un dossier, cliquez dessus avec le bouton droit de la souris et cliquez sur Renommer le dossier.
- ▶ Pour supprimer un dossier, cliquez dessus avec le bouton droit de la souris et cliquez sur Supprimer le dossier.

Verrouillage et déverrouillage d'objets du IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez verrouiller un objet pour empêcher les autres utilisateurs de mettre à jour ses versions existantes ou de créer de nouvelles versions. Un objet verrouillé est indiqué par un symbole de cadenas sur l'icône de l'objet.

Figure 9-16
Objet verrouillé



Pour verrouiller un objet

- ▶ Dans la fenêtre de l'explorateur du IBM® SPSS® Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet souhaité.
- ▶ Cliquez sur Verrouiller.

Pour déverrouiller un objet

- ▶ Dans la fenêtre de l'explorateur du IBM SPSS Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet souhaité.
- ▶ Cliquez sur Déverrouiller.

Suppression d'objets du IBM SPSS Collaboration and Deployment Services Repository

Avant de supprimer un objet du IBM® SPSS® Collaboration and Deployment Services Repository, vous devez choisir de supprimer toutes les versions ou seulement une version particulière de l'objet.

Pour supprimer toutes les versions d'un objet :

- ▶ Dans la fenêtre de l'explorateur du IBM SPSS Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet souhaité.
- ▶ Cliquez sur Supprimer les objets.

Pour supprimer la version la plus récente d'un objet :

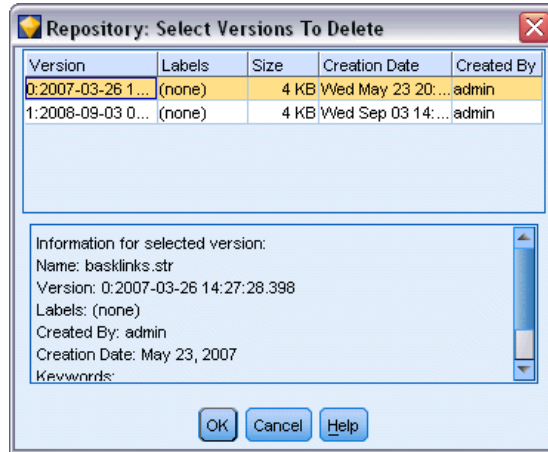
- ▶ Dans la fenêtre de l'explorateur du IBM SPSS Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet souhaité.
- ▶ Cliquez sur Supprimer.

Pour supprimer une version antérieure d'un objet :

- ▶ Dans la fenêtre de l'explorateur du IBM SPSS Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet souhaité.
- ▶ Cliquez sur Supprimer les versions.

- Sélectionnez les versions à supprimer et cliquez sur OK.

Figure 9-17
Sélectionnez les versions à supprimer



Gestion des propriétés des objets du IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez contrôler différentes propriétés d'objets depuis IBM® SPSS® Modeler. Vous pouvez :

- afficher les propriétés d'un dossier
- Afficher et modifier les propriétés d'un objet
- Créer, appliquer et supprimer les étiquettes de version d'un objet

Affichage des propriétés d'un dossier

Pour visualiser les propriétés d'un dossier dans la fenêtre du IBM® SPSS® Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur le dossier requis. Cliquez sur Propriétés du dossier.

Onglet Général

Figure 9-18
Propriétés du dossier

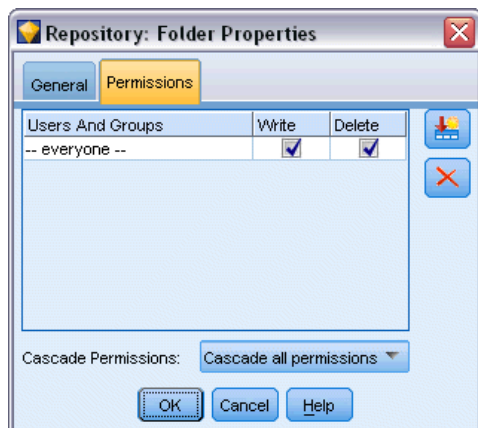


Affiche le nom du dossier et ses dates de création et de modification.

Onglet Autorisations

Indique les droits d'accès en lecture et en écriture du dossier. Tous les utilisateurs et les groupes ayant accès au dossier parent sont répertoriés. Les autorisations respectent une hiérarchie. Par exemple, si vous ne disposez pas d'un droit d'accès en lecture, vous ne pouvez pas bénéficier d'un droit d'accès en écriture. De même, si vous ne disposez pas d'un droit d'accès en écriture, vous ne pouvez pas bénéficier d'une autorisation de suppression.

Figure 9-19
Propriétés du dossier



Utilisateurs et groupes. Répertorie les utilisateurs et groupes d'utilisateurs du référentiel qui ont au moins les droits d'accès en lecture à ce dossier. Sélectionnez les cases Ecriture et Supprimer pour ajouter à un utilisateur ou à un groupe spécifique les droits d'accès à ce dossier. Cliquez sur l'icône Ajouter les utilisateurs / les groupes située à droite de l'onglet Autorisations pour attribuer l'accès à des utilisateurs et à des groupes supplémentaires. La liste des utilisateurs et des groupes disponibles est contrôlée par l'administrateur.

Droits d'accès en cascade. Choisissez une option pour contrôler la façon dont les modifications effectuées sur le dossier actuel sont appliquées à ses dossiers enfants, s'il en existe.

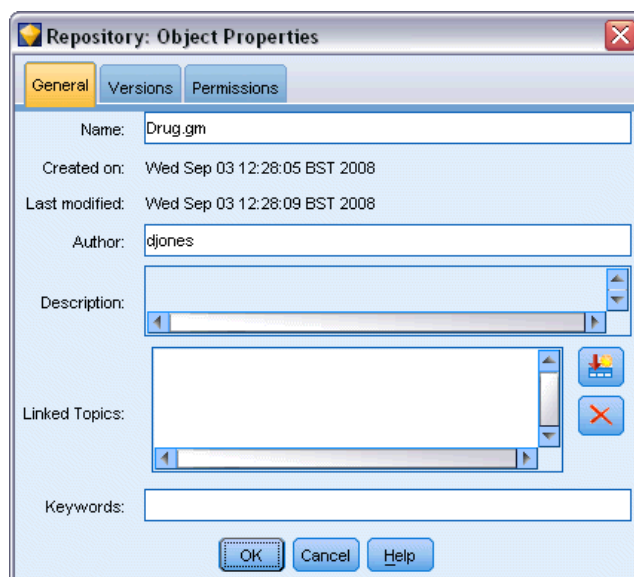
- **Tous les droits d'accès en cascade.** Met en cascade les paramètres de droits d'accès du dossier actuel pour tous les dossiers enfant et inférieurs. C'est un moyen rapide de définir les autorisations de plusieurs dossiers simultanément. Définissez les autorisations requises pour le dossier parent, puis mettez-le en cascade le cas échéant.
- **Uniquement les modifications en cascade.** Met uniquement en cascade les modifications apportées depuis la dernière validation des modifications. Par exemple, si un nouveau groupe a été ajouté et que vous voulez lui donner accès à tous les dossiers de la branche Ventes, vous pouvez lui donner accès au dossier racine Ventes, puis mettre cette modification en cascade pour tous les sous-dossiers. Tous les autres droits d'accès aux sous-dossiers existants restent inchangés.
- **Ne pas mettre en cascade.** Toutes les modifications apportées s'appliquent uniquement au dossier actuel et ne sont pas répercutées en cascade sur les dossiers enfant.

Affichage et modification des propriétés des objets

La boîte de dialogue Propriétés de l'objet du IBM® SPSS® Collaboration and Deployment Services Repository vous permet d'afficher et d'éditer les propriétés. Même si certaines propriétés ne sont pas modifiables, vous pouvez mettre un objet à jour en ajoutant une nouvelle version.

- ▶ Dans la fenêtre du IBM SPSS Collaboration and Deployment Services Repository, cliquez avec le bouton droit de la souris sur l'objet requis.
- ▶ Cliquez sur Propriétés de l'objet.

Figure 9-20
Propriétés de l'objet



Onglet Général

Nom. Nom de l'objet tel qu'il apparaît dans le IBM SPSS Collaboration and Deployment Services Repository.

Créé le. Date de création de l'objet (et non de la version).

Dernière modification. Date de modification de la version la plus récente.

Auteur. Nom de connexion de l'utilisateur.

Description : Par défaut, ce champ contient la description indiquée dans l'onglet Annotation de l'objet dans IBM® SPSS® Modeler.

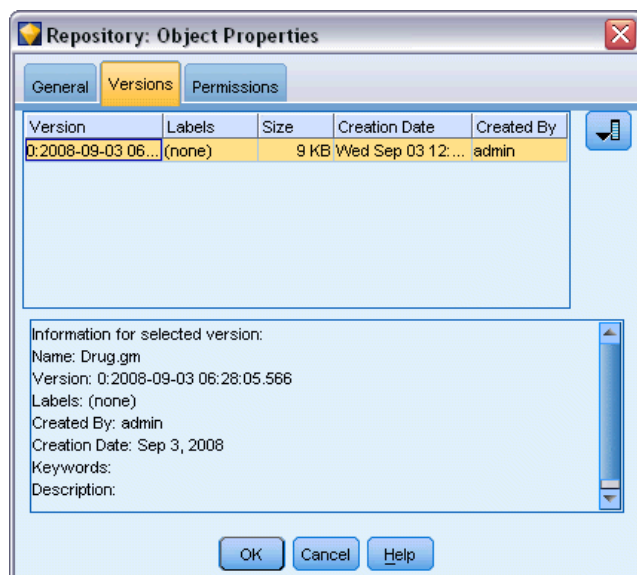
Rubriques connexes. Le IBM SPSS Collaboration and Deployment Services Repository permet l'organisation des modèles et des objets liés en rubriques le cas échéant. La liste des rubriques disponibles est définie par des utilisateurs du référentiel avec les privilèges appropriés (pour plus d'informations, consultez le *Guide de l'utilisateur de Deployment Manager*).

Mots-clés. Définissez les mots-clés dans l'onglet Annotation d'un flux, d'un modèle ou d'un objet de sortie. Les mots-clés sont séparés par des espaces et sont limités à 255 caractères. (Si les mots-clés comportent des espaces, utilisez des guillemets pour les séparer.)

Onglet Versions

Les objets stockés dans le IBM SPSS Collaboration and Deployment Services Repository peuvent avoir plusieurs versions. L'onglet Versions affiche des informations sur chaque version.

Figure 9-21
Propriétés de la version



Les propriétés suivantes peuvent être spécifiées ou modifiées pour certaines versions spécifiques d'un objet stocké :

Version. Identificateur unique de la version, généré selon l'heure à laquelle la version a été stockée.

Étiquette. Etiquette actuelle de la version. Contrairement à l'identificateur de la version, vous pouvez déplacer les étiquettes d'une version d'un objet à une autre.

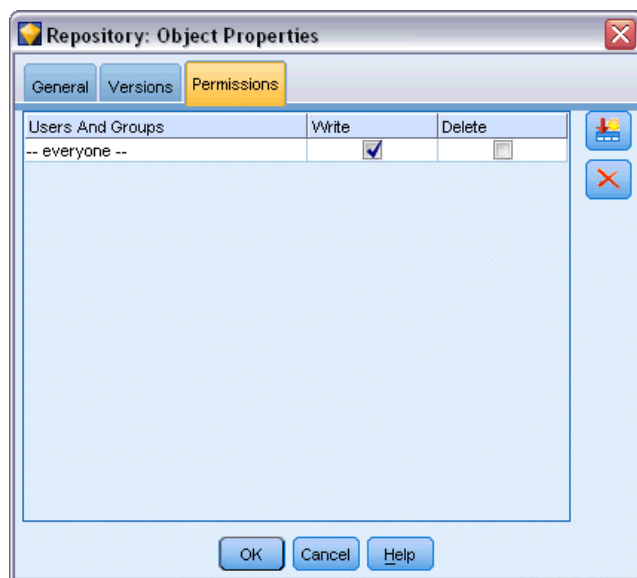
La taille de fichier, la date de création et l'auteur sont également indiqués pour chaque version.

Modifier les étiquettes. Cliquez sur l'icône Modifier les étiquettes en haut à droite de l'onglet Versions pour définir, appliquer ou supprimer des étiquettes pour des objets stockés. [Pour plus d'informations, reportez-vous à la section Gestion des étiquettes de version d'objet sur p. 186.](#)

Onglet Autorisations

L'onglet Autorisations vous permet de définir les droits d'accès en lecture et en écriture de l'objet. Tous les utilisateurs et les groupes ayant accès à l'objet actuel sont répertoriés. Les autorisations respectent une hiérarchie. Par exemple, si vous ne disposez pas d'un droit d'accès en lecture, vous ne pouvez pas bénéficier d'un droit d'accès en écriture. De même, si vous ne disposez pas d'un droit d'accès en écriture, vous ne pouvez pas bénéficier d'une autorisation de suppression.

Figure 9-22
Droits d'accès à un objet



Utilisateurs et groupes. Répertorie les utilisateurs et groupes d'utilisateurs du référentiel qui ont au moins les droits d'accès en lecture à cet objet. Sélectionnez les cases Ecriture et Supprimer pour ajouter à un utilisateur ou à un groupe spécifique les droits d'accès à cet objet. Cliquez sur l'icône Ajouter les utilisateurs / les groupes située à droite de l'onglet Autorisations pour attribuer l'accès à des utilisateurs et à des groupes supplémentaires. La liste des utilisateurs et des groupes disponibles est contrôlée par l'administrateur.

Gestion des étiquettes de version d'objet

La boîte de dialogue Modifier les étiquettes de version permet d'effectuer les opérations suivantes :

- appliquer des étiquettes à l'objet sélectionné

- Supprimer des étiquettes de l'objet sélectionné
- Définir une nouvelle étiquette et l'appliquer à l'objet

Pour appliquer des étiquettes à l'objet sélectionné

- ▶ Sélectionnez une ou plusieurs étiquettes dans la liste Etiquettes disponibles.
- ▶ Cliquez sur la flèche droite pour déplacer les étiquettes sélectionnées vers la liste Etiquettes appliquées.
- ▶ Cliquez sur OK.

Pour supprimer des étiquettes de l'objet

- ▶ Sélectionnez une ou plusieurs étiquettes dans la liste Etiquettes appliquées.
- ▶ Cliquez sur la flèche gauche pour déplacer les étiquettes sélectionnées vers la liste Etiquettes disponibles.
- ▶ Cliquez sur OK.

Pour définir une nouvelle étiquette et l'appliquer à l'objet

- ▶ Saisissez le nom de l'étiquette dans le champ Nouvelle étiquette.
- ▶ Cliquez sur la flèche droite pour déplacer la nouvelle étiquette vers la liste Etiquettes appliquées.
- ▶ Cliquez sur OK.

Déploiement de flux

Pour qu'un flux puisse être utilisé avec l'application client léger IBM® SPSS® Modeler Advantage, il doit être déployé en tant que flux (fichier *.str*) dans le référentiel.

Pour tirer le meilleur parti des fonctionnalités professionnelles de IBM® SPSS® Collaboration and Deployment Services, un flux doit être déployé en tant que scénario (fichier *.scn*) dans le référentiel. [Pour plus d'informations, reportez-vous à la section Stockage et et déploiement des objets IBM SPSS Collaboration and Deployment Services Repository sur p. 161.](#)

Pour déployer le flux actuel (méthode du menu Fichier)

- ▶ Dans le menu principal, cliquez sur :
Fichier > Stocker > Déployer
- ▶ Choisissez le type de déploiement et suivez les autres instructions de la boîte de dialogue si nécessaire.
- ▶ Cliquez sur Déployer en tant que flux pour déployer le flux à utiliser avec IBM SPSS Modeler Advantage, ou Déployer en tant que scénario pour déployer le flux à utiliser avec IBM SPSS Collaboration and Deployment Services.
- ▶ Cliquez sur le bouton Stocker. Pour plus d'informations, cliquez sur le bouton Aide.

- ▶ Continuez depuis la rubrique « Terminer le processus de déploiement ».

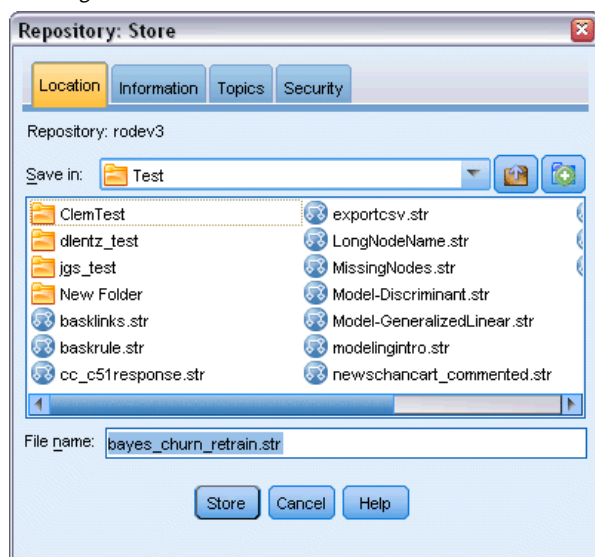
Pour déployer le flux actuel (méthode du menu Outils)

- ▶ Dans le menu principal, cliquez sur :
Outils > Propriétés du flux > Déploiement
- ▶ Choisissez le type de déploiement, remplissez le reste de l'onglet Déploiement en fonction des besoins et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Options de déploiement des flux sur p. 188.](#)

Terminer le processus de déploiement

- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.

Figure 9-23
Stockage d'un flux dans le référentiel



- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet sur p. 166.](#)

Options de déploiement des flux

L'onglet Déploiement dans la boîte de dialogue Options du flux vous permet de spécifier les options de déploiement du flux. Vous pouvez le déployer en tant que flux ou en tant que scénario.

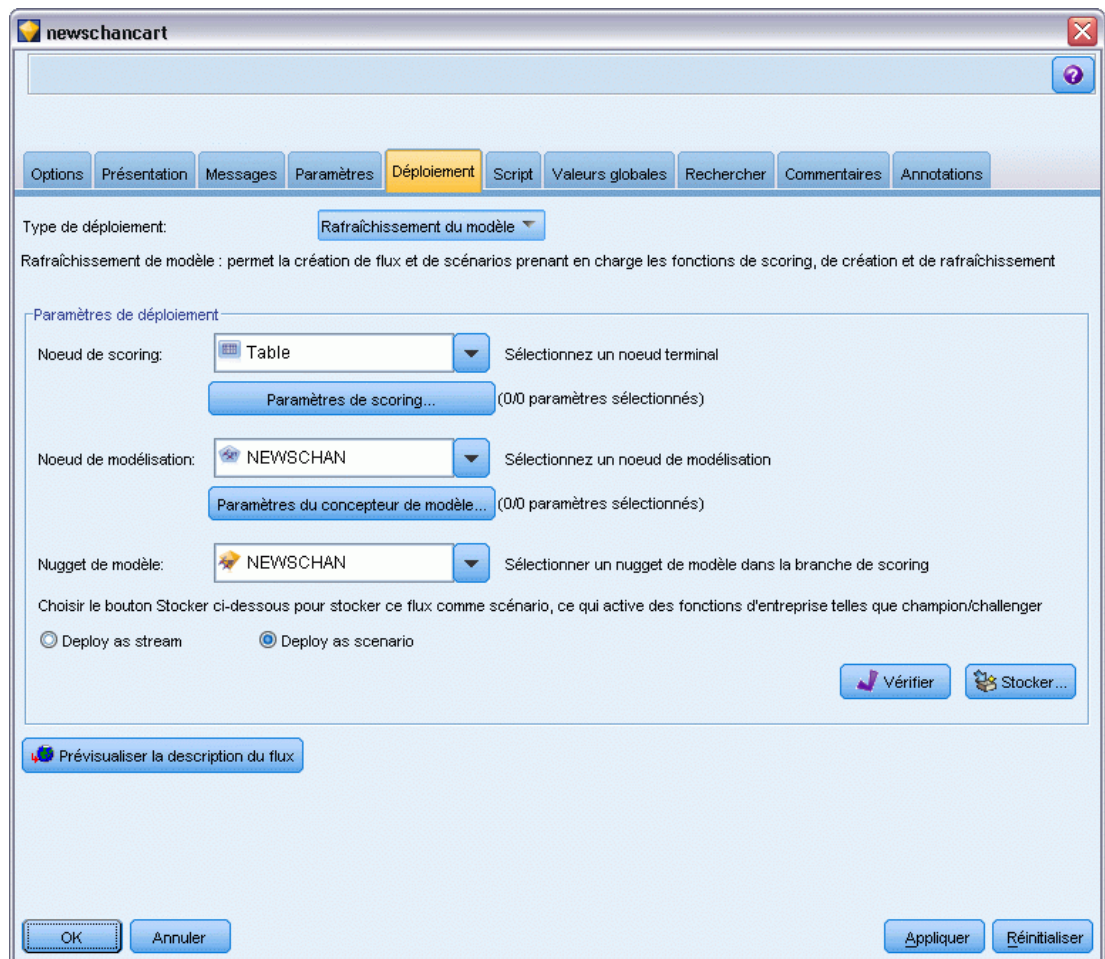
Lorsque vous déployez en tant que flux, vous pouvez ouvrir et modifier le flux dans l'application client léger IBM® SPSS® Modeler Advantage. Le flux est stocké dans le référentiel en tant que fichier avec l'extension *.str*.

Lorsque vous déployez le flux en tant que scénario, vous pouvez profiter des fonctionnalités supplémentaires disponibles avec IBM® SPSS® Collaboration and Deployment Services, comme l'accès multi-utilisateur, le scoring automatique, le rafraîchissement de modèle et l'analyse champion/challenger. Le déploiement d'un flux en tant que scénario stocke le flux dans le référentiel en tant que fichier avec l'extension *.scn*. Cela permet également d'utiliser un flux avec Predictive Applications.

A partir de cet onglet, vous pouvez aussi prévisualiser la description du flux créé par IBM® SPSS® Modeler pour le flux. [Pour plus d'informations, reportez-vous à la section Descriptions de flux dans le chapitre 5 sur p. 65.](#)

Remarque : Pour garantir un accès régulier aux données d'entreprise, un flux déployé comme scénario doit accéder à ses données source via IBM® SPSS® Collaboration and Deployment Services Enterprise View, et dans ce cas, le flux doit comprendre au moins un noeud Enterprise View source dans chaque branche de scoring ou de modélisation, suivant le cas. [Pour plus d'informations, reportez-vous à la section Noeud Enterprise View dans le chapitre 2 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Figure 9-24
Options de déploiement des flux



Type de déploiement. Sélectionnez la façon dont vous souhaitez déployer le flux. Avant d’être déployés en tant que scénarios, tous les flux nécessitent un noeud de scoring préalablement défini. Les exigences et options supplémentaires dépendent du type de déploiement.

- <aucun>. Le flux ne sera pas déployé dans le référentiel. Toutes les options sont désactivées à l’exception de l’aperçu de la description du flux.
- Uniquement le scoring. Le flux est déployé dans le référentiel lorsque vous cliquez sur le bouton Stocker. Il est possible de stocker des données à l’aide du noeud que vous avez désigné dans le champ Noeud de scoring.
- Rafraîchissement de modèle. Identique à Uniquement le scoring, mais en outre le modèle peut être mis à jour dans le référentiel à l’aide des objets que vous avez désignés dans les champs Noeud de modélisation et Nugget de modèle.

Sélectionner le noeud d’évaluation. Sélectionnez un diagramme, un noeud de sortie ou d’exportation pour identifier la branche du flux à utiliser pour le scoring des données. Bien que le flux puisse en réalité contenir n’importe quel nombre de branches, de modèles et de

noeuds terminaux valides, une seule et unique branche de scoring doit être indiquée à des fins de déploiement. C'est la condition de base requise pour déployer un flux.

Paramètres de scoring. Permet d'indiquer des paramètres pouvant être modifiés lors de l'exécution de la branche de scoring. [Pour plus d'informations, reportez-vous à la section Paramètres de scoring et de modélisation sur p. 192.](#)

Noeud de modélisation. Pour le rafraîchissement de modèle, spécifie le noeud de modélisation utilisé pour régénérer ou mettre à jour le modèle dans le référentiel. Doit être un noeud de modélisation du même type que celui spécifié pour le Nugget de modèle.

Paramètres du concepteur de modèle. Permet d'indiquer des paramètres pouvant être modifiés lors de l'exécution du noeud de modélisation. [Pour plus d'informations, reportez-vous à la section Paramètres de scoring et de modélisation sur p. 192.](#)

Nugget de modèle. Dans le cas d'un rafraîchissement de modèle, indique le nugget de modèle à mettre à jour ou à régénérer à chaque mise à jour du scénario dans le référentiel (généralement dans le cadre d'une tâche planifiée). Le modèle doit se trouver sur la branche de scoring. La branche de scoring peut comporter plusieurs modèles, mais vous ne pouvez en désigner qu'un seul. Au moment de la création du scénario, il peut s'agir en fait d'un modèle substituable mis à jour ou régénéré chaque fois que des nouvelles données sont disponibles.

Déployer comme flux. Cliquez sur cette option pour utiliser le flux avec IBM SPSS Modeler Advantage.

Déployer en tant que scénario. Cliquez sur cette option si vous souhaitez utiliser le flux avec IBM SPSS Collaboration and Deployment Services (et consultez la note qui suit).

Vérifier. Cliquez sur ce bouton pour vérifier s'il s'agit d'un flux valide pour le déploiement. Si vous déployez en tant que scénario, le flux doit contenir un noeud source Enterprise View valide ; un message d'erreur s'affiche si aucun n'est présent.

Stocker. Déploie le flux en tant que scénario si le flux est valide. Dans le cas contraire, un message d'erreur apparaît. Cliquez sur le bouton Réparer, corrigez l'erreur et réessayez.

Aperçu de la description du flux. Vous permet d'afficher le contenu de la description du flux créé par SPSS Modeler pour le flux. [Pour plus d'informations, reportez-vous à la section Descriptions de flux dans le chapitre 5 sur p. 65.](#)

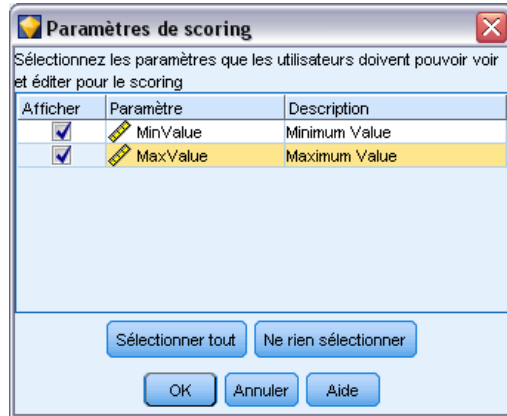
Remarque : (Déployer en tant que scénario uniquement) Plusieurs noeuds Enterprise View peuvent être utilisés dans la branche de modélisation. Dans ce cas, il est souvent préférable d'utiliser une seule connexion de données pour tous les noeuds Enterprise View de la branche. En revanche, cela est obligatoire pour l'analyse champion/challenger.

- Si la prise en charge de la fonction champion/challenger n'est pas requise, différentes connexions Enterprise View peuvent être utilisées dans la même branche, dans la mesure où les connexions ne diffèrent que par la Définition de fournisseur de données (DPD).
- Ces limites ne s'appliquent qu'au sein d'une branche donnée. Entre les branches de création de modèle et de scoring, différentes connexions Enterprise View peuvent être utilisées sans tenir compte de ces restrictions.

Paramètres de scoring et de modélisation

Lorsque vous déployez un flux dans IBM® SPSS® Collaboration and Deployment Services, vous pouvez choisir les paramètres pouvant être affichés et modifiés chaque fois que le modèle est mis à jour ou évalué. Par exemple, il est possible d'indiquer des valeurs minimum ou maximum ou une autre valeur susceptible de varier à chaque exécution d'une tâche.

Figure 9-25
Boîte de dialogue Paramètres de scoring



- Pour rendre un paramètre visible afin que vous puissiez l'afficher ou le modifier une fois le scénario déployé, sélectionnez-le dans la liste de la boîte de dialogue.

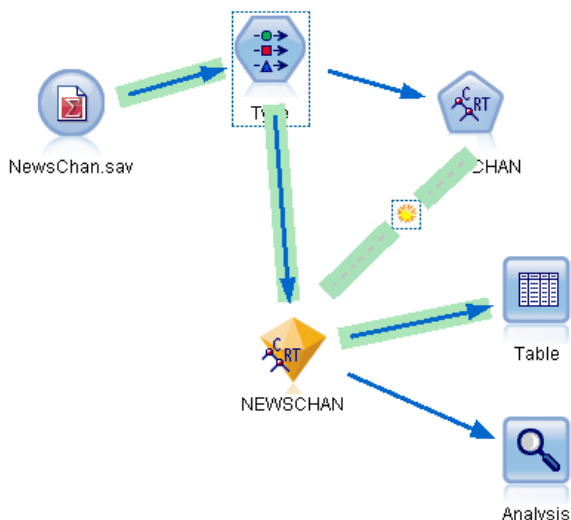
La liste des paramètres disponibles figure dans l'onglet Paramètres de la boîte de dialogue Propriétés du flux. [Pour plus d'informations, reportez-vous à la section Définition des paramètres de flux et de session dans le chapitre 5 sur p. 58.](#)

La branche de scoring

Si vous déployez un flux, une branche de ce flux doit être désignée comme **branche de scoring** (c'est-à-dire celle contenant le noeud de scoring). Lorsque vous désignez une branche comme branche de scoring, cette branche est mise en évidence dans l'espace de travail, tout comme le lien du modèle vers le nugget de la branche de scoring. Cette représentation visuelle est particulièrement utile dans les flux complexes avec plusieurs branches, où la branche de scoring n'est pas immédiatement évidente.

Remarque : Un seul flux peut être désigné comme branche de scoring.

Figure 9-26
Le flux avec la branche de scoring est mis en évidence



Si le flux contient déjà une branche de scoring définie, la nouvelle branche désignée la remplace comme branche de scoring. Vous pouvez définir la couleur de l'indication de la branche de scoring à l'aide de l'option Couleur personnalisée. [Pour plus d'informations, reportez-vous à la section Définition des options d'affichage dans le chapitre 12 sur p. 235.](#)

Vous pouvez afficher ou masquer l'indication de la branche de scoring à l'aide du bouton de la barre d'outils Afficher/Masquer le balisage des flux.

Figure 9-27
Bouton de barre d'outils Afficher/Masquer le balisage des flux



Identification de la branche de scoring pour le déploiement

Vous pouvez désigner la branche de scoring depuis le menu contextuel d'un nœud terminal ou depuis le menu Outils. Si vous utilisez le menu contextuel, le nœud de scoring est défini automatiquement dans l'onglet Déploiement des propriétés du flux.

Pour désigner une branche comme branche de scoring (menu contextuel)

- ▶ Connectez le nugget de modèle à un nœud terminal (un nœud d'exécution ou de sortie en aval du nugget).
- ▶ Faites un clic droit sur le nœud terminal.
- ▶ Dans le menu, cliquez sur Utiliser en tant que branche de scoring.

Pour désigner une branche comme branche de scoring (menu Outils)

- ▶ Connectez le nugget de modèle à un noeud terminal (un noeud d'exécution ou de sortie en aval du nugget).
- ▶ Dans le menu principal, cliquez sur :
Outils > Propriétés du flux > Déploiement
- ▶ Dans la liste des Type de déploiement, cliquez sur Scoring uniquement ou Rafraîchissement de modèle selon les besoins. [Pour plus d'informations, reportez-vous à la section Options de déploiement des flux sur p. 188.](#)
- ▶ Cliquez sur le champ Noeud de scoring et sélectionnez un noeud terminal dans la liste.
- ▶ Cliquez sur OK.

Rafraîchissement de modèle

Le rafraîchissement de modèle est le processus qui permet de reconstruire un modèle existant dans un scénario à l'aide de données plus récentes. Le flux lui-même n'est pas modifié dans le référentiel. Par exemple, le type d'algorithme et les paramètres spécifiques au flux restent les mêmes mais le modèle est de nouveau formé à partir des nouvelles données et mis à jour si la nouvelle version du modèle fonctionne mieux que l'ancienne.

Un seul nugget de modèle dans un scénario peut être défini pour être rafraîchi : il s'appelle le **modèle de rafraîchissement**. Si vous cliquez sur l'option Rafraîchissement de modèle dans l'onglet Déploiement des propriétés du flux (voir Options de déploiement des flux sur p. 188), le nugget de modèle que vous désignez à ce moment devient le modèle de rafraîchissement. Vous pouvez également désigner un modèle comme modèle de rafraîchissement à partir du menu contextuel d'un nugget de modèle. Pour ce faire, le nugget doit déjà être sur la branche de scoring.

Si vous désactivez le statut « modèle de rafraîchissement » d'un nugget, cela revient à définir le type de déploiement du scénario sur Scoring uniquement et l'onglet Déploiement de la boîte de dialogue des propriétés du flux est mis à jour en conséquence. Vous pouvez activer et désactiver ce statut au moyen de l'option Utiliser comme modèle de rafraîchissement dans le menu contextuel du nugget dans la branche de scoring actuelle.

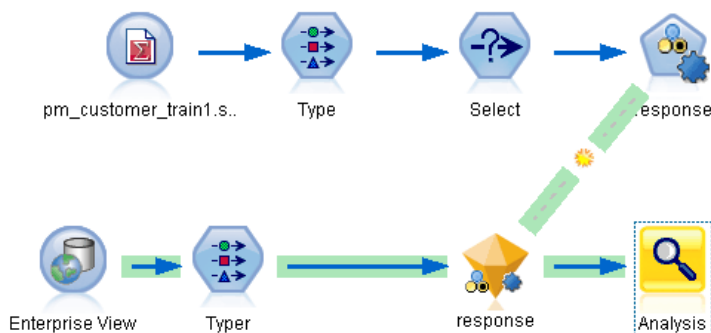
Supprimer le lien de modèle d'un nugget dans la branche de scoring supprime également le statut de « modèle de rafraîchissement » du nugget. Vous pouvez annuler la suppression du lien de modèle avec le menu Edition ou la barre d'outils ce qui restaure le statut du « modèle de rafraîchissement » du nugget.

Sélection du modèle de rafraîchissement

Tout comme la branche de scoring, le lien vers le modèle de rafraîchissement est également mis en évidence dans le flux. Le nugget de modèle choisi comme modèle de rafraîchissement, et par conséquent le lien mis en évidence, dépend du nombre de nuggets dans le flux.

Modèle unique dans le flux

Figure 9-28
Branche de scoring avec modèle unique dans le flux



Si un unique nugget de modèle lié se trouve dans la branche de scoring lorsqu'il est identifié comme tel, ce nugget devient le modèle de rafraîchissement pour le scénario.

Modèles multiples dans le flux

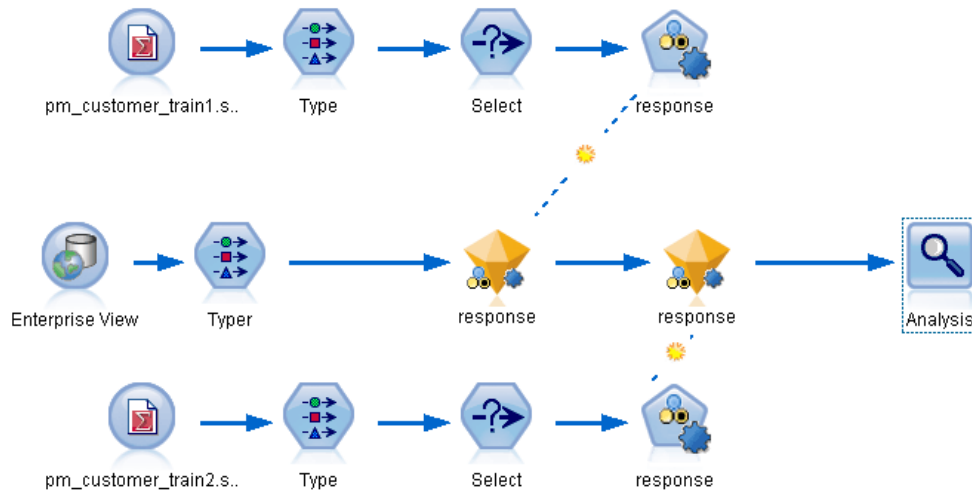
S'il existe plusieurs nuggets liés dans le flux, le modèle de rafraîchissement est choisi comme suit :

Si un nugget de modèle a été défini dans l'onglet Déploiement de la boîte de dialogue des propriétés du flux et se trouve également dans le flux, alors le nugget devient le modèle de rafraîchissement.

Si aucun nugget n'a été défini dans l'onglet Déploiement, ou si un nugget a été défini mais ne se trouve pas dans la branche de scoring, alors le nugget le plus proche du noeud terminal devient le modèle de rafraîchissement.

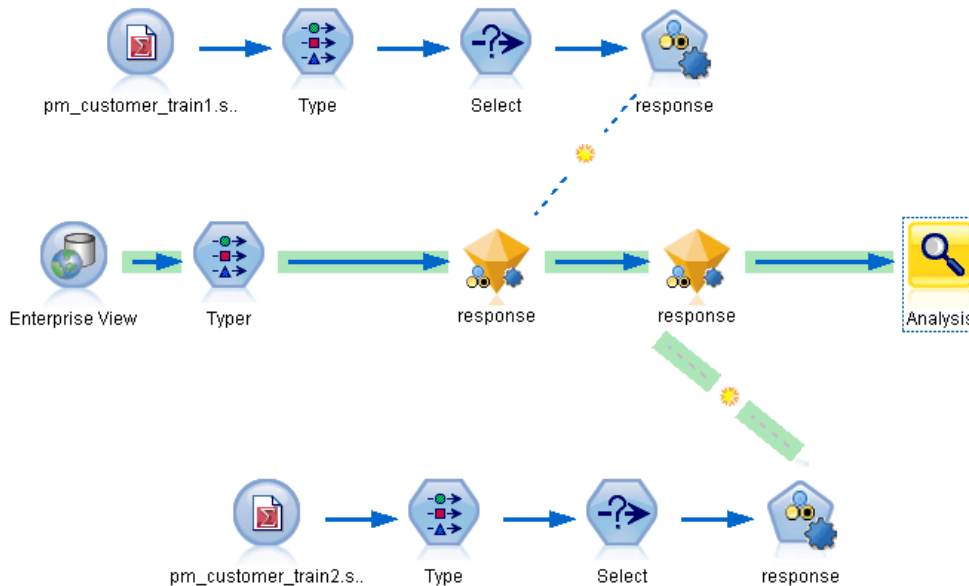
Pour illustrer cela, imaginons le flux suivant.

Figure 9-29
Branche de scoring avec plusieurs modèles dans le flux



Faites un clic droit sur le noeud Analyse et utilisez son menu pour définir la branche de scoring qui est maintenant mise en évidence. Cette action désigne également le modèle le plus proche du noeud Analyse comme modèle de rafraîchissement, comme l'indique le lien de rafraîchissement mis en évidence.

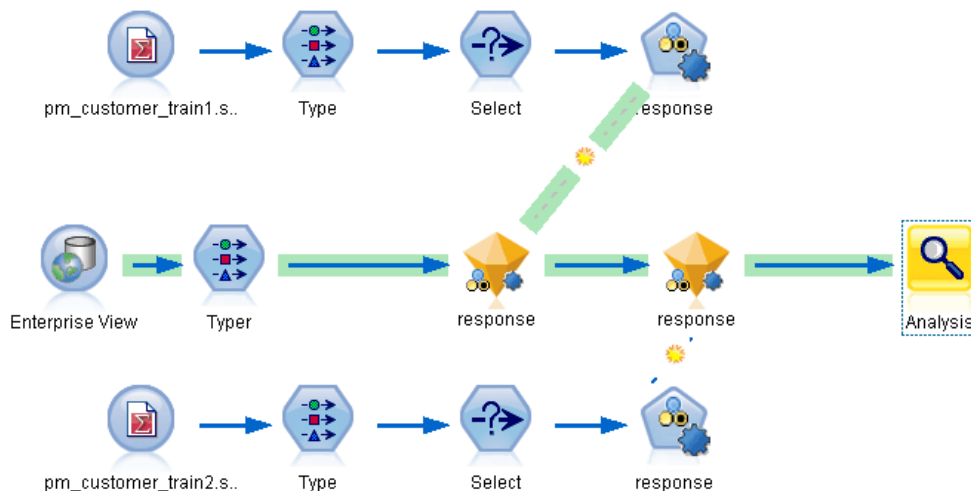
Figure 9-30
Branche de scoring mise en évidence avec modèles multiples et lien de rafraîchissement



Cependant, si vous décidez d'utiliser l'autre nugget de modèle dans le flux comme modèle de rafraîchissement, vous définissez alors dans son menu son lien de modèle pour qu'il soit utilisé comme lien de rafraîchissement.

Figure 9-31

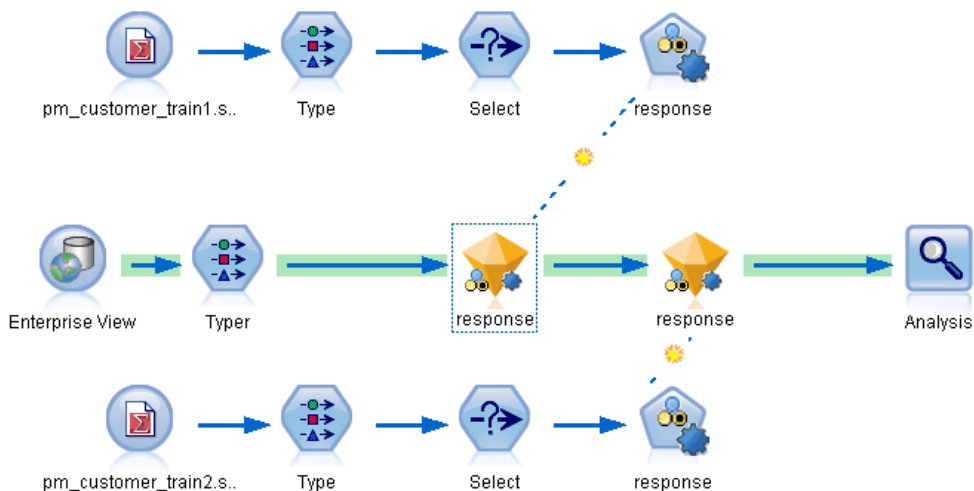
Branche de scoring avec lien de rafraîchissement transformé en premier nugget de modèle



Si ensuite, vous désélectionnez les deux liens de modèle définis comme liens de rafraîchissement, seule la branche de scoring est mise en évidence et pas les liens. Le type de déploiement est défini sur Scoring uniquement.

Figure 9-32

Branche de scoring avec modèles multiples et aucun lien de rafraîchissement



Remarque : Vous pouvez choisir de définir un des liens sur le statut Remplacer mais pas les deux. Dans ce cas, le nugget de modèle choisi comme modèle de rafraîchissement est celui qui contient un lien de rafraîchissement et qui est le plus proche du noeud terminal lorsque la branche de scoring est désignée.

Aucun modèle dans le flux

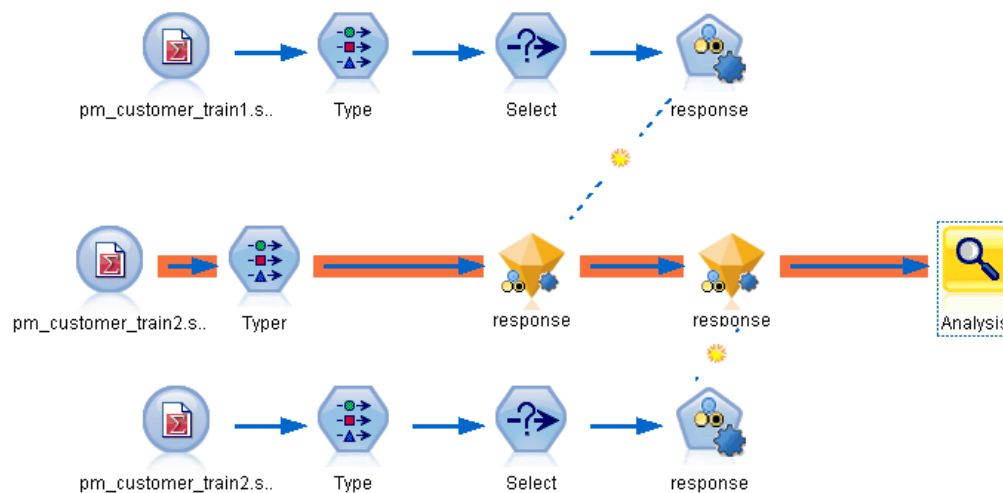
S'il n'existe pas de modèles dans le flux, ou uniquement des modèles sans lien de modèle, le type de déploiement est défini sur Scoring uniquement.

Vérification des erreurs dans une branche de scoring

Lorsque vous désignez la branche de scoring, des erreurs y sont recherchées, afin de ne pas se retrouver avec un noeud Enterprise View dans le flux lors du déploiement du scénario.

Figure 9-33

Branche de scoring avec erreurs



Si une erreur est trouvée, la branche de scoring est mise en évidence dans la couleur d'erreur des branches de scoring et un message d'erreur apparaît. Vous pouvez définir la couleur de l'erreur à l'aide de l'option Couleur personnalisée. [Pour plus d'informations, reportez-vous à la section Définition des options d'affichage dans le chapitre 12 sur p. 235.](#)

Si une erreur est trouvée, suivez les instructions suivantes :

- ▶ Corrigez l'erreur en fonction du contenu du message d'erreur.
- ▶ Effectuez l'une des actions suivantes :
 - Cliquez avec le bouton droit de la souris sur le noeud terminal et cliquez sur Vérifier le scénario dans le menu contextuel.
 - Dans le menu principal, cliquez sur : Outils > Propriétés du flux > Déploiement

puis cliquez sur le bouton Vérifier.

- ▶ Si nécessaire, répétez ce processus jusqu'à ne plus trouver d'erreurs.

Exportation vers des applications externes

À propos de l'exportation vers des applications externes

IBM® SPSS® Modeler offre divers mécanismes pour exporter la totalité du processus de Data mining vers des applications externes. Ainsi, les tâches de préparation des données et de création des modèles que vous effectuez dans SPSS Modeler peuvent également être utilisées dans une application externe.

La section précédente a montré comment déployer des flux dans un référentiel IBM SPSS Collaboration and Deployment Services pour bénéficier de l'accès multi-utilisateur, de la planification de tâches et d'autres fonctionnalités. De même, les flux SPSS Modeler peuvent aussi être utilisés en conjonction avec :

- IBM® SPSS® Modeler Advantage
- Les applications de Predictive Applications 5.0
- Les applications de Predictive Applications 4.x
- Les applications qui peuvent importer et exporter des fichiers au format PMML

Pour des informations supplémentaires sur l'utilisation des flux avec IBM SPSS Modeler Advantage, consultez Ouverture d'un flux dans IBM SPSS Modeler Advantage sur p. 201.

Pour exporter un flux à utiliser avec Predictive Applications 5.0, suivez les instructions de déploiement en tant que scénario. [Pour plus d'informations, reportez-vous à la section Déploiement de flux dans le chapitre 9 sur p. 187.](#)

Pour les détails concernant le regroupement des flux pour le déploiement des versions 4.x de Predictive Applications, reportez-vous à Assistant des applications prédictives 4.x sur p. 201.

Pour de plus amples informations sur les modèles d'exportation et d'importation comme fichiers PMML, la possibilité de partager des modèles avec toutes les autres applications qui prennent en charge ce format, reportez-vous à Importation et exportation de modèles au format PMML sur p. 209.

Remarque : Le produit Predictive Applications a été remplacé par IBM® SPSS® Decision Management. La prise en charge de Predictive Applications sera supprimée dans une version future de SPSS Modeler.

Ouverture d'un flux dans IBM SPSS Modeler Advantage

Les flux IBM® SPSS® Modeler peuvent s'utiliser avec l'application client léger IBM® SPSS® Modeler Advantage. Bien qu'il soit possible de créer des applications personnalisées complètes au sein de IBM SPSS Modeler Advantage, vous pouvez aussi utiliser un flux déjà créé dans SPSS Modeler comme base d'un flux de travail de l'application.

Pour ouvrir un flux dans IBM SPSS Modeler Advantage :

- ▶ Déployez le flux dans le référentiel IBM® SPSS® Collaboration and Deployment Services et veillez à cliquer sur l'option Déployer en tant que flux. [Pour plus d'informations, reportez-vous à la section Déploiement de flux dans le chapitre 9 sur p. 187.](#)
- ▶ Cliquez sur le bouton de la barre d'outils Ouvrir dans IBM SPSS Modeler Advantage ou dans le menu principal, cliquez sur :
Fichier > Ouvrir dans IBM SPSS Modeler Advantage
- ▶ Indiquez, si nécessaire, des paramètres de connexion au référentiel. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 162.](#) Pour les informations de port, de mot de passe et d'autres détails concernant la connexion, contactez votre administrateur système local.

Remarque : le logiciel IBM SPSS Modeler Advantage doit également être installé sur le serveur du référentiel.

- ▶ Dans le référentiel : Boîte de dialogue Stocker, sélectionnez le dossier où vous souhaitez enregistrer l'objet, spécifiez toutes les autres informations que vous souhaitez enregistrer et cliquez sur le bouton Stocker. [Pour plus d'informations, reportez-vous à la section Définition des propriétés d'un objet dans le chapitre 9 sur p. 166.](#)

Ainsi, IBM SPSS Modeler Advantage sera lancé et le flux sera déjà ouvert. Le flux est fermé dans SPSS Modeler.

Assistant des applications prédictives 4.x

Remarque : Le produit Predictive Applications a été remplacé par IBM® SPSS® Decision Management. La prise en charge de Predictive Applications sera supprimée dans une version future de IBM® SPSS® Modeler.

L'assistant des applications prédictives 4.x vous permet de regrouper les flux de SPSS Modeler en vue de leur déploiement avec les versions 4.x des applications prédictives, notamment PredictiveMarketing version 2.0 ou ultérieure. Toutes les opérations de manipulation et de modélisation de données effectuées dans SPSS Modeler peuvent rapidement être regroupées et enregistrées sous la forme d'une solution de scoring. Une fois déployée, l'application vous permet d'ajouter des modèles SPSS Modeler à votre solution de campagne.

Par exemple, une analyste de données souhaite segmenter sa clientèle et développer un modèle de valeur spécifique pour l'argent, l'or et le platine à utiliser avec PredictiveMarketing ou une autre application. Elle peut développer ce modèle dans SPSS Modeler, puis l'exporter afin de

s'en servir dans PredictiveMarketing, qui utilise le modèle SPSS Modeler en arrière-plan pour recommander l'offre de la plus grande valeur.

Pour accéder à l'assistant

- ▶ Dans le menu principal, cliquez sur :
Outils > Assistant des applications prédictives 4.x

Avant d'utiliser l'assistant des applications prédictives

Les informations suivantes visent à présenter l'intégration entre IBM® SPSS® Modeler et d'autres applications prédictives. Avant d'exporter un flux via l'assistant, passez en revue les mécanismes de l'intégration et les conditions préalables en vue de la publication.

Fonctionnement de l'intégration

Dans des conditions normales, voici comment vous pouvez utiliser SPSS Modeler pour étendre les fonctions de Data mining et de déploiement fournies avec les applications prédictives :

- ▶ Ouvrez l'application prédictive. A l'aide des options de Customer View Builder, exportez le modèle de données UCV (Unified Customer View affichage client unifié) en tant que fichier XML. Relevez l'emplacement du fichier XML car vous en aurez besoin pour travailler dans SPSS Modeler.
- ▶ Ensuite, définissez des noeuds source dans SPSS Modeler pour accéder à toutes les sources de données (bases de données, fichiers plats, etc.) contenant les champs de l'UCV (répertoriés dans le fichier XML exporté précédemment). Vous pouvez inclure l'ensemble des champs référencés par l'UCV ou n'utiliser que ceux dont vous avez besoin pour les modèles que vous créez. En général, vous devez disposer de plusieurs noeuds source dans SPSS Modeler pour accéder aux données de modélisation.
- ▶ Utilisez SPSS Modeler pour fusionner des données, effectuer des transformations ou des calculs nécessaires à votre travail de Data mining.
- ▶ A un stade précis du flux, veillez à ajouter un noeud Typer et à le nommer *UCV*. Ce noeud Typer *UCV* est utilisé à des fins d'orientation durant la modélisation et pour vérifier que vos données correspondent aux informations de champ définies dans le fichier XML. Il est judicieux de comparer les paramètres du noeud Typer aux spécifications d'attribut du fichier XML généré précédemment. [Pour plus d'informations, reportez-vous à la section Etape 3 : Sélection d'un noeud UCV sur p. 205.](#)
- ▶ Ensuite, passez au type de modèle que vous créez dans SPSS Modeler. Si vous déployez un modèle de valeur, tel qu'un réseau de neurones, vous pouvez exporter des prévisions binaires (par exemple, Attrition True (vrai)/False (faux)) en tant que propension, ce qui rend la prévision comparable à celles des modèles générés par l'application. [Pour plus d'informations, reportez-vous à la section Exportation de prévisions binaires en tant que scores de propension sur p. 203.](#)
- ▶ Une fois que vous êtes satisfait du modèle et que vous avez converti les confiances en propensions, ajoutez un noeud terminal à la branche de déploiement du flux. De nombreux utilisateurs se

servent d'un noeud Table, mais un noeud terminal suffit amplement. Vérifiez que les champs à afficher dans l'application externe apparaissent au niveau du noeud terminal. En d'autres termes, avant d'utiliser ce noeud terminal, vous pouvez être amené à filtrer les champs que vous ne souhaitez pas déployer.

- ▶ En outre, assurez-vous que les champs de prévision générés par le modèle sont instanciés avant d'être exportés. Si nécessaire, vous pouvez pour cela ajouter un noeud Typer entre le modèle généré et le noeud « exporté » terminal.
- ▶ Avant d'utiliser l'assistant, votre dernière tâche consiste à exécuter un test pour vérifier que le flux est prêt à être déployé.

Le flux est prêt à être déployé. Vous pouvez accéder à l'assistant des applications prédictives à partir du menu Outils de SPSS Modeler. Suivez les différentes étapes de l'assistant décrites dans la présente documentation pour créer un logiciel de déploiement Modeler (.cdp) contenant les informations de flux et les métadonnées nécessaires à la publication dans l'environnement en temps réel.

Exportation de prévisions binaires en tant que scores de propension

Avant d'effectuer une exportation à l'aide de l'assistant des applications prédictives, étudiez si la sortie de scoring du modèle IBM® SPSS® Modeler est en accord avec les prévisions générées par l'application prédictive. Dans de nombreux cas, vous pouvez exporter des prédictions binaires en tant que scores de propension, ce qui vous permet de comparer avec précision la fiabilité des prévisions dans différents modèles.

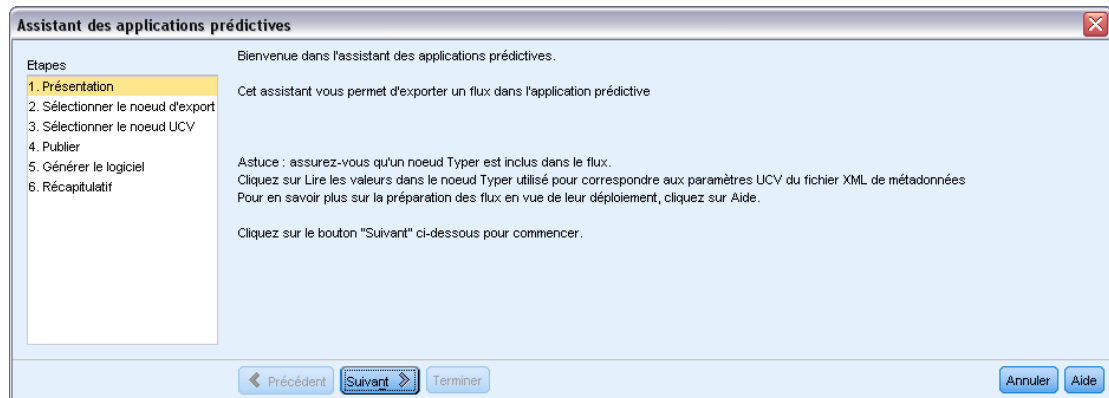
- Un **modèle de valeur** crée un **score de propension** unique pour chaque enregistrement qui évalue la probabilité d'un résultat (*oui* ou *non*) sur une échelle de 0 à 1. Par exemple, un modèle d'attrition peut produire un score se situant entre 1 (attrition probable) et 0 (attrition improbable). Les scores de propension n'étant pas des probabilités, un score de 0,5 ne signifie pas nécessairement 50 % de probabilité d'attrition, ni même deux fois plus de probabilités que pour un score de 0,25. Cela indique une *plus grande probabilité* d'attrition que pour un score de 0,4. Les scores de propension permettent de réaliser des classements ou de rechercher, par exemple, les 10 % de clients présentant la plus forte probabilité d'attrition.
- Un **modèle d'offre** crée des valeurs de prévision et de confiance qui, lorsqu'elles seront soumises à l'application, seront utilisées comme attributs virtuels dans l'UCV. Ces valeurs d'intervalle peuvent être numériques ou continues.

Plusieurs modèles SPSS Modeler permettent de générer des prévisions binaires, parmi lesquels les modèles de réseau de neurones, d'arbre décision et de régression logistique. Si vous déployez un modèle de valeur, il se peut que vous deviez exporter les scores de propension avant de procéder à l'exportation. Les scores de propension peuvent être activés sur l'onglet Analyser du noeud de modélisation ou sur l'onglet Paramètres du nugget de modèle généré. [Pour plus d'informations, reportez-vous à la section Options d'analyse des noeuds de modélisation dans le chapitre 3 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Etape 1 : Présentation de l'assistant des applications prédictives

Lorsque vous ouvrez l'assistant des applications prédictives pour la première fois, un écran de bienvenue apparaît et vous dirige vers le processus de regroupement des composants de flux nécessaires.

Figure 10-1
Assistant des applications prédictives

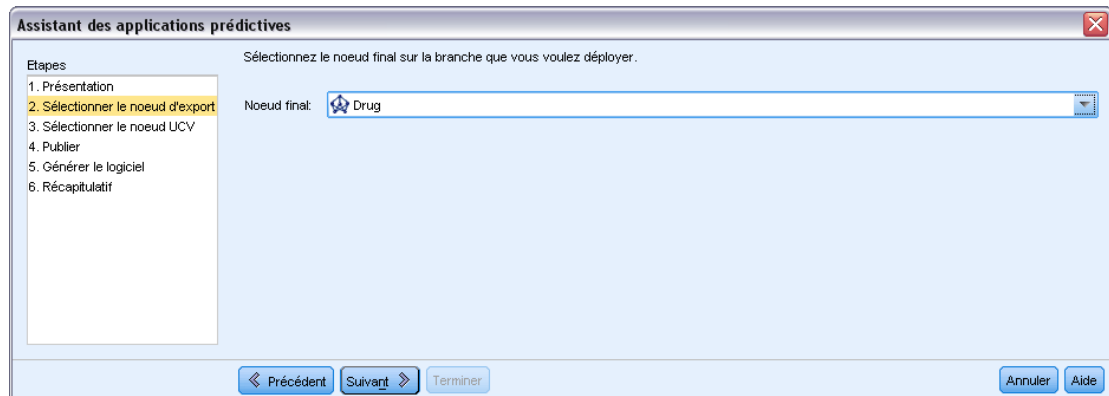


Les autres étapes de l'assistant vous guident tout au long du processus de génération d'un logiciel à déployer dans l'environnement en temps réel. Avant de poursuivre, servez-vous de la liste de contrôle préalable pour vous assurer que le flux est prêt à être déployé. [Pour plus d'informations, reportez-vous à la section Avant d'utiliser l'assistant des applications prédictives sur p. 202.](#)

Etape 2 : Sélection d'un noeud terminal

A l'étape 2, vous pouvez spécifier un noeud terminal du flux représentant la branche de scoring ou la zone de scoring du flux.

Figure 10-2
Sélection d'un noeud terminal



Il est important de distinguer les noeuds terminaux portant le même nom car la liste de l'assistant fournit uniquement le nom et le type de chaque noeud terminal du flux. Pour éviter toute confusion, attribuez un nom unique aux noeuds terminaux du flux.

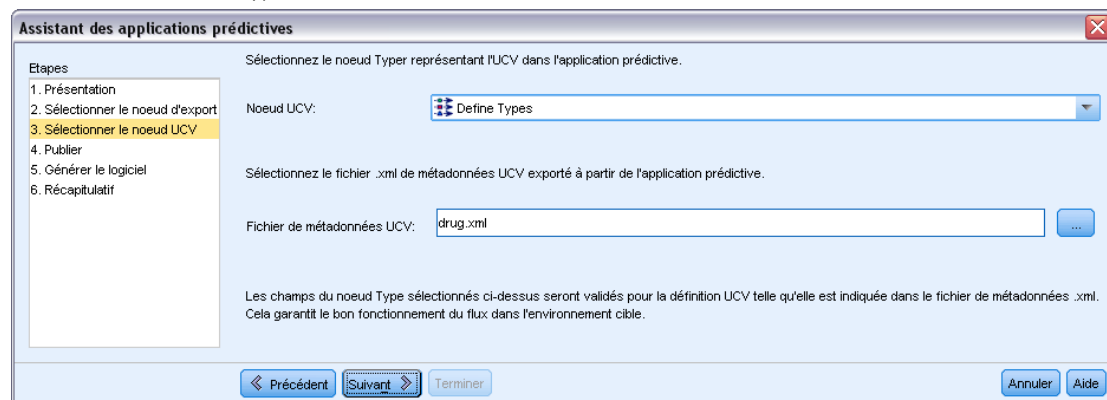
En outre, assurez-vous que seuls les champs à afficher dans l'environnement de l'application apparaissent au niveau du noeud terminal. En d'autres termes, avant d'utiliser ce noeud terminal, vous pouvez être amené à filtrer les champs que vous ne souhaitez pas déployer.

Etape 3 : Sélection d'un noeud UCV

A l'étape 3, vous fournissez deux éléments d'information essentiels à la publication, à savoir un noeud UCV et un fichier de métadonnées UCV.

Figure 10-3

Sélection d'un noeud *Typier* utilisé comme noeud UCV et d'un fichier de métadonnées UCV (fichier XML)



Noeud UCV. Un noeud UCV est un noeud *Typier* du flux que vous utilisez pour vérifier que toutes les données correspondent aux définitions de l'UCV. Ces spécifications sont stockées dans un fichier XML (exporté précédemment à partir de Customer View Builder). Lorsque vous cliquez sur *Suivant*, l'assistant valide automatiquement les paramètres du noeud *Typier* en fonction des spécifications du fichier XML indiquées ici via l'option *Fichier de métadonnées UCV*.

Fichier de métadonnées UCV. Un fichier de métadonnées UCV représente le fichier XML que vous avez généré précédemment depuis Customer View Builder. Le fichier XML que vous sélectionnez ici comporte des attributs de données nécessaires au déploiement dans l'application.

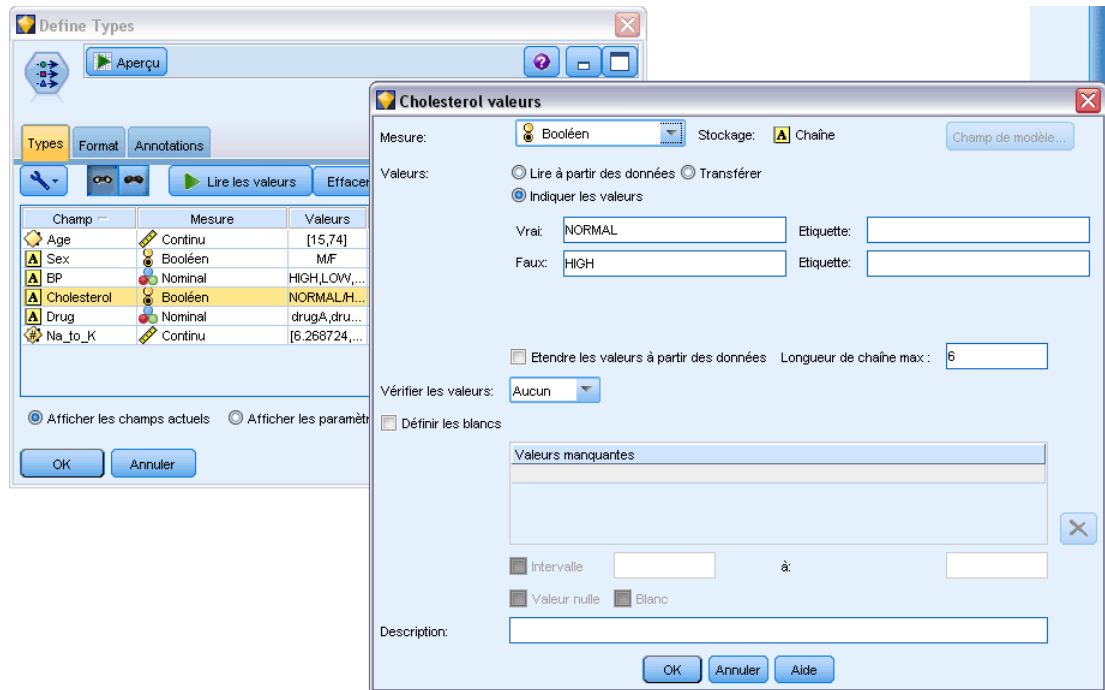
Erreurs de non-correspondance des données

Si l'assistant génère des erreurs de non-correspondance des données, revenez au flux IBM® SPSS® Modeler et vérifiez les spécifications du noeud *Typier*. Comparez les informations de champ de ce noeud *Typier* à celles incluses dans le fichier XML créé à partir de Customer View Builder. (Vous pouvez ouvrir le fichier XML dans un navigateur de texte, comme le Bloc-notes.) *Les paramètres du noeud Typier correspondent-ils aux attributs UCV ?* Par exemple, le fichier XML peut indiquer qu'un champ appelé *Cholestérol* doit être utilisé par l'UCV et qu'il contient des valeurs de chaîne :

```
<UcvAttribute Name="Cholesterol" Domain="String"/>
```

Dans le flux SPSS Modeler, vérifiez dans les paramètres du noeud *Typier* que seuls les champs utiles à l'environnement en temps réel sont exportés depuis SPSS Modeler.

Figure 10-4
Vérification du stockage et des paramètres du noeud Typier dans la sous-boîte de dialogue Valeurs



Mappage d'un stockage sur un domaine

Dans certains cas, le mappage des paramètres d'un noeud Typier et des spécifications UCV peut s'avérer complexe. Dans le cadre de cette intégration, le **domaine** (domain) défini dans le fichier XML UCV est équivalent au **stockage** (storage) dans SPSS Modeler. Les paramètres de champ de chacun de ces éléments sont mis en correspondance lors de l'exportation.

Table 10-1
Mappage d'un stockage et d'un domaine

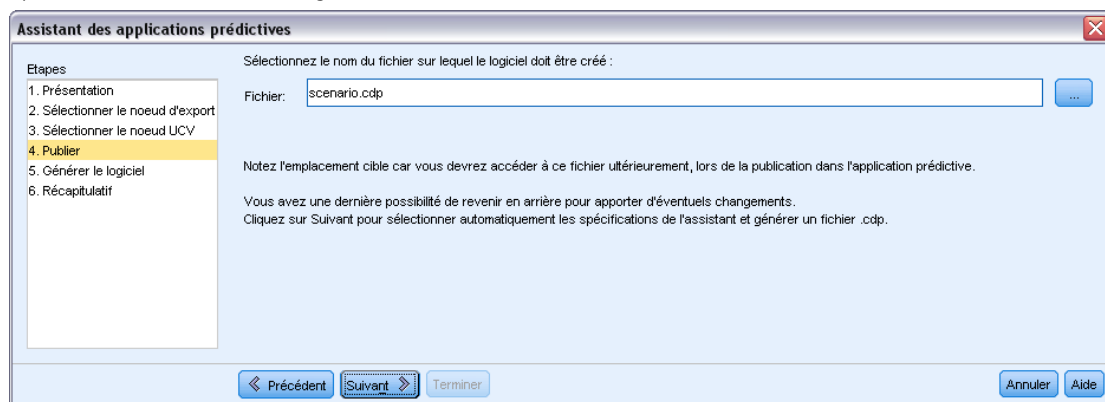
SPSS Modeler Stockage	Domaines d'application
Chaîne	Chaîne Caractère Bit
Réel	Float Double Décimal
Entier	Long Entier
Date	Date
Horodatage	Horodatage
Temps	Aucune correspondance

Vous pouvez modifier le type de stockage à l'aide des fonctions de conversion, telles que `to_integer`, dans un noeud Calculer ou Remplacer. [Pour plus d'informations, reportez-vous à la section Conversion du stockage à l'aide du noeud Remplacer dans le chapitre 4 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Etape 4 : Spécification d'un logiciel

A l'étape 4, vous devez indiquer le nom du logiciel de flux, ainsi que l'endroit où il sera enregistré. Le fait de suivre cette étape permet d'enregistrer plusieurs objets en tant que groupe appelé logiciel **Modeler Deployment Package**, doté de l'extension `.cdp`.

Figure 10-5
Spécification d'un nom de logiciel

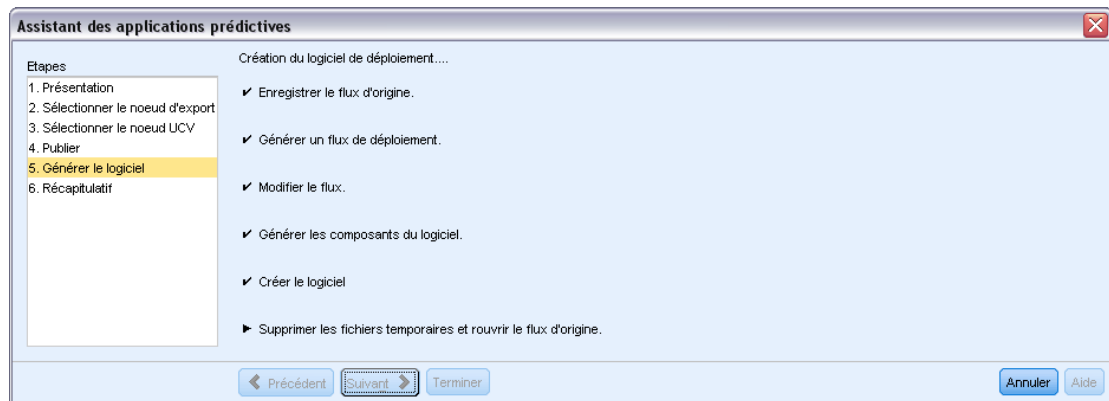


Cliquez sur Suivant pour vérifier automatiquement les métadonnées de flux et vos spécifications. Si toutes ces données sont fournies correctement, le fichier `.cdp` est généré. Relevez l'emplacement cible car vous devrez ultérieurement accéder à ce fichier lors de la publication dans l'application.

Etape 5 : Création du logiciel

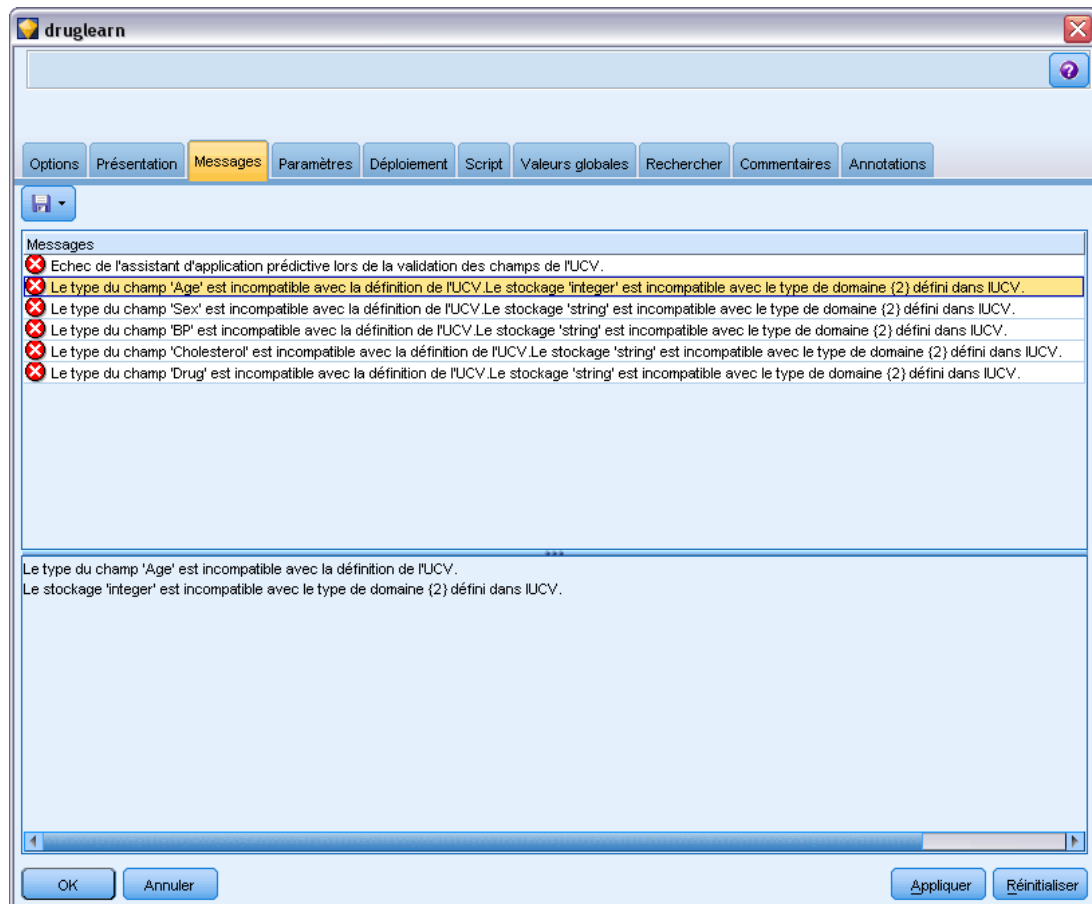
A ce stade, l'assistant vérifie automatiquement les informations de flux, ainsi que les spécifications déterminées dans les écrans précédents.

Figure 10-6
Création du logiciel de déploiement



Les noms et types de champ du flux IBM® SPSS® Modeler sont vérifiés en fonction du fichier XML généré à partir de l'UCV. Si des informations de champ ne coïncident pas, l'assistant peut se fermer automatiquement et afficher les messages d'erreur correspondants.

Figure 10-7
Messages d'erreur liés à des métadonnées incompatibles



En cas d'erreur, revenez au flux SPSS Modeler et vérifiez les éléments suivants :

- Vérifiez que tous les noms de champ du noeud UCV (noeud Typer du flux) figurent dans le fichier XML qui définit l'UCV. Assurez-vous également que leurs types (domaines) sont compatibles.
- Relevez la casse des noms de champ car SPSS Modeler et les autres applications peuvent distinguer les majuscules des minuscules. Toutefois, l'ordre des champs n'a aucune importance.

Remarque : La taille du logiciel *.cdp* créé doit être inférieure à 5 Ko. En particulier, si la liste des champs générés est longue, vous devez en supprimer une partie de la description pour respecter cette limite.

Etape 6 : Récapitulatif

Une fois le logiciel IBM® SPSS® Modeler Deployment Package créé, votre travail dans SPSS Modeler est terminé.

Etape suivante

Par la suite, vous pouvez importer le modèle (fichier du logiciel *.cdp*) dans Interaction Builder.

Importation et exportation de modèles au format PMML

PMML (Predictive Model Markup Language - Langage de balisage de modèle de prévision) est un format XML permettant de décrire les modèles de Data mining et de statistiques, notamment les entrées des modèles, les transformations utilisées avant la préparation des données pour le Data mining et les paramètres définissant les modèles eux-mêmes. IBM® SPSS® Modeler peut importer et exporter le format PMML, permettant ainsi de partager des modèles avec d'autres applications qui prennent en charge ce format, tel que IBM® SPSS® Statistics.

Remarque : Une licence pour IBM® SPSS® Modeler Server est nécessaire pour exporter au format PMML.

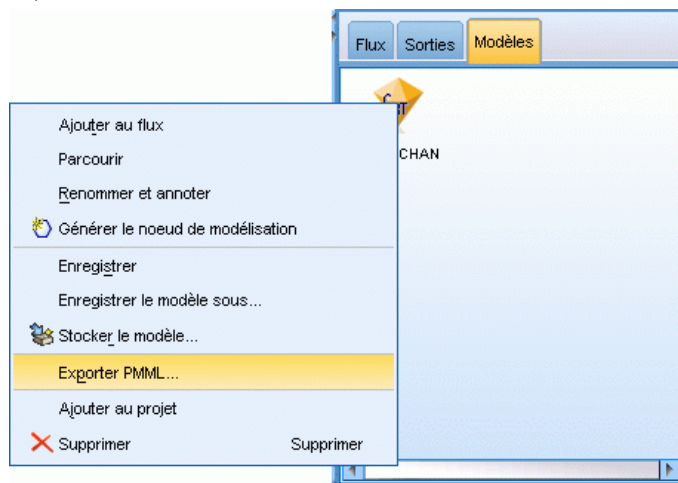
Pour plus d'informations sur le langage PMML, reportez-vous au site Web du Data mining group (<http://www.dmg.org>).

Pour exporter un modèle

L'exportation PMML est prise en charge pour la plupart des types de modèle générés dans SPSS Modeler. [Pour plus d'informations, reportez-vous à la section Types de modèle prenant en charge le format PMML sur p. 212.](#)

- ▶ Cliquez avec le bouton droit de la souris sur un nugget de modèle dans la palette des modèles. (Vous pouvez aussi double-cliquer sur un nugget de modèles dans l'espace de travail et sélectionner le menu Fichier.)
- ▶ Dans le menu, cliquez sur Exporter PMML.

Figure 10-8
Exportation d'un modèle au format PMML



- Dans la boîte de dialogue Exporter (ou Enregistrer), choisissez un répertoire cible et un nom unique pour le modèle.

Remarque : Vous pouvez modifier les options d'exportation PMML dans la boîte de dialogue Options utilisateur. Dans le menu principal, cliquez sur :
Outils > Options > Options utilisateur

et cliquez sur l'onglet PMML.

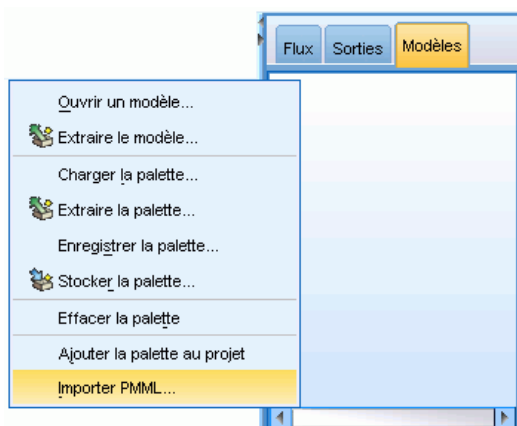
Pour plus d'informations, reportez-vous à la section Définition des options d'exportation PMML dans le chapitre 12 sur p. 240.

Pour importer un modèle enregistré au format PMML

Les modèles exportés au format PMML à partir de SPSS Modeler ou d'une autre application peuvent être importés dans la palette de modèles. Pour plus d'informations, reportez-vous à la section Types de modèle prenant en charge le format PMML sur p. 212.

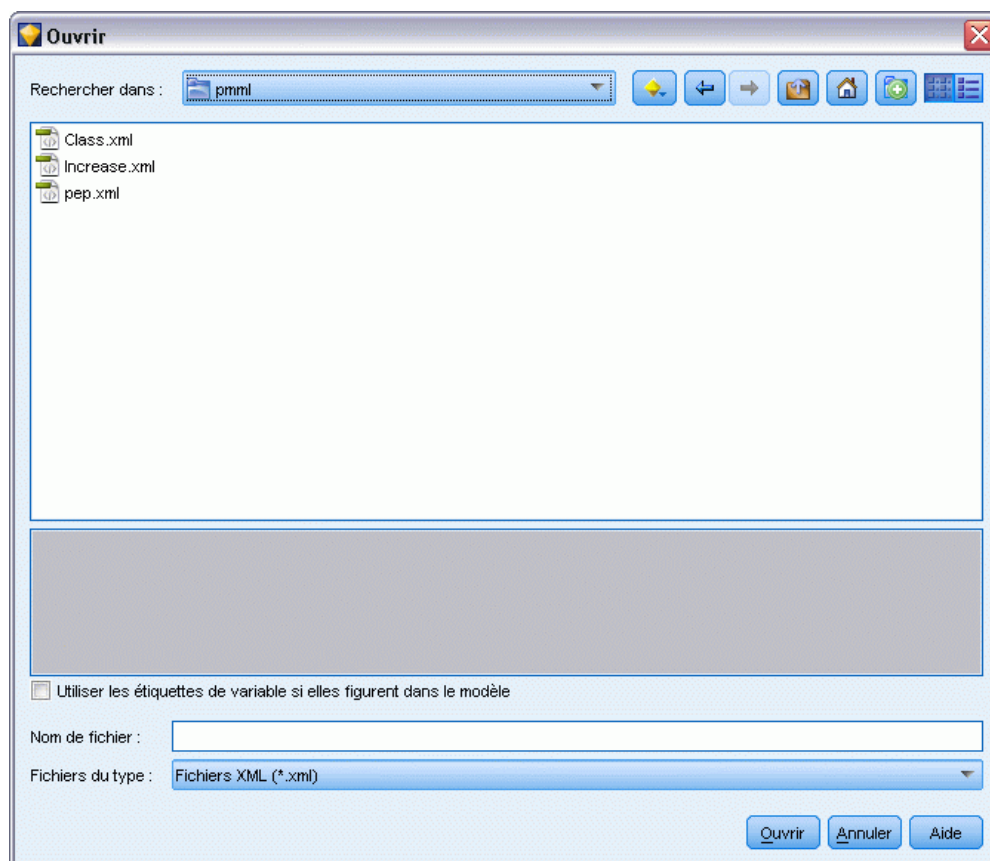
- Cliquez avec le bouton droit de la souris sur cette palette, puis sélectionnez Importer PMML dans le menu.

Figure 10-9
Importation d'un modèle au format PMML



- ▶ Sélectionnez le fichier à importer et indiquez les options requises pour les étiquettes de variables.
- ▶ Cliquez sur Ouvrir.

Figure 10-10
Sélection du fichier XML pour un modèle enregistré au format PMML



Utilisez des étiquettes de variables si elles sont présentes dans le modèle. Le langage PMML (Predictive Model Markup Language) permet de spécifier les noms et les étiquettes (par exemple, ID Référence pour *IDRéf*) des variables du dictionnaire de données. Sélectionnez cette option pour utiliser des étiquettes de variables si elles apparaissent dans le langage PMML initialement exporté.

Si vous avez sélectionné l'option d'étiquette de variable alors que le langage PMML ne contient pas d'étiquette de variable, le nom des variables est utilisé.

Types de modèle prenant en charge le format PMML

Exportation en PMML

SPSS Modeler modèles. Les modèles suivants créés dans IBM® SPSS® Modeler peuvent être exportés au format PMML 4.0 :

- Arbre C&RT
- QUEST
- CHAID
- Régression linéaire
- R. neurones
- C5.0
- Régression logistique
- Genlin
- SVM
- Bayes Net
- A priori
- Carma
- Séquence
- Nuées dynamiques
- Kohonen
- TwoStep
- KNN
- Modèle Statistics

Le modèle suivant créé dans SPSS Modeler peut être exporté au format PMML 3.2 :

- Liste de décisions

Modèles natifs de base de données. Pour les modèles générés à l'aide d'algorithmes natifs de base de données, l'exportation PMML est disponible pour les modèles IBM InfoSphere Warehouse uniquement. Les modèles créés avec Analysis Services de Microsoft ou Oracle Data Miner ne peuvent pas être exportés. Notez également que les modèles IBM exportés au format PMML ne peuvent pas être réimportés dans SPSS Modeler. [Pour plus d'informations, reportez-vous](#)

à la section Modélisation de base de données - Présentation dans le chapitre 2 dans *Guide d'exploration de base de données IBM SPSS Modeler 14.2*.

Importation de PMML

SPSS Modeler peut importer et évaluer les modèles PMML générés par les versions actuelles de tous les produits IBM® SPSS® Statistics, y compris les modèles exportés depuis SPSS Modeler, et les modèles ou transformations PMML générés par SPSS Statistics 17.0 ou une version ultérieure. En d'autres termes, il s'agit de tout PMML que le moteur d'évaluation peut évaluer, à l'exception des éléments suivants :

- Les modèles Apriori, Carma et Détection des anomalies ne peuvent pas être importés.
- Les modèles PMML ne seront peut-être pas parcourus après importation dans SPSS Modeler, même s'ils peuvent être utilisés pour le scoring. (Notez que cela inclut les modèles qui ont été exportés à l'origine de SPSS Modeler. Pour contourner cet obstacle, exportez le modèle au format de fichier de modèle généré [*.gm] plutôt qu'au format PMML.)
- Les modèles IBM InfoSphere Warehouse exportés au format PMML ne peuvent pas être importés.
- La validation limitée se produit lors de l'importation, mais la validation complète s'effectue lors de l'évaluation du modèle. Par conséquent, il est possible que l'importation réussisse mais que l'évaluation échoue ou produise des résultats incorrects.

Projets et rapports

Introduction aux projets

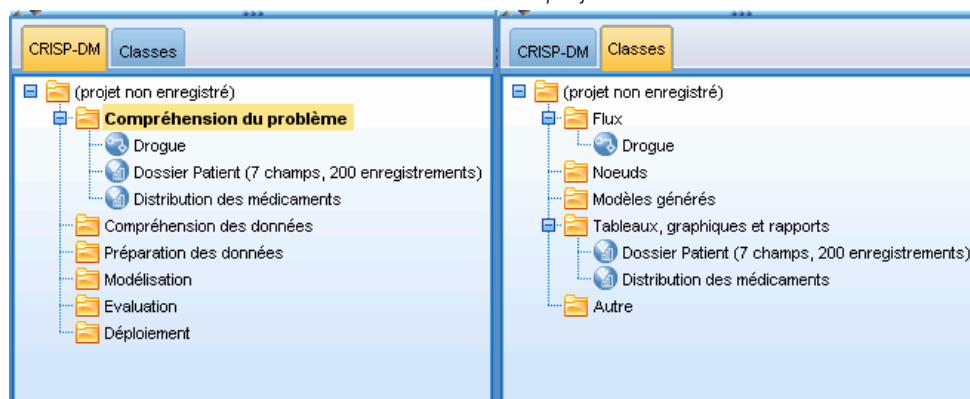
Un **projet** est un groupe de fichiers lié à un travail de Data mining. Les projets incluent des flux de données, des graphiques, des modèles générés, des rapports, ainsi que tout élément créé dans IBM® SPSS® Modeler. Au premier coup d'oeil, il peut sembler que les projets SPSS Modeler ne sont qu'un simple moyen d'organiser les sorties, mais ils sont en fait capables de bien plus. A l'aide des projets, vous pouvez notamment :

- Annoter chacun des objets du fichier de projet.
- Utiliser la méthodologie CRISP-DM dans vos tâches de Data mining. Les projets contiennent également le système d'aide CRISP-DM qui fournit des détails et des exemples réels relatifs au Data mining combiné à la méthodologie CRISP-DM.
- Ajouter au projet des objets créés hors de SPSS Modeler, tels qu'un diaporama PowerPoint utilisé pour présenter vos objectifs de Data mining ou vos livres blancs sur les algorithmes que vous envisagez d'utiliser.
- Générer à la fois des rapports de mise à jour simples et détaillés, sur la base de vos annotations. Il est possible de générer ces rapports au format HTML afin d'en faciliter la publication sur l'intranet de votre entreprise.

Remarque : Si le panneau de projet n'est pas visible dans la fenêtre SPSS Modeler, cliquez sur Project dans le menu Affichage.

Vous pouvez visualiser les objets ajoutés à un projet de deux façons : A l'aide du **mode Classes** et du **mode CRISP-DM**. Tous les éléments que vous ajoutez à un projet sont ajoutés aux deux modes et vous pouvez basculer entre ces modes pour créer l'environnement de travail le plus efficace.

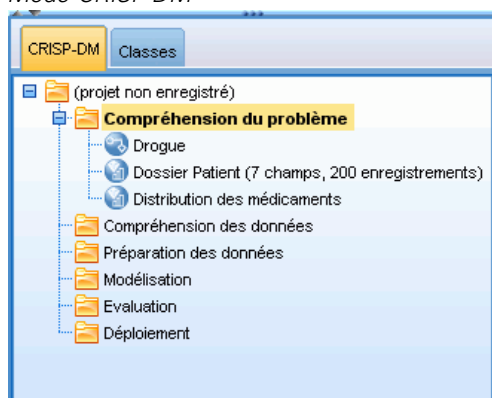
Figure 11-1
Mode CRISP-DM et mode Classes d'un fichier de projet



Mode CRISP-DM

Grâce à la prise en charge de CRISP-DM (Cross-Industry Standard Process for Data Mining), les projets IBM® SPSS® Modeler fournissent une méthodologie confirmée et non exclusive permettant d'organiser les éléments de vos travaux de Data mining. CRISP-DM utilise six phases pour décrire le processus du début (collecte des exigences commerciales) à la fin (déploiement de vos résultats). Même si certaines phases n'impliquent généralement pas d'actions dans SPSS Modeler, le panneau de projet intègre les six phases de sorte que vous disposiez d'un emplacement central dédié au stockage et au suivi de tous les éléments associés au projet. Par exemple, la phase de compréhension du problème implique généralement la collecte des exigences et une réunion entre collègues afin de déterminer les objectifs plutôt que le traitement des données dans SPSS Modeler. Le panneau de projet vous permet de stocker vos notes résultant de ce type de réunion dans le dossier *Compréhension du problème* pour vous y reporter ultérieurement et pour l'inclure dans les rapports.

Figure 11-2
Mode CRISP-DM



La vue CRISP-DM dans le panneau de projet comporte son propre système d'aide afin de vous guider à travers le cycle de vie du Data mining. Dans SPSS Modeler, vous pouvez accéder à ce système d'aide en cliquant sur Aide CRISP-DM dans le menu Aide.

Remarque : Si le panneau de projet n'est pas visible dans la fenêtre, cliquez sur Project dans le menu Affichage.

Définition de la phase par défaut du projet

Les objets ajoutés à un projet le sont dans une phase par défaut de CRISP-DM. En d'autres termes, il est nécessaire d'organiser les objets manuellement selon la phase de Data mining dans laquelle vous les utilisez. Il est judicieux de définir le dossier par défaut en fonction de la phase dans laquelle vous travaillez actuellement.

Pour sélectionner la phase à utiliser par défaut :

- ▶ En mode CRISP-DM, cliquez avec le bouton droit de la souris sur le dossier de la phase à définir comme phase par défaut.
- ▶ Dans le menu, cliquez sur Définir par défaut.

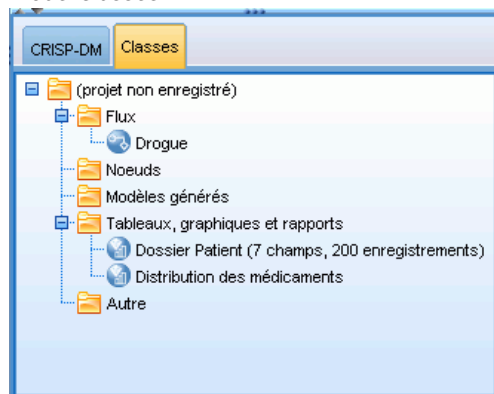
Le dossier par défaut est affiché en caractères gras.

Mode Classes

Le mode Classes du panneau de de projet organise votre travail dans IBM® SPSS® Modeler en catégories, selon les types d'objet créés. Vous pouvez ajouter des objets enregistrés à l'une des catégories suivantes :

- Flux
- Noeuds
- Modèles
- Tableaux, graphiques, rapports
- Autres (fichiers n'appartenant pas à SPSS Modeler, tels que des diaporamas ou des livres blancs applicables à votre travail de Data mining)

Figure 11-3
mode Classes



L'ajout d'objets au mode Classes entraîne également leur ajout dans le dossier de la phase par défaut du mode CRISP-DM.

Remarque : Si le panneau de projet n'est pas visible dans la fenêtre, cliquez sur Project dans le menu Affichage.

Création d'un projet

Un projet est avant tout un fichier qui contient des références à tous les fichiers que vous associez au projet. En d'autres termes, les éléments du projet sont enregistrés à la fois individuellement et en tant que référence dans le fichier de projet (.cpj). Etant donné cette structure référentielle, il est nécessaire de prendre en considération les points suivants :

- Les éléments de projet doivent d'abord être enregistrés individuellement avant d'être ajoutés à un projet. Si un élément n'est pas enregistré, vous serez invité à l'enregistrer avant de l'ajouter au projet en cours.

- Les objets mis à jour individuellement, tels que les flux, sont également mis à jour dans le fichier de projet.
- Lorsque vous déplacez ou supprimez manuellement des objets (tels que des flux, des noeuds ou des objets de sortie) du système de fichiers, les liens du fichier de projet ne sont plus valides.

Création d'un projet

Vous pouvez facilement créer des projets dans la fenêtre IBM® SPSS® Modeler. Vous pouvez créer un projet à partir d'un projet ouvert, ou fermer le projet existant et créer entièrement un projet.

- ▶ Dans le menu principal, cliquez sur :
Fichier > Projet > Nouveau projet...

Ajout à un projet

Après avoir créé ou ouvert un projet, vous pouvez ajouter des objets, tels que des flux de données, des noeuds et des rapports, à l'aide de plusieurs méthodes.

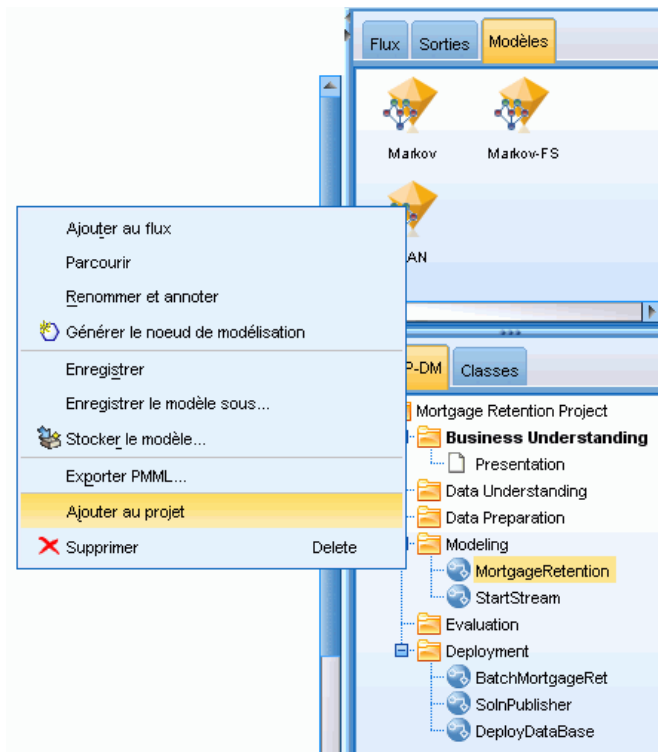
Ajout d'objets à partir des gestionnaires

A l'aide des gestionnaires situés dans la partie supérieure droite de la fenêtre IBM® SPSS® Modeler, vous pouvez ajouter des flux ou des sorties.

- ▶ Sélectionnez un objet, tel qu'un tableau ou un flux, dans l'un des onglets des gestionnaires.
- ▶ Cliquez avec le bouton droit de la souris sur l'élément et cliquez sur Ajouter au projet.
Si l'objet a été préalablement enregistré, il sera automatiquement ajouté au dossier d'objets approprié (en mode Classes) ou au dossier de la phase par défaut (en mode CRISP-DM).
- ▶ Vous pouvez également faire glisser les objets des gestionnaires vers le panneau des projets.

Remarque : vous pouvez être invité à enregistrer d'abord l'objet. Au moment de l'enregistrement, veillez à ce que l'option Ajouter le fichier au projet soit sélectionnée dans la boîte de dialogue Enregistrer. Cette sélection entraînera l'ajout automatique de l'objet au projet après l'enregistrement.

Figure 11-4
Ajout d'éléments à un projet



Ajout de noeuds à partir de l'espace de travail

Vous pouvez ajouter des noeuds individuels à partir de l'espace de travail de flux à l'aide de la boîte de dialogue Enregistrer.

- ▶ Sélectionnez un noeud dans l'espace de travail.
- ▶ Cliquez avec le bouton droit de la souris sur l'élément et cliquez sur Enregistrer le noeud. Sinon, dans le menu principal, cliquez sur :
Edition > Noeud > Enregistrer le noeud...
- ▶ Dans la boîte de dialogue Enregistrer, sélectionnez Ajouter le fichier au projet.
- ▶ Nommez le noeud et cliquez sur Enregistrer.

Cette action enregistre le fichier et l'ajoute au projet. Les noeuds sont ajoutés au dossier *Noeuds* en mode Classes et au dossier de la phase par défaut en mode CRISP-DM.

Ajout de fichiers externes

Vous pouvez ajouter à un projet de nombreux types d'objet n'appartenant pas à SPSS Modeler. Cette possibilité est utile lorsque vous gérez le processus global de Data mining dans SPSS Modeler. Vous pouvez, par exemple, stocker des liens vers des données, des notes, des présentations et des graphiques dans un projet. En mode CRISP-DM, vous pouvez ajouter des

fichiers externes au dossier de votre choix. En mode Classes, les fichiers externes peuvent uniquement être enregistrés dans le dossier *Autre*.

Pour ajouter des fichiers externes à un projet :

- ▶ Faites glisser les fichiers du bureau vers le projet.

ou

- ▶ Cliquez avec le bouton droit de la souris sur le dossier cible en mode CRISP-DM ou en mode Classes.
- ▶ Dans le menu, cliquez sur Ajouter au dossier.
- ▶ Sélectionnez un fichier dans la boîte de dialogue et cliquez sur Ouvrir.

Cette opération permet d'ajouter une référence à l'objet sélectionné au sein des projets SPSS Modeler.

Transfert de projets vers le IBM SPSS Collaboration and Deployment Services Repository

Vous pouvez transférer un projet entier, avec tous ses fichiers constitutifs, dans le IBM® SPSS® Collaboration and Deployment Services Repository en une seule opération. Les objets déjà présents dans l'emplacement cible ne sont pas déplacés. Cette fonction s'utilise également dans le sens inverse : vous pouvez transférer des projets entiers du IBM SPSS Collaboration and Deployment Services Repository vers votre système de fichiers local.

Remarque : Une licence distincte est requise pour accéder à un référentiel IBM® SPSS® Collaboration and Deployment Services. Pour plus d'informations, reportez-vous à <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Transfert d'un projet

Vérifiez que le projet à transférer est ouvert dans le panneau de projet.

Pour transférer un projet :

- ▶ Cliquez avec le bouton droit de la souris sur le dossier racine du projet et cliquez sur l'option Transférer le projet.
- ▶ Si un message vous le demande, connectez-vous au IBM SPSS Collaboration and Deployment Services Repository.
- ▶ Spécifiez le nouvel emplacement du projet et cliquez sur OK.

Définition des propriétés d'un projet

Vous pouvez personnaliser le contenu et la documentation d'un projet en utilisant la boîte de dialogue des propriétés du projet. Pour accéder aux propriétés du projet :

- ▶ Cliquez avec le bouton droit de la souris sur un objet ou un dossier dans le panneau de projet et choisissez Propriétés du projet.
- ▶ Cliquez sur l'onglet Projet pour spécifier les informations élémentaires sur le projet.

Figure 11-5
Définition des propriétés d'un projet

Objets du projet	Comptages
Flux	0
Noeuds	0
Super noeuds	0
Modèles générés	1
Tables, graphiques et rapports	3
Autre	0
TOTAL	4

Créé(e). Indique la date de création du projet (non modifiable).

Récapitulatif. Vous pouvez entrer un récapitulatif de votre projet de Data mining. Il sera affiché dans le rapport du projet.

Contenu. Répertorie le type des composants référencés par le fichier de projet, ainsi que leur nombre (non modifiable).

Enregistrer les objets non enregistrés en tant que. Indique si les objets non enregistrés doivent être sauvegardés dans le système de fichiers local ou stockés dans le référentiel. [Pour plus d'informations, reportez-vous à la section A propos du IBM SPSS Collaboration and Deployment Services Repository dans le chapitre 9 sur p. 159.](#)

Mettre à jour les références d'objet lors du chargement du projet. Sélectionnez cette option pour mettre à jour les références du projet sur ses composants. *Remarque* : les fichiers ajoutés à un projet ne sont pas enregistrés dans le fichier de projet lui-même. Une référence aux fichiers est en fait stockée dans le projet. En d'autres termes, le fait de déplacer ou de supprimer un fichier entraîne la suppression de cet objet du projet.

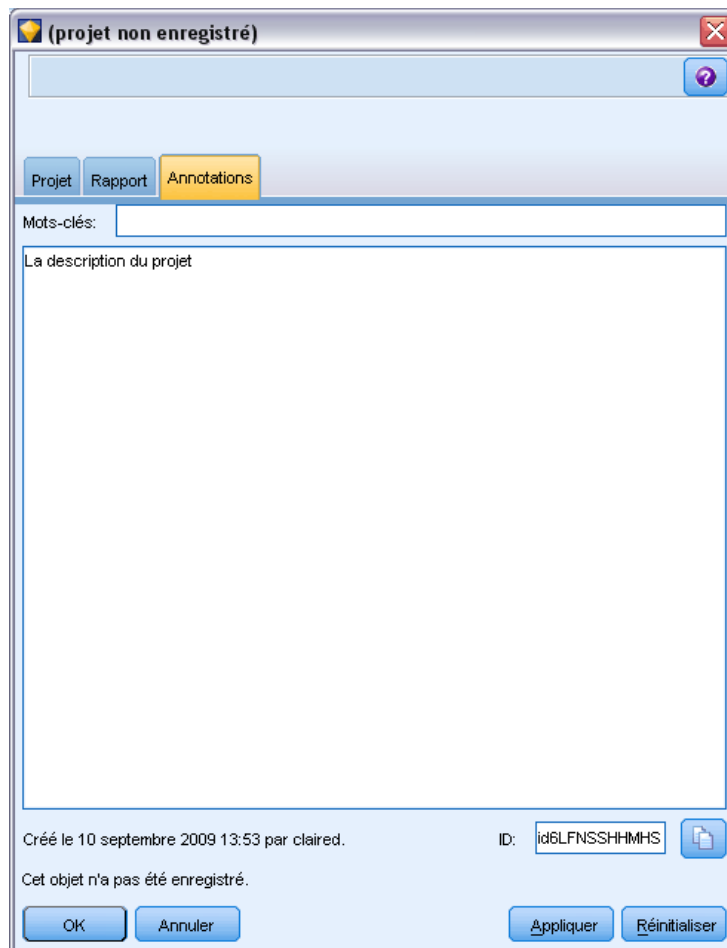
Annotation d'un projet

Le panneau de projet met à votre disposition différentes méthodes d'annotation de vos travaux de Data mining. Les annotations relatives au projet sont souvent utilisées pour assurer le suivi des objectifs et décisions de façon globale, tandis que les annotations relatives au dossier ou au noeud fournissent des détails supplémentaires. L'onglet Annotations offre assez d'espace pour consigner des détails relatifs au projet, tels que l'exclusion de données contenant des éléments manquants irrécupérables, ou d'hypothèses formulées durant l'exploration des données et pouvant s'avérer utiles.

Pour annoter un projet :

- ▶ Sélectionnez le dossier de projet en mode CRISP-DM ou en mode Classes.
- ▶ Cliquez avec le bouton droit de la souris sur le dossier et sélectionnez Propriétés du projet.
- ▶ Cliquez sur l'onglet Annotations.

Figure 11-6
Onglet Annotations dans la boîte de dialogue des propriétés du projet



- Entrez des mots-clés et un texte décrivant le projet.

Annotations et propriétés du dossier

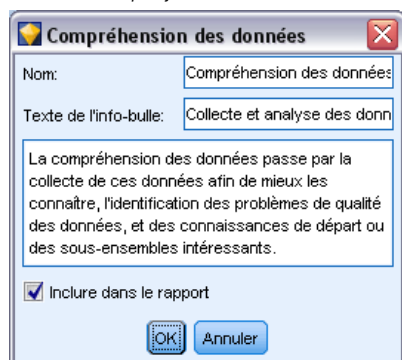
Vous pouvez annoter des dossiers de projet individuels (à la fois en mode CRISP-DM et en mode Classes). En mode CRISP-DM, ces annotations peuvent s'avérer extrêmement efficaces pour documenter les objectifs de votre entreprise pour chaque phase du Data mining. Par exemple, en utilisant l'outil Annotation pour le dossier *Compréhension du problème*, vous pouvez inclure une formulation telle que « L'objectif commercial de cette étude est de réduire le score d'attrition parmi les clients les plus importants ». Ce texte pourra alors être automatiquement inclus dans le rapport du projet si vous sélectionnez Inclure dans le rapport.

Pour annoter un dossier :

- Sélectionnez un dossier dans le panneau de projet.
- Cliquez avec le bouton droit de la souris sur le dossier et cliquez sur Propriétés du dossier.

En mode CRISP-DM, les dossiers sont annotés avec un résumé de l'objectif de chaque phase, ainsi que des lignes directrices sur l'exécution des travaux de Data mining. Vous pouvez éditer ou supprimer ces annotations.

Figure 11-7
Dossier de projet avec annotation CRISP-DM



Nom. Cette zone affiche le nom du champ sélectionné.

Texte de l'info-bulle. Permet de créer des info-bulles personnalisées qui apparaissent lorsque vous positionnez le pointeur de la souris sur un dossier de projet. Ces info-bulles sont utiles en mode CRISP-DM, par exemple, pour fournir un aperçu des objectifs de chaque phase ou pour marquer l'état d'une phase, tel que « En cours » ou « Terminé ».

Champ Annotation. Utilisez ce champ pour les annotations plus longues pouvant être rassemblées dans le rapport du projet. Le mode CRISP-DM inclut une description de chaque phase de Data mining dans l'annotation, mais vous pouvez évidemment personnaliser cet élément selon votre projet.

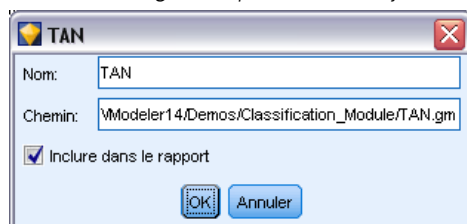
Inclure dans le rapport. Pour inclure l'annotation dans les rapports, sélectionnez Inclure dans le rapport.

Propriétés de l'objet

Vous pouvez visualiser les propriétés des objets et choisir d'inclure des objets individuels dans le rapport du projet. Pour accéder aux propriétés des objets :

- ▶ Cliquez avec le bouton droit de la souris sur un objet dans le panneau du projet.
- ▶ Dans le menu, cliquez sur Propriétés de l'objet.

Figure 11-8
Boîte de dialogue Propriétés de l'objet



Nom. Cette zone indique le nom de l'objet enregistré.

Chemin. Cette zone indique l'emplacement de l'objet enregistré.

Inclure dans le rapport. Cochez cette case si vous souhaitez que les détails de l'objet soient intégrés dans un rapport.

Fermeture d'un projet

Lorsque vous quittez IBM® SPSS® Modeler ou que vous ouvrez un nouveau projet, le fichier du projet existant (fichier *.cpj*) est fermé.

Certains fichiers associés à ce projet (tels que les flux, les noeuds ou les graphiques) peuvent rester ouverts. Si vous souhaitez garder ces fichiers ouverts, choisissez Non à l'affichage du message ... Souhaitez-vous enregistrer et fermer ces fichiers ?

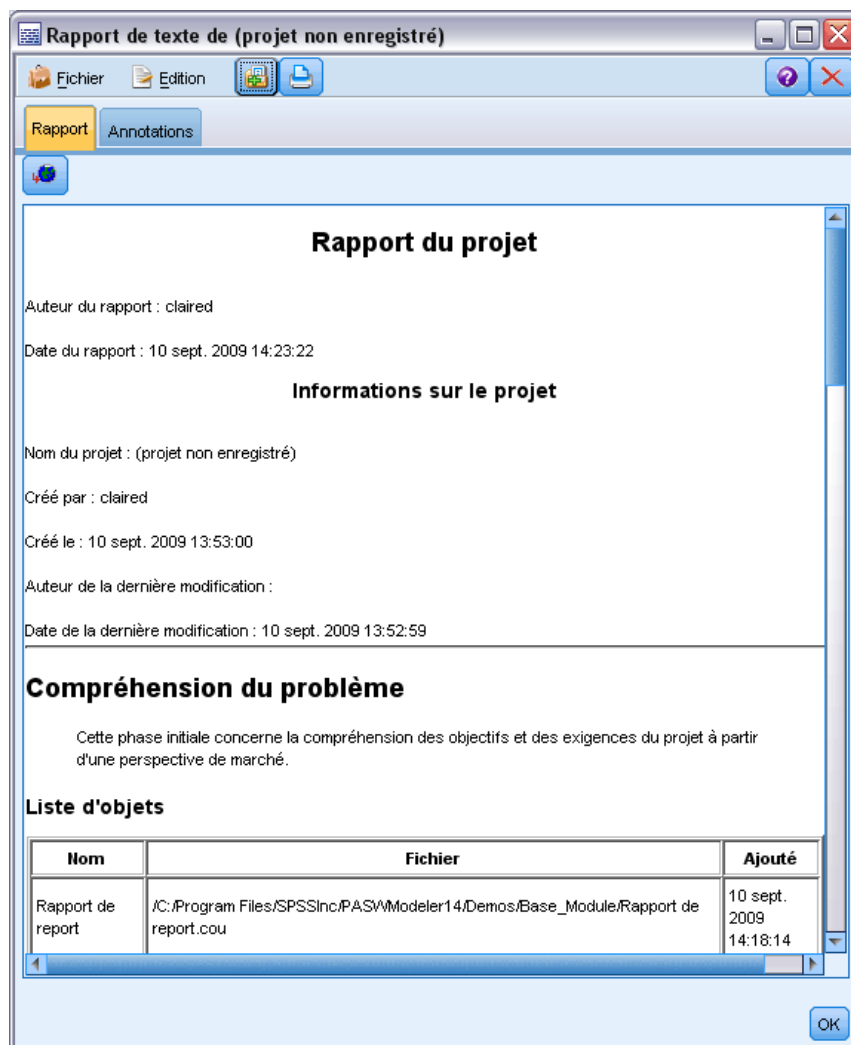
Si vous modifiez et enregistrez des fichiers associés après la fermeture d'un projet, ces nouvelles versions seront incluses dans le projet lors de sa prochaine ouverture. Pour empêcher cette action, retirez le fichier du projet ou enregistrez-le sous un nom différent.

Génération d'un rapport

L'une des fonctions les plus utiles des projets est la capacité à générer des rapports sur la base des éléments et des annotations du projet. Ce composant est essentiel pour un travail de Data mining efficace, comme l'explique la section relative à la méthodologie CRISP-DM. Vous pouvez générer un rapport directement dans l'un des divers types de fichier disponibles ou dans une fenêtre de sortie à l'écran, pour consultation immédiate. Vous pouvez alors imprimer, enregistrer ou afficher le rapport dans un navigateur Web. Vous avez également la possibilité de remettre les rapports enregistrés à d'autres personnes de votre entreprise.

Un même projet donne souvent lieu à la création de plusieurs rapports au cours du processus de Data mining ; ils sont distribués à toutes les personnes impliquées dans le projet. Le rapport sélectionne des informations sur les objets référencés à partir du fichier de projet et à partir de toutes les annotations créées. Vous pouvez créer des rapports en mode Classes ou en mode CRISP-DM.

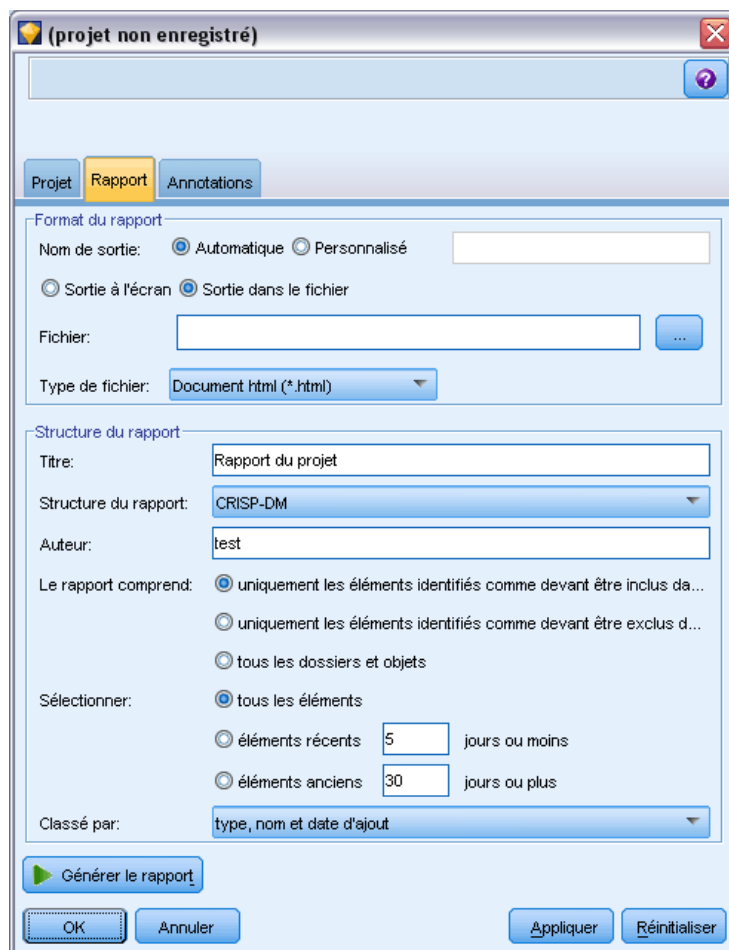
Figure 11-9
Fenêtre du rapport généré



Pour générer un rapport :

- ▶ Sélectionnez le dossier de projet en mode CRISP-DM ou en mode Classes.
- ▶ Cliquez avec le bouton droit de la souris sur le dossier et cliquez sur Rapport du projet.
- ▶ Spécifiez les options du rapport et cliquez sur Générer le rapport.

Figure 11-10
Sélection des options pour un rapport



Les options de la boîte de dialogue du rapport offrent différentes façons de générer le type de rapport souhaité :

Nom de sortie. Spécifiez le nom de la fenêtre de sortie si vous souhaitez afficher les données du rapport à l'écran. Vous pouvez indiquer un nom personnalisé ou autoriser IBM® SPSS® Modeler à attribuer automatiquement un nom à la fenêtre.

Sortie à l'écran. Sélectionnez cette option pour générer et afficher le rapport dans une fenêtre de sortie. Notez que vous avez la possibilité d'exporter le rapport vers différents types de fichier à partir de la fenêtre de sortie.

Sortie dans le fichier. Sélectionnez cette option pour générer et enregistrer le rapport au format indiqué dans la liste Type de fichier.

Fichier. Indiquez le nom de fichier du rapport généré. Les fichiers sont enregistrés par défaut dans le répertoire de SPSS Modeler `\bin`. Utilisez le bouton ... pour indiquer un autre emplacement.

Type de fichier. Les types de fichier disponibles sont les suivants :

- **Document HTML.** Le rapport est enregistré sous la forme d'un fichier HTML unique. S'il contient des graphiques, ils sont enregistrés comme des fichiers PNG référencés par le fichier HTML. Lors de la publication du rapport sur Internet, veillez à envoyer le fichier HTML et les images référencées.
- **Document texte.** Le rapport est enregistré sous la forme d'un fichier texte unique. S'il contient des graphiques, seuls le nom de fichier et le chemin d'accès sont inclus dans le rapport.
- **Document MS Word.** Le rapport est enregistré sous la forme d'un document unique, les graphiques étant intégrés directement dans le document.
- **Document MS Excel.** Le rapport est enregistré sous la forme d'une feuille de calcul unique, les graphiques étant intégrés directement dans la feuille de calcul.
- **Document MS PowerPoint.** Chaque étape apparaît sur une nouvelle diapositive. Les graphiques sont intégrés directement dans les diapositives PowerPoint.
- **Objet de sortie.** Quand il est ouvert dans SPSS Modeler, ce fichier (.cou) est le même que celui obtenu avec l'option Sortie à l'écran dans le groupe Format de rapport.

Remarque : Pour effectuer une exportation vers un fichier Microsoft Office, l'application correspondante doit être installée.

Titre : Spécifiez le titre du rapport.

Structure du rapport. Sélectionnez CRISP-DM ou Classes. Le mode CRISP-DM fournit un rapport de situation contenant des résumés de la situation globale, ainsi que des détails sur chaque phase du Data mining. Le mode Classes est un mode orienté objet qui est davantage approprié au suivi interne des données et des flux.

Auteur. Le nom d'utilisateur par défaut est affiché, mais vous pouvez le modifier.

Le rapport comprend. Sélectionnez la méthode d'introduction des objets dans le rapport. Sélectionnez tous les dossiers et objets pour inclure tous les éléments ajoutés au fichier de projet. Vous pouvez également inclure des éléments en fonction de la sélection ou non de l'option Inclure dans le rapport dans les propriétés de l'objet. Pour vérifier les éléments non consignés, vous pouvez également choisir d'inclure uniquement les éléments identifiés comme devant être exclus (l'option Inclure dans le rapport n'est pas sélectionnée).

Sélectionner. Cette option vous permet de fournir des mises à jour du projet en sélectionnant uniquement les éléments récents du rapport. Vous pouvez également assurer un suivi des questions plus anciennes et non résolues en définissant des paramètres pour les éléments anciens. Sélectionnez tous les éléments pour exclure le paramètre temps du rapport.

Ordre d'affichage : Vous pouvez sélectionner une combinaison des caractéristiques d'objet suivantes afin de les classer au sein d'un dossier :

- **Type.** Permet de grouper les objets par type.
- **Nom.** Permet de classer les objets par ordre alphabétique.
- **Date d'ajout.** Permet de trier les objets en fonction de la date d'ajout au projet.

Enregistrement et exportation de rapports générés

Les rapports générés à l'écran sont affichés dans une nouvelle fenêtre de sortie. Les graphiques qu'ils contiennent apparaissent sous la forme d'images en ligne.

Le nombre total de noeuds dans chaque flux est présenté dans une liste du rapport. Ces nombres se trouvent sous chacun des en-têtes suivants, qui utilisent la terminologie IBM® SPSS® Modeler et non la terminologie CRISP-DM.

- **Lecteurs de données.** Noeuds source. [Pour plus d'informations, reportez-vous à la section Aperçu dans le chapitre 2 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)
- **Rédacteurs de données.** noeuds d'exportation. [Pour plus d'informations, reportez-vous à la section Présentation des noeuds d'exportation dans le chapitre 7 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)
- **Concepteurs du modèle.** Concevoir ou modéliser des noeuds. [Pour plus d'informations, reportez-vous à la section Description des noeuds de modélisation dans le chapitre 3 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)
- **Utilisateurs du modèle.** Modèles générés, également appelés nuggets. [Pour plus d'informations, reportez-vous à la section Nuggets de modèle dans le chapitre 3 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)
- **Concepteurs de sortie.** Noeuds de graphique ou de sortie. [Pour plus d'informations, reportez-vous à la section Fonctions communes des noeuds Graphiques dans le chapitre 5 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#) [Pour plus d'informations, reportez-vous à la section Présentation des noeuds de sortie dans le chapitre 6 dans *IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie*.](#)
- **Autre.** Tous les autres noeuds associés au projet. Par exemple, ceux disponibles dans l'onglet Ops sur champs ou l'onglet Ops sur lignes de la palette des noeuds.

Pour enregistrer un rapport :

- ▶ Dans le menu Fichier, cliquez sur Enregistrer.
- ▶ Spécifiez un nom de fichier.
Le rapport est enregistré sous la forme d'un objet de sortie.

Pour exporter un rapport :

- ▶ Dans le menu Fichier, cliquez sur Exporter et indiquez le type de fichier dans lequel effectuer l'exportation.
- ▶ Spécifiez un nom de fichier.
Le rapport est enregistré dans le format choisi.

Vous pouvez effectuer l'exportation vers les types de fichier suivants :

- HTML
- Texte
- Microsoft Word

- Microsoft Excel
- Microsoft PowerPoint

Remarque : Pour effectuer une exportation vers un fichier Microsoft Office, l'application correspondante doit être installée.

Utilisez les boutons situés en haut de la fenêtre pour :

- Imprimer le rapport.
- Afficher le rapport au format HTML dans un navigateur Web externe.

Figure 11-11

Rapport affiché dans un navigateur Web

The screenshot shows a web browser window with a report. The report has a title 'Business Understanding' and a sub-header 'Liste d'objets'. It contains a table with one row of data. Below the table, there are sections for 'Presentation', 'Data Understanding', and 'Data Preparation', each with a brief description.

Nom	Fichier	Ajouté
Presentation	/C:/SPSS Screenshooting/Modeler_Shots_Kit/Data_Files/Mortgage Retention/Presentation.ppt	10 août 2004 18:31:32

Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective.

Liste d'objets

Presentation

Data Understanding

The data understanding involves collecting data in order to become familiar with the data, identify data quality problems and identify initial insights or interesting subsets.

Data Preparation

This phase covers all activities involved in constructing the final datasets to be used in the modeling phase.

Personnaliser IBM SPSS Modeler

Options de personnalisation de IBM SPSS Modeler

Vous pouvez effectuer un certain nombre d'opérations pour personnaliser IBM® SPSS® Modeler en fonction de vos besoins. Cette personnalisation consiste avant tout à définir des options utilisateur spécifiques telles que l'allocation de mémoire, les répertoires par défaut et l'utilisation du son et de la couleur. Vous pouvez aussi personnaliser la palette Noeuds située au bas de la fenêtre SPSS Modeler.

Définition des options de IBM SPSS Modeler

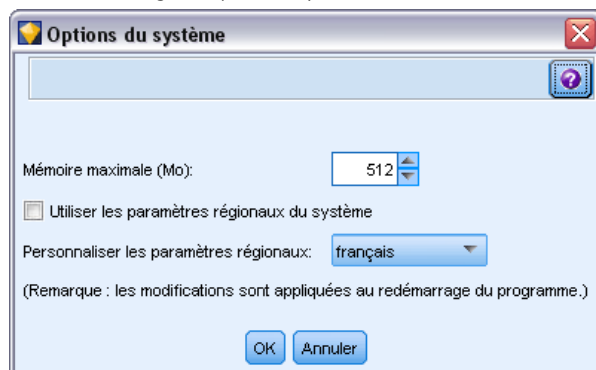
Vous pouvez personnaliser et définir les options de IBM® SPSS® Modeler de plusieurs façons :

- Définissez les options système, telles que l'utilisation de la mémoire et les paramètres régionaux, en cliquant sur Options système dans le menu Outils > Options.
- Définissez les options utilisateur, telles que les polices, les optimisations et les avertissements, en cliquant sur Options utilisateur dans le menu Outils > Options.
- Indiquez l'emplacement des applications utilisées avec SPSS Modeler en cliquant sur Programmes externes dans le menu Outils > Options.
- Indiquez les répertoires par défaut utilisés dans SPSS Modeler en cliquant sur Définir le répertoire ou Définir le répertoire du serveur dans le menu Fichier.

Options du système

Vous pouvez indiquer la langue ou les paramètres régionaux à utiliser avec IBM® SPSS® Modeler en cliquant sur Options système dans le menu Outils > Options. Vous pouvez également définir la quantité maximale de mémoire de SPSS Modeler. Notez que les changements effectués dans cette boîte de dialogue ne prennent effet qu'après avoir redémarré SPSS Modeler.

Figure 12-1
Boîte de dialogue Options système



Mémoire maximale. Cette option permet d'imposer une limite en mégaoctets à la mémoire utilisée par SPSS Modeler. Sur certaines plates-formes, SPSS Modeler limite la taille de son traitement pour réduire l'impact sur les ordinateurs surchargés ou disposant de ressources limitées. Si vous travaillez avec de grandes quantités de données, une erreur de « mémoire insuffisante » risque d'être générée. Vous pouvez alléger la charge de la mémoire en sélectionnant un nouveau seuil.

Utiliser les paramètres régionaux du système. Cette option est sélectionnée par défaut et paramétrée sur Anglais (Etats-Unis). Désélectionnez-la pour choisir une autre langue dans la liste des langues et des paramètres régionaux disponibles.

Gestion de la mémoire

Outre le paramètre Mémoire maximale indiqué dans la boîte de dialogue Options système, vous pouvez optimiser l'utilisation de la mémoire de plusieurs manières :

- Vous pouvez créer un cache sur tout noeud non terminal afin que les données soient lues à partir du cache, plutôt qu'extraites de la source de données lorsque vous exécutez le flux de données. Vous allégez ainsi la charge de la mémoire pour les ensembles de données volumineux. [Pour plus d'informations, reportez-vous à la section Options de mise en cache des noeuds dans le chapitre 5 sur p. 47.](#)
- Vous pouvez ajuster l'option Nombre maximal des membres pour les champs nominaux de la boîte de dialogue des propriétés du flux. Cette option permet de spécifier le nombre maximal de membres pour les champs nominaux. Au-delà de ce nombre, le niveau de mesure du champ devient *Sans type*. [Pour plus d'informations, reportez-vous à la section Définition des options de flux dans le chapitre 5 sur p. 51.](#)
- Vous pouvez forcer IBM® SPSS® Modeler à libérer de la mémoire en cliquant dans l'angle inférieur droit de la fenêtre, où figurent la mémoire utilisée par SPSS Modeler et la quantité allouée (xxMo / xxMo). Lorsque vous cliquez sur cette zone, elle prend une couleur plus foncée et la quantité de mémoire allouée diminue. La région reprend sa couleur normale lorsque SPSS Modeler a libéré toute la mémoire possible.

Définition des répertoires par défaut

Vous pouvez indiquer le répertoire par défaut des navigateurs de fichiers et de la sortie en sélectionnant Définir le répertoire ou Définir le répertoire du serveur dans le menu Fichier.

- **Définir le répertoire.** Vous pouvez utiliser cette option pour définir le répertoire de travail. Le répertoire de travail par défaut est basé sur le chemin d'installation de votre version de IBM® SPSS® Modeler ou sur le chemin de ligne de commande utilisé pour lancer SPSS Modeler. En mode local, le répertoire de travail est le chemin d'accès utilisé pour toutes les opérations effectuées côté client et pour les fichiers de sortie (s'ils sont référencés avec des chemins d'accès relatifs).
- **Définir le répertoire du serveur.** L'option Définir le répertoire du serveur du menu Fichier est activée lorsqu'il existe une connexion à un serveur distant. Utilisez cette option pour définir le répertoire par défaut de tous les fichiers serveur et fichiers de données indiqués pour l'entrée ou la sortie. Le répertoire du serveur par défaut est *\$CLEO/data*, où *\$CLEO* est le répertoire dans lequel la version serveur de SPSS Modeler est installée. Vous pouvez également utiliser

la ligne de commande pour remplacer ce répertoire par défaut : utilisez le commutateur `-server_directory` avec l'argument de ligne de commande `modelerclient`.

Définition des options utilisateur

Vous pouvez définir les options générales de IBM® SPSS® Modeler en sélectionnant Options utilisateur dans le menu Outils > Options. Ces options s'appliquent à tous les flux utilisés dans SPSS Modeler.

Les types d'option suivants peuvent être définis en cliquant sur l'onglet correspondant :

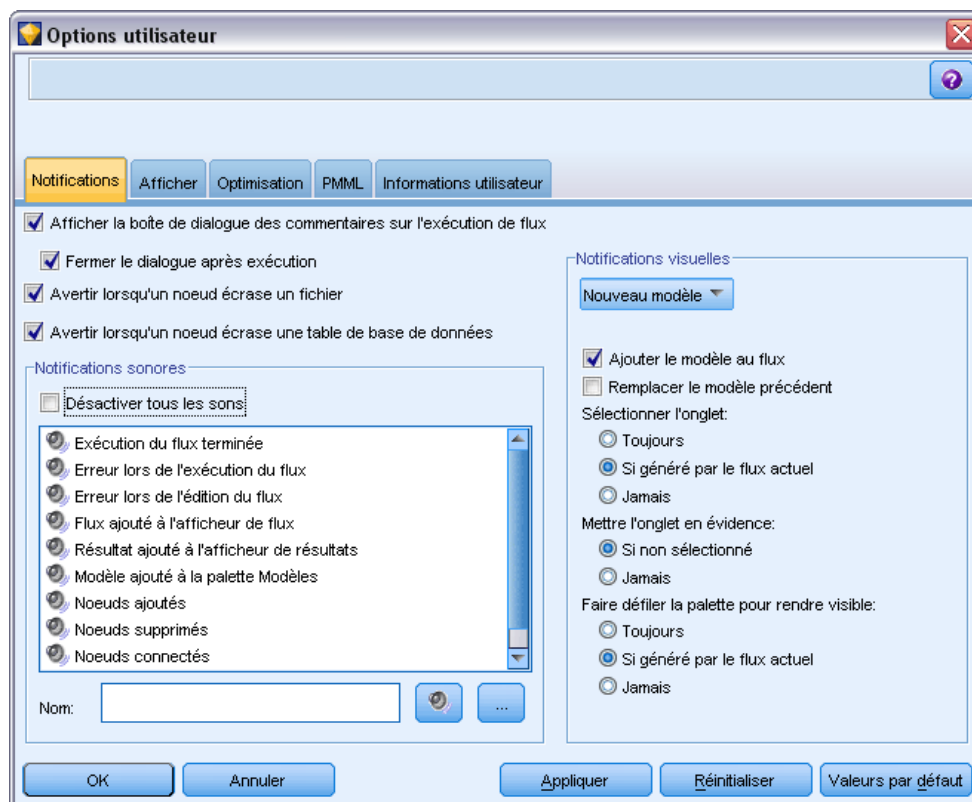
- Options de notification, telles que le remplacement d'un modèle et les messages d'erreur.
- Options d'affichage, telles que les couleurs des graphiques et des arrière-plans.
- Options d'optimisation, telles que la génération SQL et la réécriture de flux.
- Options d'exportation PMML utilisées lors de l'exportation de modèles au format PMML (Predictive Model Markup Language).
- Informations relatives à l'utilisateur et à l'auteur (vos nom, initiales et adresse électronique, par exemple). Ces informations peuvent être affichées dans l'onglet Annotations pour les noeuds et les autres objets créés.

Pour définir des options propres à un flux, telles que des séparateurs décimaux, des formats de date et d'heure et des scripts de flux, utilisez la boîte de dialogue des propriétés du flux, accessible à partir des menus Fichier et Outils.

Définition d'options de notification

L'onglet Notifications de la boîte de dialogue Options utilisateur vous permet de définir différentes options concernant le type et l'occurrence des avertissements et des fenêtres de confirmation dans IBM® SPSS® Modeler. Vous pouvez également définir le comportement des onglets Sorties et Modèles dans le panneau des gestionnaires lorsque vous générez de nouveaux modèles et sorties.

Figure 12-2
Boîte de dialogue Options utilisateur, onglet Notifications.



Afficher la boîte de dialogue des commentaires sur l'exécution de flux. Sélectionnez cette option pour afficher une boîte de dialogue comprenant un indicateur de progression lorsque l'exécution du flux est en cours depuis trois secondes. La boîte de dialogue contient également des détails sur les objets de sortie créés par le flux.

- **Fermez la boîte de dialogue une fois terminée.** Par défaut, la boîte de dialogue se ferme lorsque l'exécution du flux est terminée. Désélectionnez cette case si vous voulez que la boîte de dialogue reste visible lorsque l'exécution du flux se termine.

Avertir lorsqu'un noeud écrase un fichier. Sélectionnez cette option pour être averti par un message d'erreur lorsque des opérations de noeud écrasent un fichier existant.

Avertir lorsqu'un noeud écrase une table de base de données. Sélectionnez cette option pour être averti par un message d'erreur lorsque des opérations de noeud écrasent une table de base de données existante.

Notifications sonores

Utilisez la liste pour indiquer si vous souhaitez être averti par un son lorsqu'un événement ou une erreur se produit. Il existe un grand nombre de sons disponibles. Utilisez le bouton Lecture (haut-parleur) pour émettre le son sélectionné. Utilisez le bouton (...) pour sélectionner un son.

Remarque : les fichiers *.wav* utilisés pour créer des sons dans SPSS Modeler sont stockés dans le répertoire */media/sounds* du dossier d'installation.

- **Désactiver tous les sons.** Sélectionnez cette option pour désactiver la notification sonore pour tous les événements.

Notifications visuelles

Les options de ce groupe sont utilisées pour indiquer le comportement des onglets Sorties et Modèles du panneau des gestionnaires en haut à droite de l'affichage lorsque vous générez de nouveaux éléments. Sélectionnez *Nouveau modèle* ou *Nouvelle sortie* dans la liste pour indiquer le comportement de l'onglet correspondant.

Les options suivantes sont disponibles pour le format *Nouveau modèle* :

Ajout de modèle à un flux. Si cette option est sélectionnée (par défaut), un nouveau modèle est ajouté au flux, ainsi qu'♦♦ l'onglet Modèles, dès que le modèle est construit. Dans le flux, le modèle est affiché avec un lien pointant sur le noeud de modélisation à partir duquel le modèle a été créé. Si vous décochez cette case, le modèle est ajouté uniquement à l'onglet Modèles.

Remplacer le modèle précédent. Si cette option est sélectionnée (elle l'est par défaut), cela écrase le modèle existant à partir de ce flux dans l'onglet Modèles et sur l'espace de travail du flux. Si cette case est décochée, le modèle est ajouté aux modèles existants sur l'onglet et sur l'espace de travail. *Remarque* : ce paramètre est remplacé par le paramètre de remplacement du modèle sur un lien de modèle. [Pour plus d'informations, reportez-vous à la section Remplacement d'un modèle dans le chapitre 3 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Les options suivantes sont disponibles pour le format *Nouvelle sortie* :

Avertir lorsque les sorties excèdent [n]. Précisez s'il faut afficher un avertissement lorsque le nombre d'éléments de l'onglet Sortie dépasse une quantité prédéfinie. La quantité par défaut est 20 ; toutefois, vous pouvez la modifier au besoin.

Les options suivantes sont disponibles dans tous les cas :

Sélectionner l'onglet. Indiquez si vous souhaitez accéder à l'onglet Sorties ou Modèles lorsque l'objet correspondant est généré pendant l'exécution du flux.

- Sélectionnez *Toujours* pour accéder à l'onglet correspondant dans le panneau des gestionnaires.
- Sélectionnez *Si généré par le flux actuel* pour accéder à l'onglet correspondant uniquement pour les objets générés par le flux actuellement visible dans l'espace de travail.
- Sélectionnez *Jamais* pour empêcher que le logiciel accède à l'onglet correspondant afin de vous avertir des sorties ou des modèles générés.

Mettre l'onglet en évidence. Indiquez si vous souhaitez mettre en évidence l'onglet Sorties ou Modèles dans le panneau des gestionnaires lorsque de nouvelles sorties ou de nouveaux modèles sont générés.

- Sélectionnez *Si non sélectionné* pour mettre en évidence l'onglet correspondant (s'il n'est pas déjà sélectionné) lorsque de nouveaux objets sont générés dans le panneau des gestionnaires.
- Sélectionnez *Jamais* pour empêcher que le logiciel mette en évidence l'onglet correspondant afin de vous avertir d'objets générés.

Faire défiler la palette pour rendre visible (Nouveau modèle uniquement). Indiquez si vous souhaitez faire défiler automatiquement l'onglet Modèles du panneau des gestionnaires pour afficher le modèle le plus récent.

- Sélectionnez *Toujours* pour activer le défilement.
- Sélectionnez *Si généré par le flux actuel* pour activer le défilement uniquement pour les objets générés par le flux actuellement visible dans l'espace de travail.
- Sélectionnez *Jamais* pour empêcher que le logiciel n'active le défilement automatique de l'onglet Modèles.

Ouvrir une fenêtre (Nouvelle sortie uniquement). Sélectionnez s'il faut ouvrir automatiquement une fenêtre de sortie lors de la génération.

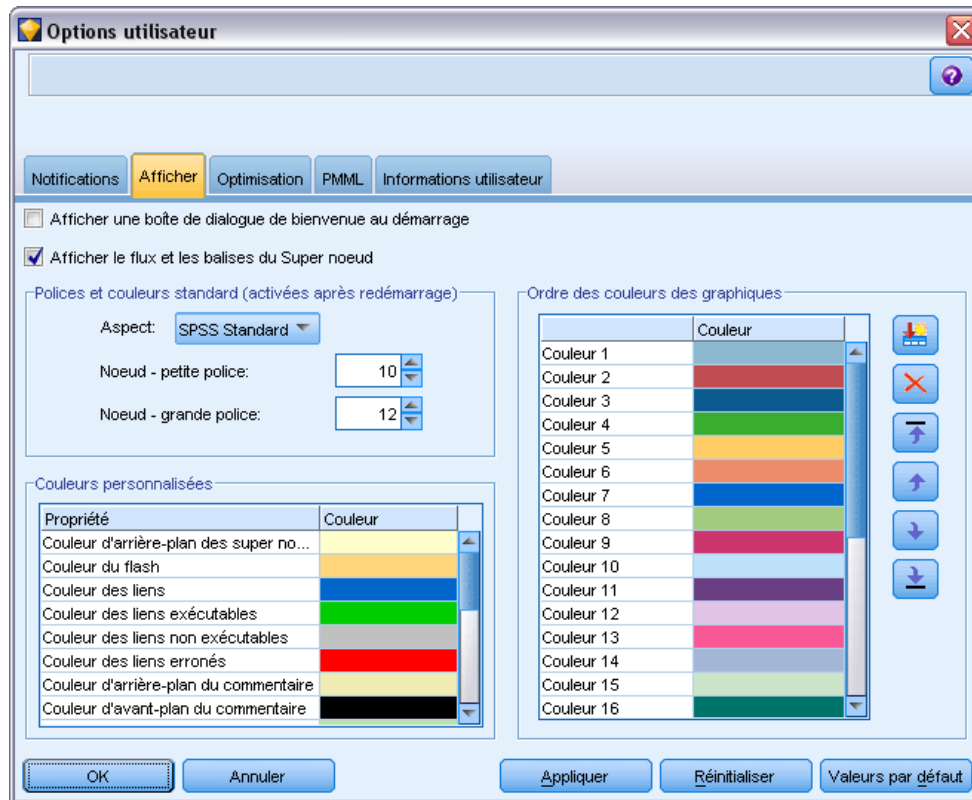
- Sélectionnez *Toujours* pour qu'une nouvelle fenêtre de sortie soit automatiquement ouverte.
- Sélectionnez *Si généré par le flux actuel* pour ouvrir une nouvelle fenêtre pour la sortie générée par le flux actuellement visible dans l'espace de travail.
- Sélectionnez *Jamais* pour empêcher que le logiciel ouvre automatiquement une nouvelle fenêtre pour la sortie générée.

Cliquez sur *Valeurs par défaut* pour revenir aux paramètres système par défaut de cet onglet.

Définition des options d'affichage

L'onglet *Afficher* de la boîte de dialogue *Options utilisateur* vous permet de définir les options d'affichage des polices et des couleurs dans IBM® SPSS® Modeler.

Figure 12-3
Boîte de dialogue Options utilisateur, onglet Afficher.



Afficher la boîte de dialogue de bienvenue au démarrage. Sélectionnez cette option pour que la boîte de dialogue de bienvenue s'affiche au démarrage. Cette boîte de dialogue contient des options permettant de lancer le tutoriel d'exemples de l'application, d'ouvrir un flux de démonstration ou un flux ou un projet existant, ou encore pour créer un nouveau flux.

Afficher les balisages de flux et de super noeud. Si cette option est sélectionnée, le balisage (s'il en existe) des flux et des super noeuds est affiché par défaut. Le balisage inclut la mise en surbrillance des commentaires de flux, des liens de modèle et des branches de scoring.

Polices & couleurs standard (effectives au redémarrage). Les options de cette boîte de dialogue permettent de définir le dessin de l'écran de SPSS Modeler, le modèle de couleur ainsi que la taille des polices affichées. Les options sélectionnées ici ne sont appliquées qu'après la fermeture et le redémarrage de SPSS Modeler.

- **Aspect.** Vous permet de sélectionner un modèle de couleurs et un dessin d'écran. Vous pouvez sélectionner

SPSS Standard (par défaut), une conception commune aux produits IBM SPSS.

SPSS Classic, une conception familière pour les utilisateurs des versions précédentes de SPSS Modeler.

Windows, un dessin Windows qui permet d'obtenir un contraste plus élevé dans l'espace de travail de flux et les palettes.

- **Noeud - petite police.** Indiquez la taille de police à utiliser dans les palettes de noeuds et pour les petits noeuds affichés dans l'espace de travail de flux.
- **Noeud - grande police.** Indiquez la taille de police à utiliser pour les grands noeuds (standard) affichés dans l'espace de travail de flux.

Remarque : Vous pouvez définir la taille de nœud d'un flux dans l'onglet Présentation de la boîte de dialogue des propriétés du flux.

Couleurs personnalisées. Ce tableau dresse la liste des couleurs sélectionnées actuellement utilisées pour les différents éléments d'affichage. Vous pouvez modifier la couleur actuelle de chaque élément répertorié dans la liste en double-cliquant sur la ligne correspondante dans la colonne Couleur et en sélectionnant la couleur dans la liste. Pour définir une couleur personnalisée, faites défiler la liste jusqu'en bas et cliquez sur l'entrée Couleur....

Ordre des couleurs des graphiques. Ce tableau répertorie les couleurs actuellement utilisées pour l'affichage des nouveaux graphiques. L'ordre des couleurs reflète celui dans lequel elles sont utilisées dans le graphique. Par exemple, si un champ nominal utilisé comme superposition de couleurs contient quatre valeurs uniques, seules les quatre premières couleurs répertoriées ici seront utilisées. Vous pouvez modifier la couleur actuelle de chaque élément répertorié dans la liste en double-cliquant sur la ligne correspondante dans la colonne Couleur et en sélectionnant la couleur dans la liste. Pour définir une couleur personnalisée, faites défiler la liste jusqu'en bas et cliquez sur l'entrée Couleur.... Les changements apportés ici n'affectent pas les graphiques créés précédemment.

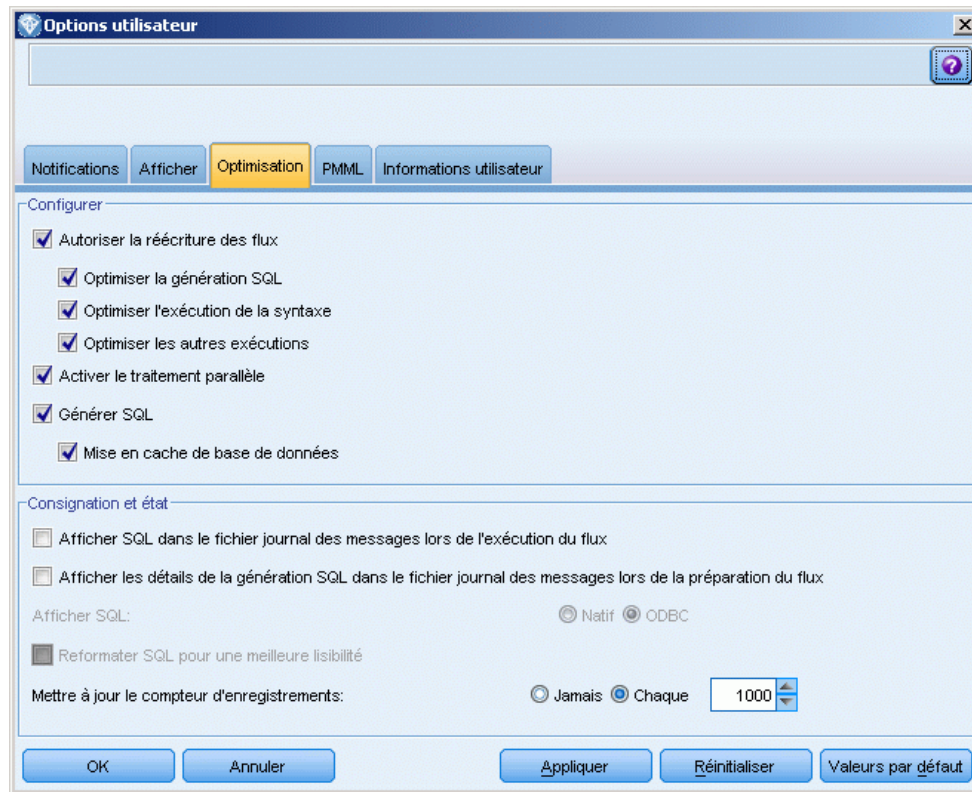
Cliquez sur Valeurs par défaut pour revenir aux paramètres système par défaut de cet onglet.

Définition des options d'optimisation

L'onglet Optimisation de la boîte de dialogue Options utilisateur vous permet d'optimiser les performances du flux. Les paramètres de performances et d'optimisation de IBM® SPSS® Modeler Server (si ce dernier est utilisé) annulent ceux définis pour le client. *Cet onglet n'est disponible que si la connectivité SPSS Modeler Server a été activée sur l'ordinateur IBM® SPSS® Modeler.*

Remarque : La modélisation de base de données et l'optimisation SQL requièrent l'activation de la connectivité à SPSS Modeler Server sur l'ordinateur SPSS Modeler. Avec ce paramètre activé, vous pouvez accéder aux algorithmes de la base de données, effectuer le push back de SQL directement depuis SPSS Modeler et accéder à SPSS Modeler Server. Pour vérifier l'état de la licence actuelle, cliquez sur A propos de dans le menu Aide de SPSS Modeler. [Pour plus d'informations, reportez-vous à la section Connexion au IBM SPSS Modeler Server dans le chapitre 3 sur p. 9.](#)

Figure 12-4
Boîte de dialogue Options utilisateur, onglet Optimisation.



Remarque : La prise en charge des répercussions et de l'optimisation SQL dépend du type de base de données utilisé. Pour obtenir les informations les plus récentes sur les bases de données et pilotes ODBC pris en charge et testés pour une utilisation avec IBM® SPSS® Modeler 14.2, consultez les matrices de compatibilité des produits sur le site Web de support technique de l'entreprise à l'adresse <http://www.ibm.com/support>.

Autoriser la réécriture des flux. Sélectionnez cette option pour activer la réécriture de flux dans SPSS Modeler. Deux types de réécriture sont disponibles. Vous pouvez sélectionner l'une de ces options ou les deux. La réécriture de flux réorganise les noeuds d'un flux en arrière-plan, sans pour autant modifier sa sémantique, afin qu'il soit exécuté de manière plus efficace.

- **Optimiser la génération SQL.** Cette option permet de réorganiser les noeuds du flux afin d'augmenter le nombre d'opérations pouvant être répercutées, à l'aide d'une génération SQL, pour y être exécutées dans la base de données. Lorsqu'il détecte un noeud qui ne peut pas être affiché sous forme de code SQL, le programme d'optimisation recherche plus en avant tout noeud qui peut l'être et qu'il est possible de ramener avant le noeud problématique sans risquer de modifier la sémantique du flux. Non seulement la base de données peut effectuer les opérations plus efficacement que SPSS Modeler, mais de telles répercussions contribuent également à réduire le volume des données renvoyées à SPSS Modeler pour être traitées. Vous réduisez ainsi le trafic sur le réseau et accélérez les opérations de flux. Veuillez noter que la case Générer SQL doit être sélectionnée pour que l'optimisation SQL ait le moindre effet.

Pour plus d'informations, reportez-vous à la section [Optimisation SQL](#) dans le chapitre 6 dans *Administration et guide des performances de IBM SPSS Modeler Server 14.2*.

- **Optimiser l'exécution de syntaxe.** Cette méthode de réécriture de flux augmente l'efficacité des opérations qui incorporent plus d'un noeud contenant une syntaxe IBM® SPSS® Statistics. L'optimisation est réalisée en combinant des commandes de syntaxe dans une opération unique, au lieu d'exécuter chacune d'elle en tant qu'opération séparée.
- **Optimiser les autres exécutions.** Cette méthode de réécriture de flux augmente l'efficacité des opérations exécutées (c'est-à-dire des opérations dont l'exécution ne peut pas être déléguée à la base de données). L'optimisation est obtenue en réduisant dès que possible la quantité de données dans le flux. Tout en conservant l'intégrité des données, le flux est réécrit de façon à rapprocher les opérations de la source de données, ceci afin de réduire le volume des données en aval pour les opérations coûteuses, telles que les jointures.

Activer le traitement parallèle. Lorsque vous exécutez ce programme sur un ordinateur doté de plusieurs processeurs, cette option permet au système d'équilibrer la charge entre les différents processeurs et d'augmenter ainsi les performances. Le traitement parallèle peut s'avérer bénéfique en cas d'utilisation de plusieurs noeuds ou des noeuds individuels suivants : C5.0, Fusionner (par clé), Trier, Discrétiser (méthodes des rangs et des quantiles) et Agréger (à l'aide d'un ou de plusieurs champs-clés).

Générer SQL. Sélectionnez cette option pour activer la génération SQL et répercuter ainsi les opérations de flux dans la base de données à l'aide du code SQL pour générer les processus d'exécution ; cette opération peut améliorer les performances. Pour augmenter encore les performances, vous pouvez également sélectionner l'option [Optimiser la génération SQL](#), afin d'accroître au maximum le nombre d'opérations répercutées dans la base de données. Si les opérations d'un noeud ont été retransmises à la base de données, le noeud apparaît en violet au cours de l'exécution du flux.

- **Mise en cache de base de données.** Dans le cas des flux exécutés dans la base de données qui génèrent du code SQL, les données peuvent être mises en mémoire cache en milieu de flux dans une table temporaire de la base de données, plutôt que dans le système de fichiers. Vous pouvez combiner cela à la fonction d'optimisation SQL et accroître ainsi les performances de manière significative. Par exemple, la sortie d'un flux qui fusionne plusieurs tables afin de créer une vue de Data mining peut être mise en mémoire cache et réutilisée selon les besoins. Une fois la mise en mémoire cache de base de données activée, il vous suffit de cliquer à l'aide du bouton droit de la souris sur un noeud non terminal afin de mettre les données en mémoire cache à ce stade du flux ; le cache est automatiquement créé dans la base de données lors de l'exécution suivante du flux. Le code SQL peut ainsi être généré pour les noeuds en aval, ce qui améliore encore les performances. Vous pouvez également désactiver cette option si besoin est, par exemple lorsque des règles ou des permissions empêchent l'écriture des données dans la base de données. Si la mise en mémoire cache de base de données ou l'optimisation SQL n'est pas activée, le cache est écrit dans le système de fichiers. [Pour plus d'informations, reportez-vous à la section Options de mise en cache des noeuds dans le chapitre 5 sur p. 47.](#)

Remarque : En raison de différences d'implémentation SQL mineures, les flux exécutés dans une base de données peuvent renvoyer des résultats légèrement différents que ceux renvoyés lors de leur exécution dans SPSS Modeler. Pour ces mêmes raisons, ces différences peuvent également varier en fonction du fournisseur de base de données.

Afficher SQL dans le fichier journal des messages lors de l'exécution du flux. Détermine l'insertion dans le journal des messages du code SQL généré lors de l'exécution du flux.

Afficher les détails de la génération SQL dans le fichier journal des messages lors de la préparation du flux. Au cours de la prévisualisation du flux, détermine l'insertion dans le journal des messages d'un aperçu du code SQL susceptible d'être généré.

Afficher SQL. Indique si le code SQL affiché dans le journal doit contenir des fonctions SQL natives ou des fonctions ODBC standard (générées par SPSS Modeler au format {fn FUNC(...)}). Les premières reposent sur certaines fonctionnalités du pilote ODBC, qui ne sont peut-être pas implémentées. Ainsi, cette commande n'aurait aucun impact dans SQL Server.

Reformater SQL pour une meilleure lisibilité. Indique si le SQL affiché dans le journal doit être formaté pour être lu.

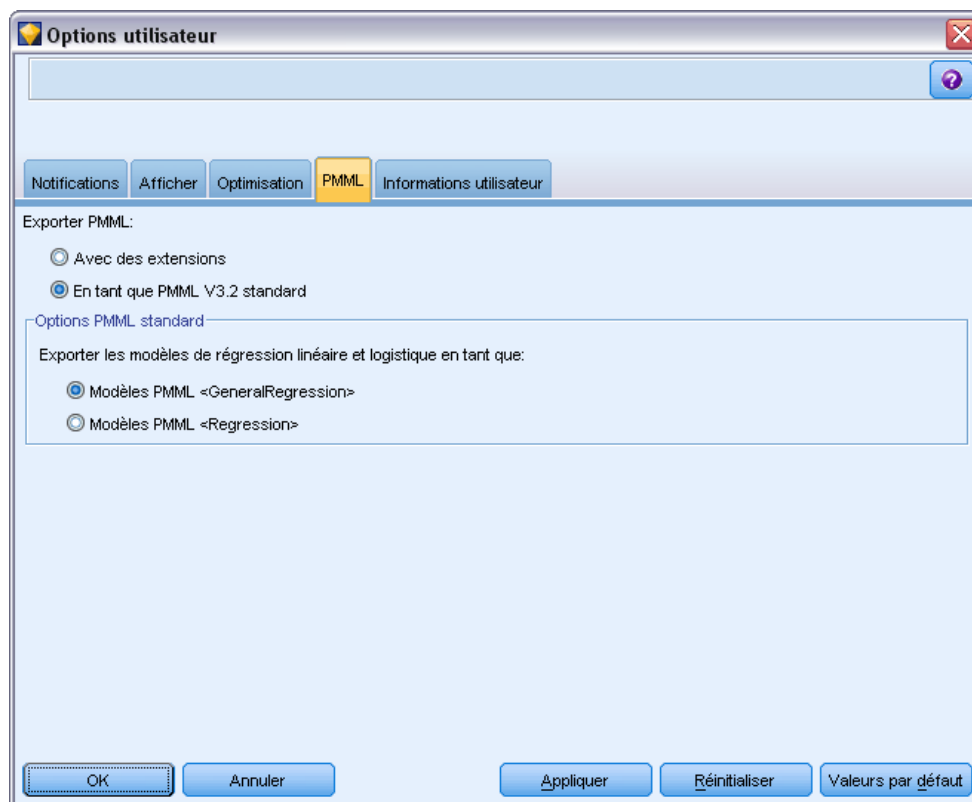
Mettre à jour le compteur d'enregistrements. Indique le moment où les enregistrements doivent être enregistrés à mesure qu'ils arrivent dans les noeuds terminaux. Indiquez le chiffre déterminant la mise à jour de l'état tous les N enregistrements.

Cliquez sur Valeurs par défaut pour revenir aux paramètres système par défaut de cet onglet.

Définition des options d'exportation PMML

Dans l'onglet PMML, vous pouvez déterminer la manière dont IBM® SPSS® Modeler exporte des modèles au format PMML. [Pour plus d'informations, reportez-vous à la section Importation et exportation de modèles au format PMML dans le chapitre 10 sur p. 209.](#)

Figure 12-5
Boîte de dialogue Options utilisateur, onglet PMML.



Exporter PMML. Vous pouvez configurer ici les variations du langage PMML les plus adaptées à votre application cible.

- Sélectionnez l'option avec les extensions pour autoriser les extensions PMML dans les cas de figure spéciaux où aucun équivalent PMML standard n'existe. Dans la plupart des cas, cela produira le même résultat qu'avec le langage PMML standard.
- Sélectionnez PMML V3.2 standard pour exporter les données PMML qui se conforment au maximum à la norme PMML.

Options PMML standard. Lorsque vous sélectionnez l'option PMML standard..., vous avez le choix entre deux méthodes valides pour exporter des modèles de régression linéaire et logistique :

- En tant que modèles PMML <GeneralRegression>
- En tant que modèles PMML <Regression>

Définition des informations utilisateur

Informations sur l'utilisateur/auteur. Les informations que vous saisissez ici peuvent être affichées dans l'onglet Annotations des noeuds et des autres objets que vous créez.

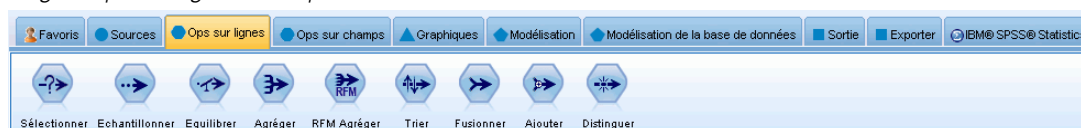
Personnalisation de la palette Noeuds

Les flux sont créés à l'aide de noeuds. La palette Noeuds située au bas de la fenêtre IBM® SPSS® Modeler contient tous les noeuds pouvant être utilisés pour la création des flux. [Pour plus d'informations, reportez-vous à la section Palette de noeuds dans le chapitre 3 sur p. 15.](#)

Vous pouvez réorganiser la palette Noeuds de deux façons :

- Personnaliser le gestionnaire de palette. [Pour plus d'informations, reportez-vous à la section Personnalisation du gestionnaire de palette. sur p. 242.](#)
- Modifier l'affichage des onglets de la palette contenant des sous-palettes dans la palette Noeuds. [Pour plus d'informations, reportez-vous à la section Création d'une sous-palette sur p. 247.](#)

Figure 12-6
Onglet Ops sur lignes de la palette Noeuds



Personnalisation du gestionnaire de palette.

Le gestionnaire de palette peut être personnalisé en fonction de votre utilisation de IBM® SPSS® Modeler. Par exemple, si vous analysez fréquemment des données de séries temporelles à partir d'une base de données, vous souhaitez peut-être vous assurer que le noeud source Base de données, le noeud des intervalles temporels, le noeud Séries temporelles et le noeud du graphique Tracé horaire soient disponibles ensemble à partir d'un seul onglet de la palette. Le gestionnaire de palette vous permet d'apporter facilement ces modifications en créant vos onglets de palette personnalisés dans la palette de noeuds.

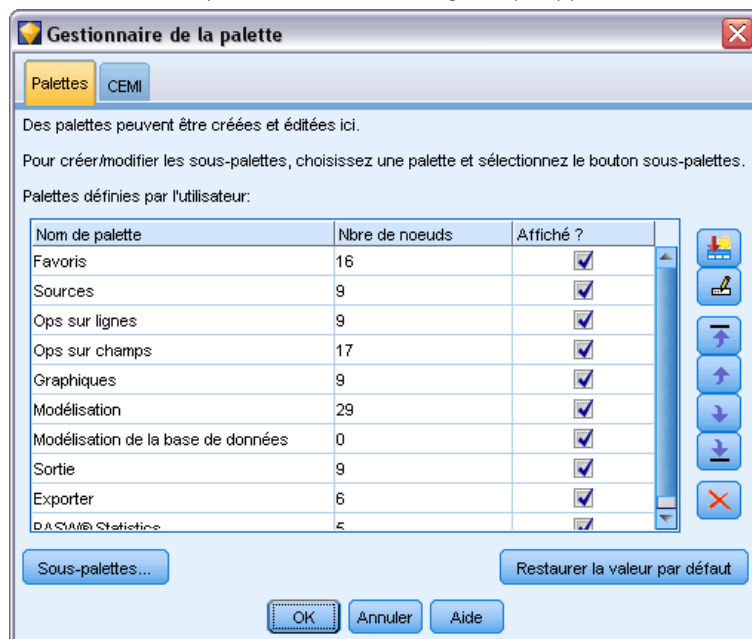
Le gestionnaire de palette vous permet d'effectuer différentes tâches :

- Contrôler quels onglets de la palette apparaissent dans la palette de noeuds sous l'espace de travail du flux.
- Modifier l'ordre d'affichage des onglets de la palette dans la palette de noeuds.
- Créer et modifier vos propres onglets de palette et des éventuelles sous-palettes associées.
- Modifier les sélections de noeud par défaut dans votre onglet Favoris.

Pour accéder au Gestionnaire de la palette :

- Dans le menu Outils, cliquez sur Gérer les palettes.

Figure 12-7
Gestionnaire de la palette affichant les onglets qui apparaissent dans la palette de noeuds.



Nom de la palette. Chaque onglet de palette disponible est répertorié, qu'il apparaisse ou non dans la palette de noeuds. Cela comprend les éventuels onglets de palette que vous avez créés. [Pour plus d'informations, reportez-vous à la section Création d'un onglet Palette sur p. 244.](#)

Nbre de noeuds. Le nombre de noeuds affichés dans chaque onglet de la palette. Un nombre élevé ici signifie qu'il vous sera peut-être plus pratique de créer des sous-palettes afin de diviser les noeuds dans l'onglet. [Pour plus d'informations, reportez-vous à la section Création d'une sous-palette sur p. 247.](#)

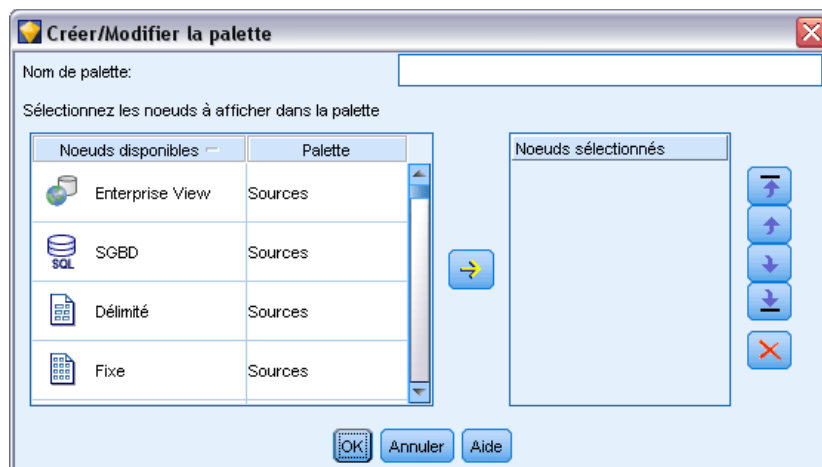
Affiché ? Sélectionnez ce champ pour afficher l'onglet palette dans la palette de noeuds. [Pour plus d'informations, reportez-vous à la section Affichage des onglets Palette dans la palette de noeuds sur p. 244.](#)

Sous-palettes. Pour sélectionner des sous-palettes à afficher dans un onglet de palette, mettez en surbrillance le Nom de la palette requis et cliquez sur ce bouton pour afficher la boîte de dialogue Sous-palettes. [Pour plus d'informations, reportez-vous à la section Création d'une sous-palette sur p. 247.](#)

Rétablir les valeurs par défaut. Pour supprimer complètement tous les changements et les ajouts apportés aux palettes et aux sous-palettes et revenir aux paramètres de palette par défaut, cliquez sur ce bouton.

Création d'un onglet Palette

Figure 12-8
Création d'un onglet Palette dans la boîte de dialogue Créer/Editer une palette



Pour créer un onglet de palette personnalisé :

- ▶ Dans le menu Outils, ouvrez le gestionnaire de la palette.
- ▶ A droite de la colonne *Affiché ?*, cliquez sur le bouton Ajouter une palette ; la boîte de dialogue Créer/Editer une palette s'affiche.
- ▶ Saisissez un Nom de la palette unique.
- ▶ Dans la zone Noeuds disponibles, sélectionnez le noeud à ajouter à l'onglet palette.
- ▶ Cliquez sur la flèche droite Ajouter un noeud pour déplacer le(s) noeud(s) sélectionné à la zone Noeuds sélectionnés. Répétez l'opération jusqu'à ce que vous ayez ajouté tous les noeuds de votre choix.

Une fois que vous avez ajouté tous les noeuds requis, vous pouvez modifier l'ordre de leur apparition dans l'onglet palette :

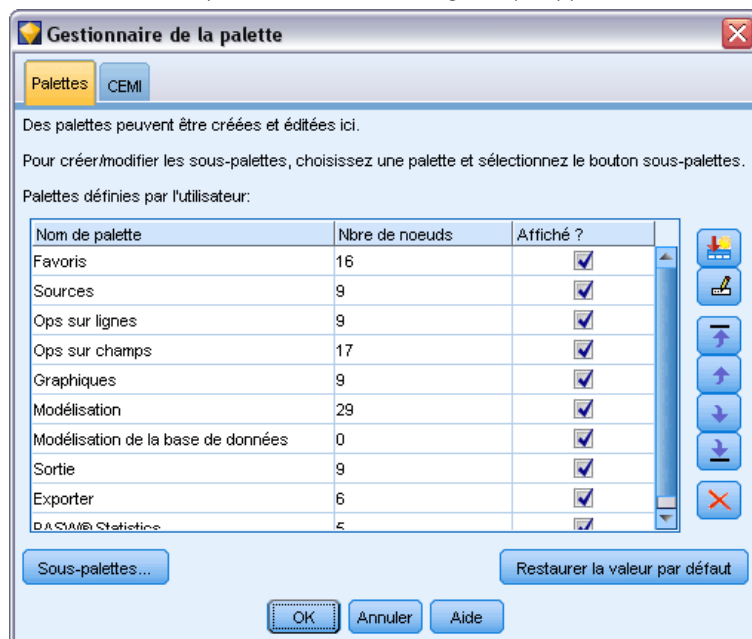
- ▶ Utilisez les boutons représentant une simple flèche pour déplacer un noeud d'un rang vers le haut ou vers le bas.
- ▶ Utilisez les boutons représentant une flèche et une ligne pour placer un noeud tout en bas ou tout en haut de la liste.
- ▶ Pour supprimer un noeud d'une palette, mettez le noeud en surbrillance et cliquez sur le bouton Supprimer à droite de la zone Noeuds sélectionnés.

Affichage des onglets Palette dans la palette de noeuds

Il peut exister dans IBM® SPSS® Modeler des options que vous n'utilisez jamais ; dans ce cas, vous pouvez utiliser le gestionnaire de palette pour masquer les onglets contenant ces noeuds.

Figure 12-9

Gestionnaire de la palette affichant les onglets qui apparaissent dans la palette de noeuds.



Pour sélectionner les onglets à afficher dans la palette de noeuds :

- ▶ Dans le menu Outils, ouvrez le gestionnaire de la palette.
- ▶ A l'aide des cases à cocher figurant dans la colonne *Affiché ?*, choisissez s'il faut inclure ou masquer chaque onglet de la palette.

Pour supprimer définitivement un onglet de la palette de noeuds, mettez le noeud en surbrillance et cliquez sur le bouton Supprimer à droite de la colonne *Affiché ?*. Une fois supprimé, un onglet de palette ne peut pas être récupéré.

Remarque : Vous ne pouvez pas supprimer les onglets par défaut de la palette fournis avec SPSS Modeler, à l'exception de l'onglet Favoris.

Modification de l'ordre d'affichage dans la palette de noeuds

Une fois que vous avez sélectionné les onglets de la palette à afficher, vous pouvez modifier leur ordre d'affichage dans la palette de noeuds :

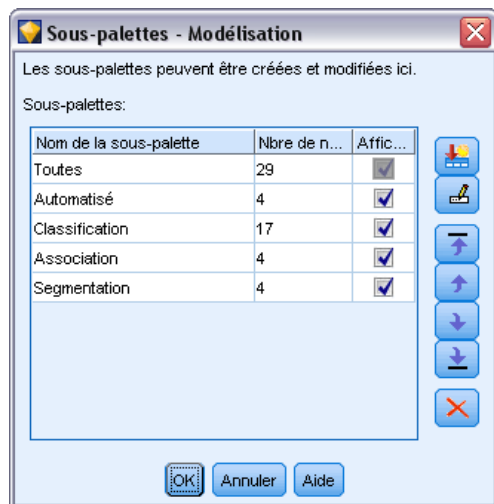
- ▶ Utilisez les boutons représentant une simple flèche pour déplacer un noeud de la palette d'un rang vers le haut ou vers le bas. Un déplacement vers le haut les déplace vers la gauche de la palette de noeuds, et vice versa.
- ▶ Utilisez les boutons représentant une flèche et une ligne pour placer un onglet de la palette tout en bas ou tout en haut de la liste. Ceux qui apparaissent en haut de la liste figureront à gauche de la palette de noeuds.

Affichage des sous-palettes sur un onglet Palette

De même que vous pouvez contrôler les onglets de palette qui s'affichent dans la palette de noeuds, vous pouvez contrôler les sous-palettes disponibles à partir de leur onglet de palette parent.

Figure 12-10

Sous-palettes disponibles pour l'onglet Modélisation de la palette



Pour sélectionner les sous-palettes à afficher dans un onglet palette :

- ▶ Dans le menu Outils, ouvrez le gestionnaire de la palette.
- ▶ Sélectionnez la palette de votre choix.
- ▶ Cliquez sur le bouton Sous-palettes ; la boîte de dialogue Sous-palettes s'affiche.
- ▶ A l'aide des cases à cocher figurant dans la colonne *Affiché ?*, choisissez s'il faut inclure chaque sous-palette dans l'onglet palette. La sous-palette Tous est toujours affichée et ne peut pas être supprimée.
- ▶ Pour supprimer définitivement une sous-palette de l'onglet palette, mettez la sous-palette en surbrillance et cliquez sur le bouton Supprimer à droite de la colonne *Affiché ?*.

Remarque : Vous ne pouvez pas supprimer les sous-palettes par défaut fournies avec l'onglet Modélisation de la palette.

Modification de l'ordre d'affichage dans l'onglet Palette

Une fois que vous avez sélectionné les onglets de la palette à afficher, vous pouvez modifier leur ordre d'affichage dans l'onglet de la palette parent :

- ▶ Utilisez les boutons représentant une simple flèche pour déplacer une sous-palette d'un rang vers le haut ou vers le bas.
- ▶ Utilisez les boutons représentant une flèche et une ligne pour placer une sous-palette tout en bas ou tout en haut de la liste.

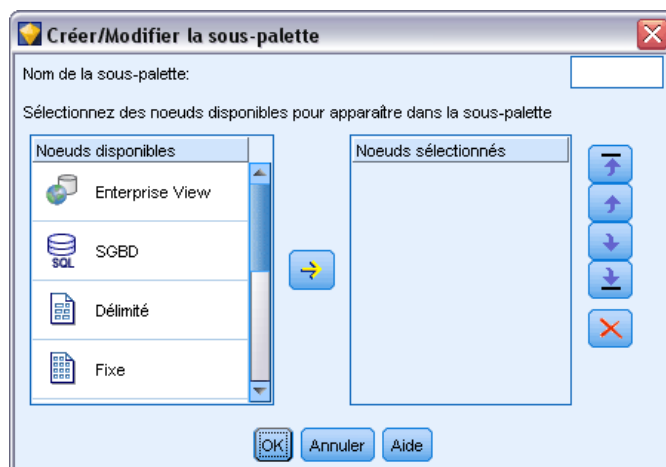
Les sous-palettes que vous créez sont affichées dans la palette de noeuds lorsque vous sélectionnez l'onglet palette parent. [Pour plus d'informations, reportez-vous à la section Modification de l'affichage d'un onglet Palette sur p. 248.](#)

Création d'une sous-palette

Comme vous pouvez ajouter tout noeud existant aux onglets de palette personnalisés que vous créez, vous sélectionnez peut-être davantage de noeuds qu'on ne peut en afficher facilement à l'écran sans défilement. Pour éviter d'avoir à défiler, vous pouvez créer des sous-palettes dans lesquelles vous placerez les noeuds choisis pour l'onglet palette. Par exemple, si vous avez créé un onglet de palette contenant les noeuds que vous utilisez le plus fréquemment pour créer vos flux, vous pourriez créer quatre sous-palettes divisant les sélections par noeud source, opérations de champ, modélisation et sortie.

Remarque : Vous ne pouvez sélectionner la sous-palette que parmi celles qui sont ajoutées à l'onglet palette parent.

Figure 12-11
Création d'une sous-palette dans la boîte de dialogue Créer/Editer une sous-palette



Pour créer une sous-palette :

- ▶ Dans le menu Outils, ouvrez le gestionnaire de la palette.
- ▶ Sélectionnez la palette à laquelle vous souhaitez ajouter les sous-palettes.
- ▶ Cliquez sur le bouton Sous-palettes ; la boîte de dialogue Sous-palettes s'affiche.
- ▶ A droite de la colonne *Affiché ?*, cliquez sur le bouton Ajouter une sous-palette ; la boîte de dialogue Créer/Editer une sous-palette s'affiche.
- ▶ Saisissez un nom de sous-palette unique.
- ▶ Dans la zone Noeuds disponibles, sélectionnez le noeud à ajouter à l'onglet sous-palette.
- ▶ Cliquez sur la flèche droite Ajouter un noeud pour déplacer un noeud sélectionné dans la zone Noeuds sélectionnés.

- Lorsque vous avez ajouté les noeuds requis, cliquez sur OK pour revenir à la boîte de dialogue Sous-palettes.

Les sous-palettes que vous créez sont affichées dans la palette de noeuds lorsque vous sélectionnez l'onglet palette parent. [Pour plus d'informations, reportez-vous à la section Modification de l'affichage d'un onglet Palette sur p. 248.](#)

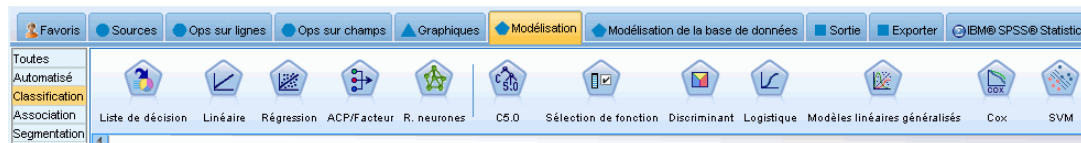
Modification de l'affichage d'un onglet Palette

En raison du grand nombre de noeuds disponibles dans IBM® SPSS® Modeler, ils peuvent ne pas être tous visibles dans les petits écrans sans défilement vers la gauche ou vers la droite de la palette de noeuds ; cela est particulièrement remarquable dans l'onglet de la palette Modélisation. Pour réduire les besoins de défilement, vous pouvez choisir d'afficher uniquement les noeuds figurant dans une sous-palette (le cas échéant). [Pour plus d'informations, reportez-vous à la section Création d'une sous-palette sur p. 247.](#)

Pour modifier les noeuds affichés dans l'onglet palette ; sélectionnez l'onglet palette puis, dans le menu de gauche, choisissez d'afficher soit tous les noeuds, soit seulement les noeuds figurant dans une sous-palette spécifique.

Figure 12-12

Onglet de la palette Modélisation illustrant la sous-palette Classification



Gestion du noeud CEMI

CEMI est désormais rejeté et a été remplacé par CLEF, qui offre un ensemble de fonctions plus flexibles et plus simples d'utilisation. Pour plus d'informations, consultez le *IBM® SPSS® Modeler 14.2 CLEF Guide du développeur* fourni avec cette version.

Remarques relatives aux performances des flux et des noeuds

Vous pouvez concevoir vos flux pour maximiser les performances en organisant les noeuds selon la configuration la plus efficace, en activant les caches de noeud le cas échéant et en tenant compte des autres critères décrits dans cette section.

En plus des critères décrits ici, des améliorations supplémentaires et plus importantes des performances peuvent généralement être obtenues en utilisant de manière efficace votre base de données, en particulier à travers l'optimisation SQL. [Pour plus d'informations, reportez-vous à la section Considérations relatives aux performances dans le chapitre 5 dans *Administration et guide des performances de IBM SPSS Modeler Server 14.2*.](#)

Ordre des noeuds

Même si vous n'utilisez pas l'optimisation SQL, l'ordre des noeuds d'un flux peut avoir une incidence sur les performances. L'objectif principal consiste à réduire au maximum le traitement en aval ; ainsi, lorsque vous disposez de noeuds qui permettent de réduire la quantité de données, placez-les au début du flux. IBM® SPSS® Modeler Server peut appliquer automatiquement plusieurs règles de réorganisation au cours de la compilation et faire avancer certains noeuds lorsque cette opération s'avère sûre. (Cette fonctionnalité est activée par défaut. Assurez-vous auprès de l'administrateur système que c'est le cas pour votre installation.)

Lorsque vous utilisez l'optimisation SQL, vous souhaitez augmenter au maximum sa disponibilité et son efficacité. L'optimisation s'arrête lorsque le flux contient une opération qui ne peut pas être exécutée dans la base de données ; par conséquent, il s'avère plus judicieux de regrouper les opérations optimisées SQL au début du flux. Grâce à cette méthode, une plus grande partie du traitement est confinée dans la base de données et une moindre quantité de données est ainsi transmise dans IBM® SPSS® Modeler.

Les opérations ci-dessous peuvent être réalisées dans la plupart des bases de données. Essayez de les regrouper au *début* du flux :

- Fusion par clé (jointure)
- Select
- Agréger
- Trier
- Echantillon
- Ajouter
- Opérations de distinction en mode *inclure* (tous les champs sont sélectionnés)

- Opérations de remplacement
- Opérations de calcul de base qui utilisent une manipulation arithmétique ou une manipulation de chaîne standard (en fonction des opérations prises en charge par la base de données)
- Binarisation

Les opérations ci-dessous ne peuvent pas être réalisées dans la plupart des bases de données. Elles doivent donc être placées dans le flux *après* celles mentionnées dans la liste précédente :

- Opérations effectuées sur des données non incluses dans une base de données (les fichiers plats, par exemple)
 - Fusion par ordre
 - Balance
 - Opérations de distinction en mode *supprimer* (seul un sous-ensemble de champs est sélectionné comme étant distinct)
 - Toute opération qui requiert un accès aux données à partir d'enregistrements autres que celui en cours de traitement
 - Calculs de champ (état et comptage)
 - opérations relatives au noeud *Historiser*
 - Opérations impliquant des fonctions "@" (séries temporelles)
 - Modes de vérification du type *Avertir* et *Stopper*
 - Construction, application et analyse de modèle
- Remarque* : Les modèles d'arbre décision, d'ensemble de règles et de régression linéaire, ainsi que ceux générés par facteur, peuvent tous générer du code SQL et donc être renvoyés vers la base de données.
- Sortie de données vers tout emplacement autre que la base de données qui traite les données

Caches de noeud

Pour optimiser l'exécution du flux, vous pouvez définir un **cache** sur n'importe quel nœud non terminal. Lorsque vous définissez un cache sur un nœud, le cache est alimenté par les données passant par le nœud lors de l'exécution suivante du flux de données. Les données sont alors lues à partir du cache (stocké sur le disque dans un répertoire temporaire) et non à partir de la source de données.

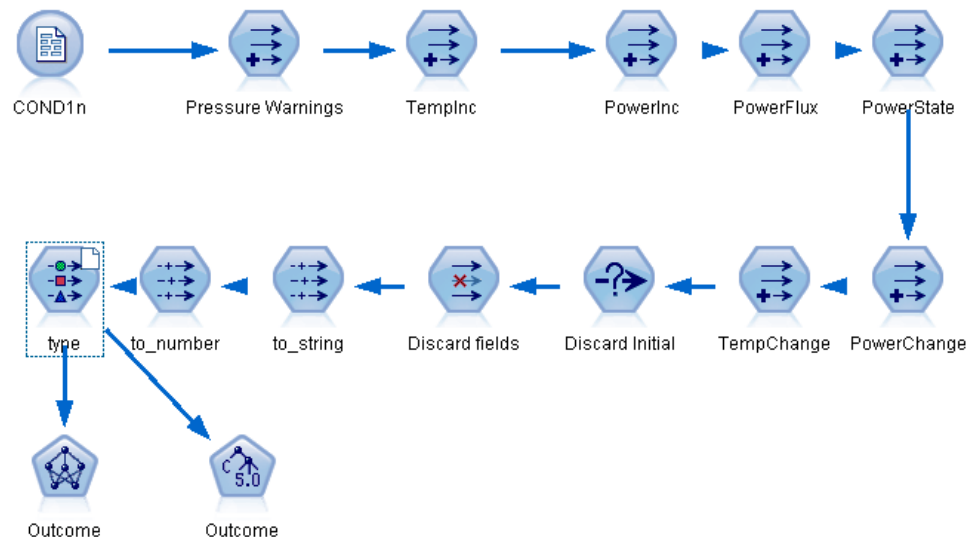
La mise en cache est plus utile après une opération de longue durée telle que le tri, la fusion ou l'agrégation. Par exemple, supposez que vous ayez un nœud source configuré pour la lecture des données relatives aux ventes à partir de la base de données et un nœud Agréger qui fournit des résultats de ventes par zone géographique. Vous pouvez définir un cache sur le nœud Agréger plutôt que sur le nœud source pour que le cache stocke uniquement les données agrégées et non l'ensemble des données.

Remarque : La mise en cache au niveau des nœuds source, qui stocke simplement une copie des données originales telle qu'elle est lue dans IBM® SPSS® Modeler, n'améliorera pas les performances dans la plupart des cas.

Lorsque la mise en cache d'un noeud est activée, une petite icône en forme de document apparaît dans l'angle supérieur droit. Une fois les données mises en cache au niveau du noeud, l'icône de document devient verte.

Figure 13-1

Mise en cache au niveau du noeud Typer pour stocker les champs nouvellement calculés



Pour activer un cache

- ▶ Dans l'espace de travail de flux, cliquez avec le bouton droit de la souris sur le noeud, puis cliquez sur Cache dans le menu.
- ▶ Dans le sous-menu de mise en cache, cliquez sur Activer.
- ▶ Vous pouvez désactiver le cache en cliquant sur le noeud avec le bouton droit de la souris et en cliquant sur Désactiver dans le sous-menu de mise en cache.

Mise en mémoire cache des noeuds dans une base de données

Dans le cas des flux exécutés dans une base de données, les données peuvent être mises en mémoire cache en milieu de flux dans une table temporaire de la base de données, plutôt que dans le système de fichiers. Vous pouvez combiner cela à la fonction d'optimisation SQL et accroître ainsi les performances de manière significative. Par exemple, la sortie d'un flux qui fusionne plusieurs tables afin de créer une vue de Data mining peut être mise en mémoire cache et réutilisée selon les besoins. La génération automatique du code SQL pour tous les noeuds en aval peut améliorer encore les performances.

Lors de l'utilisation de la mise en mémoire cache de la base de données avec des chaînes d'une longueur supérieure à 256 caractères, vérifiez qu'il existe un noeud Typer en amont depuis la mise en mémoire cache du noeud et que les valeurs des champs sont lues. Cela permet de garantir que la colonne correspondante dans la table temporaire est définie sur la largeur appropriée pour la prise en charge des chaînes.

Pour tirer profit de la mise en mémoire cache de base de données, vous devez activer les fonctions d'optimisation SQL et de mise en mémoire cache de base de données. Les paramètres d'optimisation du serveur annulent ceux définis pour le client. [Pour plus d'informations, reportez-vous à la section Définition des options d'optimisation dans le chapitre 12 sur p. 237.](#)

Une fois la mise en mémoire cache de base de données activée, il vous suffit de cliquer à l'aide du bouton droit de la souris sur un noeud non terminal afin de mettre les données en mémoire cache à ce stade du flux ; le cache est automatiquement créé dans la base de données lors de l'exécution suivante du flux. Si la mise en mémoire cache de base de données ou l'optimisation SQL n'est pas activée, le cache est écrit dans le système de fichiers.

Remarque : Les bases de données suivantes prennent en charge les tables temporaires à des fins de mise en mémoire cache : DB2, Netezza, Oracle, SQL Server et Teradata. Les autres bases de données utilisent une table standard pour la mise en mémoire cache de base de données. Si nécessaire, le code SQL peut être personnalisé pour des bases de données spécifiques : contactez l'assistance technique.

Performances : noeuds d'exécution

Trier. Le noeud Trier doit lire l'ensemble de données d'entrée dans son intégralité pour pouvoir le trier. Les données sont stockées en mémoire jusqu'à une certaine limite, et l'excédent de données est placé sur le disque. L'algorithme de tri consiste en un algorithme combinatoire : les données sont lues en mémoire jusqu'à ce que la limite soit atteinte, puis elles sont triées à l'aide d'un algorithme hybride de tri rapide. Si toutes les données tiennent en mémoire, le tri est terminé. Sinon, un algorithme de fusion/tri est appliqué. Les données triées sont écrites dans un fichier et le fragment de données suivant est lu en mémoire, trié, puis écrit sur le disque. Cette opération est répétée jusqu'à ce que toutes les données soient lues, suite à quoi les fragments triés sont fusionnés. L'opération de fusion peut nécessiter plusieurs passages sur les données stockées sur le disque. Lors d'une utilisation intensive du système, le noeud Trier place deux copies complètes des données sur le disque : l'une triée et l'autre non.

Le temps d'exécution global de l'algorithme est de l'ordre de $N \cdot \log(N)$, où N correspond au nombre d'enregistrements. Le tri en mémoire s'avère plus rapide que la fusion à partir du disque ; vous pouvez, par conséquent, réduire le temps d'exécution réel en allouant davantage de mémoire à l'opération de tri. L'algorithme s'alloue une partie de la mémoire RAM physique contrôlée par l'option de configuration *Multiplicateur d'utilisation de la mémoire* de IBM® SPSS® Modeler Server. Pour accroître la quantité de mémoire utilisée pour le tri, fournissez davantage de mémoire RAM physique ou augmentez cette valeur. Lorsque la quantité de mémoire utilisée dépasse l'ensemble de travail du processus de telle sorte qu'une partie de la mémoire est paginée sur le disque, les performances sont réduites ; en effet, le modèle d'accès mémoire de l'algorithme de tri en mémoire est aléatoire et peut entraîner une pagination excessive. L'algorithme de tri est utilisé par plusieurs autres noeuds que Trier, mais les mêmes règles de performances s'appliquent.

Discrétiser. Le noeud Discrétiser lit l'ensemble de données d'entrée dans son intégralité afin de calculer les limites d'intervalle et d'allouer des enregistrements à ces intervalles. L'ensemble de données est mis en mémoire cache pendant le calcul des limites ; il est ensuite analysé à nouveau pour procéder à l'allocation. Lorsque vous choisissez la méthode de création d'intervalles *Largeur fixe* ou *Moyenne et écart-type*, l'ensemble de données est mis en mémoire cache directement sur le disque. Ces méthodes présentent un temps d'exécution linéaire et requièrent

suffisamment d'espace disque pour stocker l'intégralité de l'ensemble de données. Si vous utilisez la méthode de création d'intervalles *Rangs* ou *Quantiles*, l'ensemble de données est trié à l'aide de l'algorithme de tri décrit précédemment, puis l'ensemble de données trié est utilisé comme cache. Le tri confère à ces méthodes un temps d'exécution de $M*N*\log(N)$, où M correspond au nombre de champs mis en intervalles et N au nombre d'enregistrements ; cette opération nécessite une quantité d'espace disque équivalant à deux fois la taille de l'ensemble de données.

La génération d'un noeud Calculer basée sur des intervalles générés permet d'améliorer les performances lors des passages suivants. Les opérations de calcul sont bien plus rapides que la création d'intervalles. [Pour plus d'informations, reportez-vous à la section Prévisualisation des intervalles générés dans le chapitre 4 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Fusion par clé (jointure). Le noeud Fusionner, lorsque vous utilisez la méthode *Clés* (qui équivaut à une jointure de base de données), trie chaque ensemble de données d'entrée en fonction des champs-clés. Cette partie de la procédure présente un temps d'exécution de $M*N*\log(N)$, où M correspond au nombre d'entrées et N au nombre d'enregistrements contenus dans l'ensemble de données d'entrée le plus volumineux ; elle nécessite une quantité suffisante d'espace disque pour pouvoir stocker tous les ensembles de données d'entrée plus une deuxième copie de l'ensemble de données le plus volumineux. Le temps d'exécution de la fusion à proprement parler est proportionnel à la taille de l'ensemble de données de sortie, qui dépend de la fréquence des clés correspondantes. Dans le cas le plus extrême, où la sortie équivaut au produit cartésien des entrées, le temps d'exécution peut avoisiner NM . Ce cas est rare—la plupart des jointures comportent bien moins de clés correspondantes. Si un ensemble de données est sensiblement plus volumineux que les autres ou si les données entrantes sont déjà triées en fonction d'un champ-clé, vous pouvez améliorer les performances du noeud par le biais de l'onglet Optimisation. [Pour plus d'informations, reportez-vous à la section Paramètres d'optimisation de la fusion dans le chapitre 3 dans IBM SPSS Modeler 14.2 Noeuds source, exécution et de sortie.](#)

Agréger. Lorsque l'option *Les clés sont adjacentes* n'est pas définie, ce noeud lit (mais ne stocke pas) l'intégralité de l'ensemble de données d'entrée avant de générer une sortie agrégée. Dans les cas les plus extrêmes, où la taille des données agrégées atteint une limite (déterminée par l'option de configuration *Multiplieur d'utilisation de la mémoire* de SPSS Modeler Server), le reste de l'ensemble de données est trié et traité comme si l'option *Les clés sont adjacentes* était activée. Lorsque cette option est définie, aucune donnée n'est stockée car les enregistrements de sortie agrégés sont générés au fur et à mesure de la lecture des données d'entrée.

Distinguer. Le noeud Distinguer stocke tous les champs-clés uniques dans l'ensemble de données d'entrée ; dans les cas où tous les champs sont des champs-clés et tous les enregistrements sont uniques, il stocke l'ensemble de données complet. Par défaut, le noeud Distinguer trie les données dans les champs-clés puis sélectionne (ou supprime) le premier enregistrement de chaque groupe. Pour les ensembles de données d'une taille plus restreinte avec un faible nombre de clés distinctes ou ceux qui ont été prétriés, vous pouvez choisir des options pour améliorer la vitesse et l'efficacité du traitement.

Type. Dans certaines instances, le noeud Typer met en mémoire cache les données d'entrée lors de la lecture des valeurs ; le cache est utilisé pour les traitements en aval. Un cache requiert une quantité suffisante d'espace disque pour pouvoir stocker l'intégralité de l'ensemble de données, mais il accélère le traitement.

Evaluation. Le noeud Evaluation doit trier les données d'entrée pour pouvoir calculer les quantiles. L'opération de tri est répétée pour chaque modèle évalué car les scores et l'ordre des enregistrements qui en découle diffèrent dans chaque cas. Le temps d'exécution équivaut à $M*N*\log(N)$, où M correspond au nombre de modèles et N au nombre d'enregistrements.

Performances : Noeuds de modélisation

Réseau de neurones et Kohonen. Les algorithmes d'apprentissage du réseau de neurones (y compris l'algorithme de Kohonen) effectuent plusieurs passages sur les données d'apprentissage. Les données sont stockées en mémoire jusqu'à une certaine limite, et l'excédent de données est placé sur le disque. Il n'est pas judicieux d'accéder aux données d'apprentissage à partir du disque ; en effet, la méthode d'accès est aléatoire et peut donc entraîner une activité excessive sur le disque. Vous pouvez désactiver l'utilisation du stockage sur le disque pour ces algorithmes et forcer le stockage en mémoire de toutes les données ; pour ce faire, sélectionnez l'option d'optimisation de la vitesse dans l'onglet Modèle de la boîte de dialogue du noeud. Si la quantité de mémoire requise pour le stockage des données est supérieure à la taille de l'ensemble de travail du processus serveur, une partie sera paginée sur le disque et les performances s'en trouveront réduites.

Lorsque l'option d'optimisation de la mémoire est activée, un pourcentage de mémoire RAM physique est alloué à l'algorithme en fonction de la valeur affectée à l'option de configuration IBM® SPSS® Modeler Server relative à la *limite (en pourcentage) de la mémoire de modélisation*. Pour consacrer davantage de mémoire aux réseaux de neurones d'apprentissage, allouez davantage de mémoire RAM ou augmentez la valeur de cette option ; notez toutefois que si vous définissez une valeur trop élevée, une pagination sera réalisée.

Le temps d'exécution des algorithmes de réseau de neurones dépend du niveau de précision requis. Vous pouvez gérer le temps d'exécution en définissant une condition d'arrêt dans la boîte de dialogue du noeud.

K-Means. L'algorithme de classification non supervisée K-Means comporte les mêmes options de gestion de l'utilisation de la mémoire que les algorithmes de réseau de neurones. Les performances relatives aux données stockées sur le disque sont toutefois supérieures car l'accès aux données s'effectue de manière séquentielle.

Performances : Expressions CLEM

Les fonctions séquentielles CLEM (fonctions "@") qui se réfèrent à nouveau au flux de données doivent stocker suffisamment de données pour permettre la consultation en aval la plus longue possible. Pour les opérations dont le degré de consultation en aval n'est pas limité, toutes les valeurs du champ doivent être stockées. Une opération est dite illimitée si la valeur de décalage n'est pas un entier littéral ; par exemple, @OFFSET(Sales, Month). La valeur de décalage correspond au nom de champ *Mois* dont la valeur est inconnue avant exécution. Le serveur doit enregistrer toutes les valeurs du champ *Ventes* afin de garantir l'exactitude des résultats. Lorsqu'une limite supérieure est connue, fournissez-la en tant qu'argument supplémentaire ; par exemple, @OFFSET(Sales, Month, 12). Cette opération indique au serveur de ne stocker au maximum que les 12 valeurs les plus récentes de *Ventes*. Les fonctions séquentielles, limitées ou non, entravent quasi systématiquement la génération SQL.

Accessibilité dans IBM SPSS Modeler

Présentation de l'accessibilité dans IBM SPSS Modeler

Cette version offre une accessibilité améliorée à tous les utilisateurs, ainsi qu'une prise en charge spécifique pour les utilisateurs souffrant de troubles visuels ou fonctionnels. La présente section décrit les fonctions et méthodes de travail à l'aide des améliorations d'accessibilité, telles que les lecteurs d'écran et les raccourcis clavier.

Types de prise en charge pour l'accessibilité

Que vous souffriez de déficience visuelle ou que vous utilisiez systématiquement le clavier pour effectuer toute manipulation, il existe diverses méthodes d'utilisation de cet outil de Data mining. Par exemple, vous pouvez créer des flux, spécifier des options et lire des résultats sans jamais vous servir de la souris. Les raccourcis clavier disponibles sont répertoriés dans les rubriques suivantes. En outre, IBM® SPSS® Modeler fournit une prise en charge quasi complète des lecteurs d'écran, tels que JAWS pour Windows. Vous pouvez également optimiser le modèle de couleurs pour obtenir un contraste supplémentaire. Ces prises en charge sont abordées dans les rubriques suivantes.

Accessibilité pour les utilisateurs ayant une déficience visuelle

De nombreuses propriétés sont à votre disposition dans IBM® SPSS® Modeler pour vous permettre d'optimiser l'utilisation du logiciel.

Options d'affichage

Vous pouvez choisir les couleurs d'affichage des graphiques. Vous pouvez également utiliser vos propres paramètres Windows dans Clementine. Vous pourrez ainsi intensifier le contraste visuel.

- ▶ Pour définir les options d'affichage, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Affichage. Les options de cet onglet vous permettent de choisir le modèle de couleurs du logiciel, les couleurs des graphiques, ainsi que la taille de police des nœuds.

Utilisation de l'avertissement sonore

Vous pouvez contrôler le type d'avertissement utilisé pour certaines opérations dans le logiciel en activant ou désactivant les sons. Par exemple, vous pouvez activer les sons pour des événements, tels que la création ou la suppression d'un nœud, ou la génération de nouveaux résultats ou modèles.

- ▶ Pour définir les options de notification, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Notifications.

Contrôle du lancement automatique de nouvelles fenêtres

L'onglet Notifications de la boîte de dialogue Options utilisateur permet également de choisir si les nouvelles sorties, comme les tableaux et les graphiques, doivent être générées dans une fenêtre séparée. Vous pouvez désactiver ces options et ouvrir une fenêtre de résultats uniquement lorsque vous le souhaitez.

- ▶ Pour définir ces options, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Notifications.
- ▶ Dans la boîte de dialogue, sélectionnez Nouvelle sortie dans la liste du groupe Notifications visuelles.
- ▶ Sous Ouvrir une fenêtre, sélectionnez Jamais.

Taille des nœuds

La taille d'affichage des nœuds peut être standard ou petite. Vous pouvez choisir la taille en fonction de vos besoins.

- ▶ Pour définir les options de taille de nœud, cliquez sur Propriétés du flux dans le menu Fichier.
- ▶ Cliquez sur l'onglet Présentation.
- ▶ Dans la liste Taille des nœuds, sélectionnez Standard.

Accessibilité pour les utilisateurs aveugles

La prise en charge proposée aux personnes aveugles dépend essentiellement de l'utilisation d'un lecteur d'écran, tel que JAWS pour Windows. Pour optimiser l'utilisation d'un lecteur d'écran avec IBM® SPSS® Modeller, vous pouvez spécifier plusieurs paramètres.

Options d'affichage

De manière générale, les lecteurs d'écran fonctionnent mieux lorsque le contraste visuel est élevé. Si vos paramètres Windows offrent un contraste élevé, utilisez-les dans Clementine.

- ▶ Pour définir les options d'affichage, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Affichage.

Utilisation de l'avertissement sonore

Vous pouvez contrôler le type d'avertissement utilisé pour certaines opérations dans le logiciel en activant ou désactivant les sons. Par exemple, vous pouvez activer les sons pour des événements, tels que la création ou la suppression d'un nœud, ou la génération de nouveaux résultats ou modèles.

- ▶ Pour définir les options de notification, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Notifications.

Contrôle du lancement automatique de nouvelles fenêtres

L'onglet Notifications de la boîte de dialogue Options utilisateur permet également de choisir si les nouvelles sorties doivent être générées dans une fenêtre séparée. Vous pouvez désactiver cette option et ouvrir une fenêtre de résultats uniquement lorsque vous en avez besoin.

- ▶ Pour définir ces options, cliquez sur Options utilisateur dans le menu Outils.
- ▶ Cliquez sur l'onglet Notifications.
- ▶ Dans la boîte de dialogue, sélectionnez Nouvelle sortie dans la liste du groupe Notifications visuelles.
- ▶ Sous Ouvrir une fenêtre, sélectionnez Jamais.

Accessibilité via le clavier

La fonctionnalité du produit est accessible à partir du clavier. Vous pouvez appuyer simultanément sur la touche Alt + la touche appropriée pour activer les menus de la fenêtre (par exemple, Alt+F pour accéder au menu Fichier) ou sur la touche Tab pour passer d'une commande à l'autre dans une boîte de dialogue. Cependant, vous découvrirez ici des problèmes spécifiques liés à chaque fenêtre principale du produit, ainsi que des astuces utiles pour naviguer dans les boîtes de dialogue.

Cette section traite de l'utilisation du clavier pour effectuer des opérations, telles que l'ouverture d'un flux, l'utilisation des boîtes de dialogue des noeuds ou l'utilisation des résultats. En outre, des listes de raccourcis clavier sont fournies pour vous offrir une navigation encore plus efficace.

Raccourcis de navigation dans la fenêtre principale

Vous réalisez la plupart de votre travail de Data mining dans la fenêtre principale de IBM® SPSS® Modeler. La zone principale est appelée l'**espace de travail de flux**. Elle permet de créer et d'exécuter les flux de données. La partie inférieure de la fenêtre contient les **palettes de noeuds** qui répertorient tous les noeuds disponibles. Les palettes sont organisées en onglets correspondant au type d'opération de Data mining pour chaque groupe de noeuds. Par exemple, les noeuds permettant d'importer les données dans SPSS Modeler sont regroupés dans l'onglet Sources. Ceux permettant de calculer, de filtrer ou de classer les champs sont regroupés dans l'onglet Ops sur champs (abréviation d'Opérations sur les champs).

La partie droite de la fenêtre contient plusieurs outils de gestion des flux, des sorties et des projets. La partie supérieure à droite contient les **gestionnaires**. Vous y trouverez trois onglets permettant de gérer les flux, les sorties, ainsi que les modèles générés. Vous pouvez accéder à ces objets en sélectionnant l'onglet et l'objet voulu dans la liste. La partie inférieure à droite contient le **panneau du projet** qui vous permet d'organiser votre travail en projets. Cette zone contient deux onglets, qui correspondent à deux modes d'affichage d'un projet. Le **mode Classes** trie les objets du projet par type, alors que le **mode CRISP-DM** trie les objets en fonction de la phase de Data mining, telle que la préparation des données ou la modélisation. Ces différents aspects de la fenêtre de SPSS Modeler sont traités dans le système d'aide, ainsi que dans le guide de l'utilisateur.

Le tableau suivant contient les raccourcis utilisés pour vous déplacer dans la fenêtre principale de SPSS Modeler et pour créer des flux. Les raccourcis des boîtes de dialogue et des sorties sont répertoriés dans les rubriques suivantes. Ces touches de raccourci sont disponibles uniquement dans la fenêtre principale.

Raccourcis de la fenêtre principale

Touche de raccourci	Fonction
Ctrl+F5	Active la palette de noeuds.
Ctrl+F6	Active l'espace de travail de flux.
Ctrl+F7	Active le panneau des gestionnaires.
Ctrl+F8	Active le panneau des projets.

Raccourcis des noeuds et des flux

Touche de raccourci	Fonction
Ctrl+N	Crée un espace de travail de flux vide.
Ctrl+O	Affiche la boîte de dialogue Ouvrir dans laquelle vous pouvez sélectionner et ouvrir un flux existant.
Ctrl+touches numériques	Permet d'activer l'onglet correspondant d'une fenêtre ou d'un panneau. Par exemple, dans une fenêtre à onglet, Ctrl+1 affiche le premier onglet à partir de la gauche, Ctrl+2 le deuxième, etc.
Ctrl+Flèche bas	S'utilise dans la palette de noeuds pour passer d'un onglet de la palette au premier noeud situé sous cet onglet.
Ctrl+Flèche haut	S'utilise dans la palette de noeuds pour se déplacer d'un noeud à l'onglet de sa palette.
Enter	Lorsqu'un noeud est sélectionné dans la palette de noeuds (y compris les modèles affinés de la palette de modèles générés), cette touche permet d'ajouter le noeud à l'espace de travail de flux. Appuyer sur Entrée lorsqu'un noeud est sélectionné dans l'espace de travail ouvre la boîte de dialogue de ce noeud.
Ctrl+Entrée	Lorsqu'un noeud est sélectionné dans la palette, ce raccourci permet d'ajouter le noeud à l'espace de travail de flux sans avoir à le sélectionner et d'activer le premier noeud de la palette.
Alt+Entrée	Lorsqu'un noeud est sélectionné dans la palette, ce raccourci permet d'ajouter le noeud à l'espace de travail de flux et de le sélectionner tout en activant le premier noeud de la palette.
Maj.+Barre d'espace	Lorsqu'un noeud ou un commentaire est activé dans la palette, ce raccourci permet de basculer entre la sélection ou la désélection de ce noeud ou de ce commentaire. Si d'autres noeuds ou commentaires sont également sélectionnés, cela a pour effet de les désélectionner.
Ctrl+Maj.+Barre d'espace	Lorsqu'un noeud ou un commentaire est activé dans le flux, ou qu'un noeud ou un commentaire est activé dans la palette, ce raccourci permet de basculer entre la sélection ou la désélection de ce noeud ou de ce commentaire. Ceci n'a aucune influence sur les autres noeuds ou commentaires sélectionnés.
Flèche gauche/droite	Si l'espace de travail de flux est sélectionné, déplace horizontalement le flux entier sur l'écran. Si l'onglet d'une palette est sélectionné, permet de se déplacer entre les onglets. Si un noeud de palette est activé, permet de se déplacer entre les noeuds de cette palette.
Flèche haut/bas	Si l'espace de travail de flux est sélectionné, déplace verticalement le flux entier sur l'écran. Si un noeud de palette est activé, permet de se déplacer entre les noeuds de cette palette. Si une sous-palette est activée, permet de se déplacer entre les autres sous-palettes de l'onglet de cette palette.

Touche de raccourci	Fonction
Alt+Flèche gauche/droite	Permet de déplacer horizontalement les noeuds et les commentaires sélectionnés dans l'espace de travail de flux dans le sens indiqué par la flèche.
Alt+Flèche haut/bas	Permet de déplacer verticalement les noeuds et les commentaires sélectionnés dans l'espace de travail de flux dans le sens indiqué par la flèche.
Ctrl+A	Sélectionne tous les noeuds d'un flux.
Ctrl+Q	Lorsqu'un noeud est activé, permet de le sélectionner, ainsi que tous les noeuds en aval et désélectionne tous les noeuds en amont.
Ctrl+W	Lorsqu'un noeud est activé, permet de le désélectionner ainsi que tous les noeuds sélectionnés en aval.
Ctrl+Alt+D	Duplique un noeud sélectionné.
Ctrl+Alt+L	Lorsqu'un nugget de modèle est sélectionné dans le flux, ouvre une boîte de dialogue Insérer pour vous permettre de charger un modèle enregistré d'un fichier .nod dans le flux.
Ctrl+Alt+R	Affiche l'onglet Annotations d'un noeud sélectionné et vous permet de renommer ce noeud.
Ctrl+Alt+U	Crée un noeud source Utilisateur.
Ctrl+Alt+C	Permet d'activer ou de désactiver le cache d'un noeud.
Ctrl+Alt+F	Vide le cache d'un noeud
Onglet	Dans l'espace de travail de flux, permet de se déplacer entre tous les noeuds source et les commentaires dans le flux actuel. Dans une sous-palette sélectionnée, bascule l'activation entre l'étiquette et le premier noeud de la sous-palette.
Maj+Tab	Effectue la même opération que la touche Tab mais dans l'ordre inverse.
Ctrl+Tab	Lorsque l'activation est sur l'espace de travail du flux, le panneau des gestionnaires ou celui des projets, déplace l'activation vers la palette de noeuds.
Une touche alphabétique quelconque	Dans le flux en cours, lorsque l'activation est sur un noeud, permet d'activer le noeud suivant dont le nom commence par la lettre de la touche sur laquelle vous avez appuyée.
F1	Ouvre le système d'Aide à une rubrique relative à l'élément activé.
F2	Permet de lancer le processus de connexion du noeud sélectionné dans l'espace de travail. Servez-vous de la touche Tab pour passer au noeud voulu dans l'espace de travail, puis appuyez sur Maj+barre d'espacement pour terminer la connexion.
F3	Permet de supprimer toutes les connexions du noeud sélectionné dans l'espace de travail.
F6	Déplace l'activation entre le panneau des gestionnaires, le panneau des projets et les palettes des noeuds.
F10	Permet d'ouvrir le menu Fichier.
Maj+F10	Ouvre le menu contextuel pour le noeud ou le flux.
Suppr	Supprime un noeud sélectionné de l'espace de travail.
Echap	Ferme le menu contextuel ou la boîte de dialogue.
Ctrl+Alt+X	Permet de développer un super noeud.
Ctrl+Alt+Z	Effectue un zoom avant dans un super noeud.
Ctrl+Alt+Maj+Z	Effectue un zoom arrière dans un super noeud.
Ctrl+E	Lorsque l'élément activé se trouve dans l'espace de travail de flux, permet d'exécuter le flux en cours.

Il existe également certaines touches de raccourci standard dans SPSS Modeler, comme Ctrl+C pour copier un élément. [Pour plus d'informations, reportez-vous à la section Utilisation de touches de raccourci dans le chapitre 3 sur p. 22.](#)

Raccourcis des boîtes de dialogue et des tableaux

Il existe plusieurs raccourcis et touches de lecteur d'écran très pratiques à utiliser dans les boîtes de dialogue et les tableaux, ainsi que dans les tableaux situés dans les boîtes de dialogue. La liste exhaustive des raccourcis clavier et de lecteur d'écran spéciaux est fournie ci-dessous.

Raccourcis des boîtes de dialogue et du Générateur de formules

Touche de raccourci	Fonction
Alt+4	Permet de fermer toutes les boîtes de dialogue ou fenêtres de sortie actives. Les sorties peuvent être récupérées dans l'onglet Sorties du panneau des questionnaires.
Ctrl+Fin	Lorsqu'une commande quelconque du Générateur de formules est activée, cette combinaison déplace le point d'insertion à la fin de l'expression.
Ctrl+1	Dans le Générateur de formules, permet d'activer la commande d'édition de l'expression.
Ctrl+2	Dans le Générateur de formules, permet d'activer la liste des fonctions.
Ctrl+3	Dans le Générateur de formules, permet d'activer la liste des champs.

Raccourcis de tableaux

Les raccourcis de tableaux sont utilisés pour les tableaux de sortie, mais également pour les commandes de tableau dans les boîtes de dialogue de noeuds tels que Typer, Filtrer et Fusionner. Vous utilisez généralement la touche Tab pour vous déplacer d'une cellule du tableau à l'autre et les touches Ctrl+Tab pour quitter la commande de tableau. *Remarque* : certains lecteurs d'écran peuvent parfois ne pas commencer immédiatement la lecture du contenu d'une cellule. Appuyez une fois ou deux sur les touches de direction pour réinitialiser le logiciel et commencer la lecture.

Touche de raccourci	Fonction
Ctrl+W	Dans les tableaux, permet de lire la brève description de la ligne sélectionnée. Par exemple, « Les valeurs de la ligne 2 sélectionnée sont sexe, booléen, m/f, etc. »
Ctrl+Alt+W	Dans les tableaux, permet de lire la longue description de la ligne sélectionnée. Par exemple, « Les valeurs de la ligne 2 sélectionnée sont champs = sexe, type = booléen, sexe = m/f, etc. »
Ctrl+D	Dans les tableaux, permet de lire la brève description de la zone sélectionnée. Par exemple, « Sélection : une ligne pour six colonnes ».
Ctrl+Alt+D	Dans les tableaux, fournit la longue description de la zone sélectionnée. Par exemple, « Sélection : une ligne pour six colonnes. Les colonnes sélectionnées sont Champs, Type, Manquant. La ligne 1 est sélectionnée. »

Touche de raccourci	Fonction
Ctrl+T	Dans les tableaux, permet de lire la brève description des colonnes sélectionnées. Par exemple, « Champs, Type, Manquant ».
Ctrl+Alt+T	Dans les tableaux, permet de lire la longue description des colonnes sélectionnées. Par exemple, « Les colonnes sélectionnées sont Champs, Type, Manquant ».
Ctrl+R	Indique le nombre d' enregistrements contenus dans le tableau.
Ctrl+Alt+R	Indique le nombre d' enregistrements contenus dans le tableau, ainsi que le nom des colonnes.
Ctrl+I	Dans les tableaux, lit les informations ou le contenu de la cellule activée.
Ctrl+Alt+I	Dans les tableaux, lit la longue description d' informations (nom de la colonne et contenu de la cellule) sur la cellule active.
Ctrl+G	Dans les tableaux, fournit de brèves informations générales sur la sélection.
Ctrl+Alt+G	Dans les tableaux, fournit de longues informations générales sur la sélection.
Ctrl+Q	Dans les tableaux, permet de passer rapidement d'une cellule à l'autre. Ctrl+Q permet de lire les descriptions longues, comme « Sexe=Féminin », lorsque vous vous déplacez dans le tableau à l'aide des touches de direction. Si vous sélectionnez de nouveau Ctrl+Q, vous passez aux descriptions courtes (contenu de la cellule).

Raccourcis pour les commentaires

Lorsque vous utilisez des commentaires à l'écran, vous pouvez utiliser les raccourcis suivants.

Touche de raccourci	Fonction
Alt+C	Permet d'afficher ou de masquer les commentaires.
Alt+M	Insère un nouveau commentaire si des commentaires sont actuellement affichés ; affiche des commentaires s'ils sont actuellement masqués.
Onglet	Dans l'espace de travail de flux, permet de se déplacer entre tous les noeuds source et les commentaires dans le flux actuel.
Enter	Lorsqu'un commentaire est activé, indique le début de la modification.
Alt+Entrée ou Ctrl+Tab	Termine la modification et enregistre les modifications.
Echap	Annule les modifications. Les modifications effectuées au cours de l'édition sont perdues.
Alt+Maj+Flèche haut	Réduit la hauteur de la zone de texte d'une cellule de grille (ou d'un pixel) si l'ajustement à la grille est activé (ou désactivé).
Alt+Maj+Flèche bas	Augmente la hauteur de la zone de texte d'une cellule de grille (ou d'un pixel) si l'ajustement à la grille est activé (ou désactivé).
Alt+Maj+Flèche gauche	Réduit la largeur de la zone de texte d'une cellule de grille (ou d'un pixel) si l'ajustement à la grille est activé (ou désactivé).
Alt+Maj+Flèche droite	Augmente la largeur de la zone de texte d'une cellule de grille (ou d'un pixel) si l'ajustement à la grille est activé (ou désactivé).

Raccourcis pour le visualiseur de classes et le Visualiseur de modèles

Les touches de raccourcis sont disponibles pour la navigation dans les fenêtres du visualiseur de classes et du Visualiseur de modèles.

Général - Visualiseur de classes et Visualiseur de modèles

Touche de raccourci	Fonction
Onglet	Active la commande de l'écran suivant.
Maj+Tab	Active la commande de l'écran précédent.
Flèche Bas	Si une liste déroulante est activée, ouvre la liste ou passe à l'élément suivant de la liste. Si un menu est activé, passe à l'élément suivant du menu. Si un graphique en miniature est activé, déplace l'activation au graphique suivant dans l'ensemble (ou au premier graphique de l'ensemble si le dernier graphique est activé).
Flèche Haut	Si une liste déroulante est ouverte, passe à l'élément précédent de la liste. Si un menu est activé, passe à l'élément précédent du menu. Si un graphique en miniature est activé, déplace l'activation au graphique précédent dans l'ensemble (ou au dernier graphique de l'ensemble si le premier graphique miniature est activé).
Enter	Ferme une liste déroulante ou exécute une sélection dans un menu ouvert.
F6	Bascule l'activation entre les panneaux de gauche et de droite de la fenêtre.
Flèches de gauche et de droite	Si un onglet est activé, passe à l'onglet précédent ou suivant. Si un menu est activé, passe au menu précédent ou suivant.
Alt+lettre	Sélectionne le bouton ou le menu dont cette lettre est soulignée dans le nom.
Echap	Ferme une liste déroulante ou un menu ouvert.

Visualiseur de classes uniquement

Le visualiseur de classes possède une vue de classes qui contient une grille de classes par caractéristiques. [Pour plus d'informations, reportez-vous à la section Vue des classes dans le chapitre 11 dans *Noeuds de modélisation IBM SPSS Modeler 14.2*.](#)

Pour sélectionner la vue des classes au lieu de la vue récapitulative du modèle :

- ▶ Appuyez à plusieurs sur la touche Tab jusqu'à ce que le bouton Vue soit sélectionné.
- ▶ Appuyez deux fois sur la flèche bas pour sélectionner Classes.

À ce stade, vous pouvez sélectionner une cellule individuelle dans la grille :

- ▶ Appuyez à plusieurs reprises sur la touche Tab jusqu'à ce que vous atteigniez la dernière icône de la barre d'outils de visualisation.

Figure A-1

Icône Afficher l'arbre de visualisation



- Appuyez une fois de plus sur la touche Tab, puis sur la barre d'espace et sur une touche fléchée.

Les raccourcis clavier suivants sont désormais disponibles :

Touche de raccourci	Fonction
Touche fléchée	Permet de passer d'une cellule individuelle à l'autre dans la grille. L'affichage Distribution des cellules dans le panneau de droite change avec l'activation.
Ctrl+, (virgule)	Sélectionne ou désélectionne la totalité de la colonne de la grille dont une cellule est activée. Pour ajouter une colonne à la sélection, utilisez les touches fléchées pour accéder à une cellule de cette colonne et appuyez de nouveau sur Ctrl+.
Onglet	Désactive la grille et active la commande de l'écran suivant.
Maj+Tab	Désactive la grille et revient à la commande de l'écran précédent.
F2	Passé en mode d'édition (cellules étiquette et description uniquement).
Enter	Enregistre les modifications d'édition et quitte le mode d'édition (cellules étiquette et description uniquement).
Echap	Quitte le mode d'édition sans enregistrer les modifications (cellules étiquette et description uniquement).

Exemple de touches de raccourci : Création de flux

Afin de clarifier le processus de création de flux pour les utilisateurs qui utilisent le clavier ou un lecteur d'écran, suivez l'exemple de création de flux sans souris ci-dessous. Dans cet exemple, vous allez créer un flux contenant un nœud Délimité, un nœud Calculer et un nœud Histogramme en procédant comme suit :

- **Démarrez SPSS Modeler.** Lorsque IBM® SPSS® Modeler démarre pour la première fois, l'onglet Favoris de la palette de nœuds est activé.
- **Ctrl+Flèche bas.** Lorsque l'onglet est activé, permet d'activer son contenu.
- **Flèche droite.** Permet d'activer le nœud Délimité.
- **Barre d'espace.** Permet de sélectionner le nœud Délimité.
- **Ctrl+Entrée.** Ajoute le nœud Délimité dans l'espace de travail du flux. Cette combinaison de touches permet également de conserver l'activation du nœud Délimité afin d'établir une connexion avec le nœud ajouté suivant.
- **Tabulation.** Active à nouveau la palette de nœuds.
- **Flèche droite (4 fois).** Permet d'accéder au nœud Calculer.
- **Barre d'espace.** Permet de sélectionner le nœud Calculer.
- **Alt+Entrée.** Permet d'ajouter le nœud Calculer à l'espace de travail et d'activer le nœud Calculer. Ce nœud est maintenant prêt à être connecté au nœud ajouté suivant.
- **Tabulation.** Active à nouveau la palette de nœuds.

- ▶ **Flèche droite (5 fois).** Permet d'activer le noeud Histogramme dans la palette.
- ▶ **Barre d'espace.** Permet de sélectionner le noeud Histogramme.
- ▶ **Entrée.** Permet d'ajouter le noeud au flux et d'activer l'espace de travail de flux.

Passez à l'exemple suivant ou enregistrez le flux si vous souhaitez essayer ultérieurement l'exemple suivant.

Exemple de touches de raccourci : Edition de noeuds

Dans cet exemple, vous utiliserez le flux créé dans l'exemple précédent. Le flux se compose d'un noeud Délimité, d'un noeud Calculer et d'un noeud Histogramme. Les instructions commencent avec l'activation du troisième noeud du flux, le noeud Histogramme.

- ▶ **Ctrl+Flèche gauche deux fois.** Permet d'activer à nouveau le noeud Délimité.
- ▶ **Entrée.** Permet d'ouvrir la boîte de dialogue Délimité. À l'aide de la touche Tab, accédez au champ Fichier et saisissez le chemin et le nom du fichier pour sélectionner ce fichier. Appuyez sur Ctrl+Tab pour accéder à la partie inférieure de la boîte de dialogue, puis au bouton OK et appuyez sur Entrée pour fermer la boîte de dialogue.
- ▶ **Ctrl+Flèche droite.** Permet d'activer le deuxième noeud, le noeud Calculer.
- ▶ **Entrée.** Permet d'ouvrir la boîte de dialogue du noeud Calculer. A l'aide de la touche Tab, sélectionnez les champs et spécifiez les conditions de calcul. Appuyez sur Ctrl+Tab pour accéder au bouton OK et appuyez sur Entrée pour fermer la boîte de dialogue.
- ▶ **Ctrl+Flèche droite.** Permet d'activer le troisième noeud, le noeud Histogramme.
- ▶ **Entrée.** Permet d'ouvrir la boîte de dialogue du noeud Histogramme. A l'aide de la touche Tab, sélectionnez les champs et spécifiez les options de graphique. Pour les listes déroulantes, appuyez sur la flèche bas pour ouvrir la liste et pour mettre en surbrillance un élément de la liste, puis appuyez sur Entrée pour sélectionner l'élément de la liste. A l'aide de la touche Tab, accédez au bouton OK et appuyez sur Entrée pour fermer la boîte de dialogue.

Vous pouvez maintenant ajouter d'autres noeuds ou exécuter le flux actuel. Gardez à l'esprit les conseils suivants lorsque vous créez des flux :

- Lorsque vous reliez manuellement des noeuds, utilisez F2 pour créer le point de départ d'une connexion, appuyez sur la touche Tab pour accéder au point de fin, puis utilisez Maj+Espace pour finaliser la connexion.
- Utilisez la touche F3 pour supprimer toutes les connexions du noeud sélectionné dans l'espace de travail.
- Lorsque le flux est créé, utilisez Ctrl+E pour exécuter le flux en cours.

La liste complète des touches de raccourci est disponible. [Pour plus d'informations, reportez-vous à la section Raccourcis de navigation dans la fenêtre principale sur p. 257.](#)

Utilisation d'un lecteur d'écran

Plusieurs lecteurs d'écran sont disponibles sur le marché. IBM® SPSS® Modeler est configuré pour prendre en charge JAWS pour Windows en utilisant le Java Access Bridge installé avec SPSS Modeler. Si JAWS est installé, démarrez-le simplement avant de lancer SPSS Modeler pour utiliser ce produit.

Compte tenu de la nature de la représentation graphique unique du processus de Data mining de SPSS Modeler, une analyse visuelle des diagrammes et des graphiques est la solution conseillée. Il est toutefois possible de comprendre les éléments et de prendre des décisions en fonction des résultats et modèles affichés textuellement avec un lecteur d'écran.

Remarque : Avec des machines clientes 64 bits, certaines caractéristiques technologiques d'assistance ne fonctionnent pas. Ceci est dû au fait que Java Access Bridge n'est pas conçu pour une utilisation en 64 bits.

Utilisation du fichier de dictionnaire dans IBM SPSS Modeler

Un fichier de dictionnaire de SPSS Modeler (*Awt.JDF*) peut être inclus dans JAWS. Pour utiliser ce fichier :

- ▶ Accédez au sous-répertoire */accessibility* de votre installation SPSS Modeler et copiez le fichier de dictionnaire (*Awt.JDF*).
- ▶ Copiez-le dans le répertoire où se trouvent vos scripts JAWS.

Un fichier appelé *Awt.JDF* existe peut-être déjà sur votre ordinateur si d'autres applications JAVA sont en cours d'exécution. Dans ce cas, vous aurez peut-être à éditer manuellement le fichier de dictionnaire pour pouvoir l'utiliser.

Utilisation d'un lecteur d'écran avec une sortie au format HTML

Lorsque vous consultez un résultat affiché au format HTML dans IBM® SPSS® Modeler à l'aide d'un lecteur d'écran, vous risquez de rencontrer des problèmes. Plusieurs types de sortie sont concernés, notamment :

- Sortie affichée dans l'onglet Options avancées pour les noeuds Régression, Régression logistique et Analyse factorielle/ACP
- Sortie du noeud Rapport

Dans chacune de ces fenêtres ou boîtes de dialogue, la barre d'outils contient un outil qui permet de lancer la sortie dans votre navigateur par défaut et qui fournit la prise en charge de lecteur d'écran standard. Vous pouvez ensuite utiliser le lecteur d'écran pour consulter les informations de sortie.

Accessibilité dans la fenêtre Arbre interactif

L'affichage standard des modèles d'arbre décision dans la fenêtre Arbre interactif peut poser problème aux lecteurs d'écran. Pour ouvrir une version accessible, dans les menus de l'arbre interactif, cliquez sur :

Affichage > Fenêtre d'accessibilité

Très semblable à la carte d'arbre standard, l'affichage obtenu peut être lu par JAWS. Vous pouvez vous déplacer vers le haut, le bas, la droite ou la gauche à l'aide des touches de direction standard. Au fur et à mesure que vous parcourez la fenêtre accessible, l'activation dans la fenêtre Arbre interactif change. Utilisez la barre d'espace pour modifier la sélection, ou les touches Ctrl+Espace pour agrandir la sélection actuelle.

Conseils d'utilisation

Il existe plusieurs astuces pour faciliter l'accès à l'environnement IBM® SPSS® Modeler. Voici des conseils généraux pour travailler dans SPSS Modeler.

- **Fermeture des zones de texte étendues.** Utilisez Ctrl+Tab pour quitter les zones de texte étendues. *Remarque* : Ctrl+Tab permet également de quitter les commandes de tableaux.
- **Utilisation de la touche Tab à la place des touches de direction.** Pour sélectionner des options dans une boîte de dialogue, utilisez la touche Tab pour accéder à ces options. Les touches de direction ne fonctionnent pas dans ce contexte.
- **Listes déroulantes.** Dans les listes déroulantes des boîtes de dialogue, vous pouvez utiliser la touche Echap ou la barre d'espace pour sélectionner un élément, puis fermer la liste. Vous pouvez également utiliser la touche Echap pour fermer une liste déroulante qui est restée active alors que vous avez accédé à une autre commande.
- **Etat de l'exécution.** Lorsque vous exécutez un flux dans une base de données volumineuse, JAWS risque de lire l'état du flux avec un peu de retard. Appuyez régulièrement sur la touche Ctrl pour actualiser le rapport sur l'état.
- **Utilisation des palettes de noeuds.** Lorsque vous accédez pour la première fois à un onglet des palettes de noeuds, JAWS lit parfois « groupbox » (zone de groupe) au lieu du nom du noeud. Dans ce cas, utilisez Ctrl+Flèche droite, puis Ctrl+Flèche gauche pour réinitialiser le lecteur d'écran et écouter le nom du noeud.
- **Lecture des menus.** Lors de la première ouverture d'un menu, il arrive que JAWS ne puisse pas lire le premier élément de menu. Si vous pensez que cette situation s'est produite, utilisez la flèche vers le bas, puis la flèche vers le haut pour entendre le nom du premier élément du menu.
- **Menus en cascade.** JAWS ne lit pas le premier niveau d'un menu en cascade. Si vous entendez une rupture dans l'énoncé pendant que vous vous déplacez dans un menu, appuyez sur la flèche droite pour entendre le nom des éléments du menu enfant.

En outre, si vous avez installé IBM® SPSS® Text Analytics, les conseils suivants peuvent vous rendre l'interface utilitaire interactive plus accessible.

- **Entrée dans une boîte de dialogue.** Vous devrez peut-être appuyer sur la touche Tab pour activer la première commande lorsque vous accédez à une boîte de dialogue.
- **Fermeture des zones de texte étendues.** Utilisez Ctrl+Tab pour fermer les zones de texte étendues et accéder à la commande suivante. *Remarque* : Ctrl+Tab permet également de quitter les commandes de tableaux.

- **Saisie de la première lettre pour trouver un élément dans la liste de l'arbre.** Lorsque vous recherchez un élément dans les panneaux des catégories ou des résultats extraits, ou dans l'arbre de bibliothèque, vous pouvez taper la première lettre de l'élément si le panneau est actif. Cela permet de sélectionner l'élément suivant commençant par la lettre saisie.
- **Listes déroulantes.** Dans les listes déroulantes des boîtes de dialogue, vous pouvez utiliser la barre d'espace pour sélectionner un élément, puis fermer la liste.

D'autres astuces utiles sont indiquées dans les rubriques suivantes.

Interférence avec d'autres logiciels

Lors des tests de IBM® SPSS® Modeler avec des lecteurs d'écran comme JAWS, notre équipe de développement a découvert que l'emploi de SMS (Systems Management Server) dans votre entreprise risquait de provoquer des interférences lorsque JAWS lit des applications Java, telles que SPSS Modeler. Ce problème est résolu si vous désactivez SMS. Pour plus d'informations sur le SMS, consultez le site Web de Microsoft.

JAWS et Java

Les différentes versions de JAWS offrent divers niveaux de prise en charge des applications logicielles Java. Bien que IBM® SPSS® Modeler fonctionne avec toutes les versions récentes de JAWS, certaines versions risquent tout de même de rencontrer de légers problèmes lorsqu'elles sont utilisées avec des systèmes Java. Visitez le site Web de JAWS pour Windows à l'adresse <http://www.FreedomScientific.com>.

Utilisation des graphiques dans IBM SPSS Modeler

Il est difficile pour un lecteur d'écran d'interpréter l'affichage visuel des informations, comme les histogrammes, les graphiques d'évaluation, les courbes et les diagrammes de dispersion. Remarquez cependant que les graphiques Relations et les proportions peuvent être affichés en utilisant un récapitulatif textuel disponible dans la fenêtre de sortie.

Prise en charge d'Unicode

Prise en charge d'Unicode dans IBM SPSS Modeler

IBM® SPSS® Modeler prend totalement en charge Unicode pour IBM® SPSS® Modeler et IBM® SPSS® Modeler Server. Ainsi, il est possible d'échanger des données avec d'autres applications Unicode, y compris les bases de données multilingues, et ce, sans perdre d'informations suite à une conversion vers ou depuis un code propre aux paramètres régionaux.

- SPSS Modeler stocke les données Unicode en interne, et peut lire et créer les données multilingues stockées au format Unicode dans les bases de données, sans aucune perte.
- SPSS Modeler peut lire et écrire les fichiers texte ayant fait l'objet d'un codage UTF-8. Les fonctions d'importation/exportation de fichiers texte utilisent par défaut le code régional, mais prennent en charge UTF-8 en guise d'alternative. Ce paramètre peut être spécifié dans les noeuds d'importation/exportation de fichiers. Le codage par défaut peut également être modifié dans la boîte de dialogue Propriétés du flux. [Pour plus d'informations, reportez-vous à la section Définition des options de flux dans le chapitre 5 sur p. 51.](#)
- Les fichiers de données Statistics, SAS et de texte stockés dans le code régional sont convertis au format UTF-8 au cours de l'importation, puis une nouvelle fois durant l'exportation. Lors de l'écriture d'un fichier, si certains caractères Unicode ne figurent pas dans le jeu de caractères régional, ils sont remplacés et un message d'avertissement apparaît. Cela ne doit se produire que si les données sont importées depuis une source de données qui prend en charge Unicode (une base de données ou un fichier texte UTF-8), et qui inclut les caractères d'autres paramètres régionaux ou jeux de caractères.
- Les images IBM® SPSS® Modeler Solution Publisher font l'objet d'un codage UTF-8 et peuvent être littéralement transférées entre plates-formes, quels que soient les paramètres régionaux.

A propos d'Unicode

La norme Unicode a pour objectif d'assurer le codage cohérent d'un texte multilingue pour qu'il puisse être partagé facilement entre les frontières, quels que soient les paramètres régionaux et les applications. La norme Unicode, dont la version actuelle est 4.0.1, définit un jeu de caractères constituant un sur-ensemble de tous les jeux de caractères utilisés aujourd'hui à l'échelle internationale, et attribue à chaque caractère un nom et un code caractère uniques. Les caractères et les codes caractère correspondants sont identiques à ceux du jeu universel de caractères défini par la norme ISO-10646. Pour plus d'informations, consultez la [page d'accueil d'Unicode \(http://www.unicode.org\)](http://www.unicode.org).

Avis

This information was developed for products and services offered worldwide.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

Le paragraphe suivant ne s'applique pas au Royaume Uni ni à aucun pays dont la réglementation locale est incompatible avec les dispositions suivantes : SPSS INC., AN IBM COMPANY, FOURNIT CETTE PUBLICATION « EN L'ÉTAT » SANS GARANTIE D'AUCUNE SORTE, EXPLICITE OU IMPLICITE, Y COMPRIS, MAIS SANS SE LIMITER À, LES GARANTIES TACITES DE NON CONTREFAÇON, DE CONFORMITÉ OU D'APTITUDE À UN USAGE PARTICULIER. Certains états n'autorisent pas les clauses de non garantie explicites ou implicites lors de certaines transactions, par conséquent, il est possible que cette disposition ne vous concerne pas.

Ces informations peuvent contenir des erreurs techniques ou typographiques. Ces informations sont modifiées périodiquement ; ces modifications seront incluses dans les nouvelles versions de cette publication. SPSS Inc. a le droit d'apporter des améliorations et/ou des modifications au(x) produit(s) et/ou au(x) programme(s) décrits dans cette publication à tout moment et sans avertissement préalable.

Toute référence, contenue dans ces informations, à des sites Internet autres que ceux de SPSS et IBM est fournie pour des raisons de commodité uniquement et ne peut, en aucune sorte, être considérée comme de la publicité pour ces sites Internet. Le matériel de ces sites Internet ne fait pas partie du matériel de ce produit SPSS Inc. et l'utilisation de ces sites Internet se fait à vos risques et périls.

Lorsque vous envoyez des informations à IBM ou SPSS, vous octroyez à IBM et SPSS le droit non exclusif d'utiliser ou de distribuer ces informations de la façon dont IBM ou SPSS le juge approprié sans aucune obligation envers vous.

Les informations sur des produits autres que SPSS ont été obtenues auprès des fournisseurs de ces produits, à partir de leurs annonces publiées ou d'autres sources publiques disponibles. SPSS n'a pas effectué de test sur ces produits et ne peut confirmer la validité de leurs performances, de leur compatibilité ou de toute autre prétention associée aux produits autres que SPSS. Les questions concernant les performances des produits autres que SPSS devront être adressées aux fournisseurs de ces produits.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Ces informations contiennent des exemples de données et de rapports utilisées dans les opérations quotidiennes de l'entreprise. Pour les illustrer le plus précisément possible, ces exemples contiennent les noms de personnes, d'entreprises, de marques et de produits. Tous ces noms sont fictifs et toute ressemblance avec des noms et adresses utilisés par une entreprise réelle ne serait que pure coïncidence.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Marques commerciales

IBM, le logo IBM et [ibm.com](http://www.ibm.com) sont des marques commerciales de IBM Corporation, déposées dans de nombreuses juridictions à travers le monde. Une liste mise à jour des marques commerciales IBM est disponible sur Internet à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

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

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques déposées ou des marques commerciales de Adobe Systems Incorporated aux États-Unis et/ou dans d'autres pays.

IT Infrastructure Library est une marque déposée du Central Computer and Telecommunications Agency qui fait désormais partie du Office of Government Commerce.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium et Pentium sont des marques commerciales ou déposées de Intel Corporation ou de ses filiales aux États-Unis ou dans d'autres pays.

Linux est une marque déposée de Linus Torvalds aux États-Unis, dans d'autres pays, ou les deux.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux États-Unis, dans d'autres pays, ou les deux.

ITIL est une marque déposée et une marque communautaire déposée du Office of Government Commerce et a été enregistrée auprès du Bureau américain des brevets et des marques de commerce.

UNIX est une marque déposée de The Open Group aux États-Unis et dans d'autres pays.

Cell Broadband Engine est une marque de Sony Computer Entertainment, Inc. aux États-Unis, dans d'autres pays ou les deux et est par conséquent sous contrat de licence.

Java et les marques et logo Java sont des marques de Sun Microsystems, Inc. aux États-Unis, dans d'autres pays, ou les deux.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

D'autres noms de produits ou de services peuvent être des marques de IBM, SPSS ou d'autres entreprises.

Index

- accessibilité, 255, 267
 - astuces dans IBM SPSS Modeler, 266
 - exemple, 263–264
 - fonctionnalités dans IBM SPSS Modeler, 255
- activer des noeuds , 43
- ajout
 - à un projet, 216
- ajout de connexions à IBM SPSS Modeler Server, 11–12
- Analyse champion/challenger, 161, 188
- analyse en arbre
 - applications standard, 26
- annotation
 - flux, 70, 78
 - noeuds, 70, 78
- annotations
 - dossier, 222
 - projet, 221
- Annotations
 - conversion en commentaires, 77
- annuler, 19
- aperçu
 - données de noeud, 50
- applications, 26
- applications de Data mining, 27
- applications prédictives, 188
- applications standard, 26
- apprentissage automatique, 25
- arbres de décision
 - accessibilité, 265
- arrêter l'exécution, 19
- Assistant des applications prédictives, 162, 201, 204
 - conditions préalables au flux, 202
- astuces
 - pour l'accessibilité, 266
 - utilisation générale, 89
- attribute, 25
- attribution d'un nom à des noeuds et à des flux, 78
- Automatisation, 99
- avertissements, 56
 - définition des options, 232
- avis juridiques, 269

- barre d'outils, 19
- blanc
 - suppression dans les chaînes, 107, 139
- blancs, 93–94, 108
- boîte de dialogue d'accueil, 236
- boîte de dialogue de démarrage, 236
- bouton central de la souris
 - simulation, 22, 40
- branches, modélisation et scoring, 70, 192–193, 198

- cache
 - activation, 47, 231, 250
 - définition d'un cache, 45
 - enregistrement, 49
 - options des noeuds, 47, 250
 - vidage, 49, 51
- caractère barre oblique inversée dans les expressions CLEM, 124
- caractères, 122, 124
- caractères spéciaux
 - suppression dans les chaînes, 107
- chaînes, 122, 124
 - manipulation dans les expressions CLEM, 107
 - mise en correspondance, 107
 - remplacement, 107
- champs, 25, 122, 124
 - affichage des valeurs, 116
 - dans les expressions CLEM, 116
- champs du modèle, 88
- champs essentiels, 84, 87
- champs non mappés, 84
- champs obligatoires, 88
- champs Temps
 - conversion , 148
- chargement
 - états, 83
 - noeuds, 83
- classes , 18, 214, 216
- CLEM, 113
 - création d'expressions, 114
 - Exemples, 102
 - expressions, 105, 122
 - fonctions, 115
 - Introduction, 24, 99
 - Langage, 122
 - types de données, 123–124
 - vérification des expressions, 118
- CLEM expressions
 - paramètres, 58, 106
- CLEM fonctions
 - aléatoire, 139
 - chaîne, 139
 - comparaison, 132
 - fonctions spéciales, 157
 - global, 155
 - informations, 130
 - logique, 134
 - numérique, 135
 - probabilité, 137
 - séquence, 149, 151
 - sur les bits, 137
 - trigonométrique, 136
 - valeurdate, 145
 - Valeurs manquantes, 97
 - valeurs nulles et non renseignées, 156
- client
 - répertoire par défaut, 231
- codage de texte, 53
- codage par défaut du flux, 53

- codage UTF-8, 53, 268
- coller, 19
- commande set, 58, 106
- commentaires
 - lister tout sur un flux, 76
 - raccourcis clavier, 261–262
 - sur des noeuds et des flux, 70
- compatibilité 508, 255
- concaténation de chaînes, 131
- conditions, 105
- connexion à IBM SPSS Modeler Server, 9
- connexion unique, 11, 163
- connexion unique, IBM SPSS Collaboration and Deployment Services Repository, 159, 162
- connexions
 - à IBM SPSS Collaboration and Deployment Services Repository, 162–163
 - à IBM SPSS Modeler Server, 9, 11–12
 - groupe de serveurs, 12
- conseils
 - utilisation générale, 89
- conventions, 130
- Coordinateur de processus, 12
- COP, 12
- copier, 19
- couleurs
 - définition, 235
- couper, 19
- création d'une palette personnalisée, 244
 - création de sous-palettes, 247
- CRISP-DM, 18, 214
 - mode d'affichage des projets, 215
- data mining, 25
 - exemples d'application, 37
 - stratégie, 28
- dates
 - conversion, 148
 - Manipulation, 148
- Décimales
 - Formats d'affichage, 53
- découverte de connaissances, 25
- défaut
 - phase de projet, 215
- défilement
 - définition des options, 54
- degrés
 - unités de mesure, 52
- déploiement de scénarios, 188
- désactiver des noeuds, 43, 45
- descriptions de flux, 65, 68
- déverrouillage d'objets du IBM SPSS Collaboration and Deployment Services Repository, 181
- diagrammes
 - Enregistrement des résultats, 82
- distribution Chi-deux
 - fonctions de probabilité, 137
- distribution *F*
 - fonctions de probabilité, 137
- Distribution normale
 - fonctions de probabilité, 137
- distribution *t*
 - fonctions de probabilité, 137
- documentation, 2
- Données
 - aperçu, 50
 - données parasites, 27
- dossiers, IBM SPSS Collaboration and Deployment Services Repository, 180, 182
- DTD, 210
- encoding, 53, 268
- enregistrement
 - états, 80–81
 - flux, 80
 - noeuds, 80
 - objets de sortie, 82
 - objets multiples, 81
- enregistrements, 25
 - Valeurs manquantes, 95
- ensembles, 51
- ensembles à catégories multiples
 - dans les expressions CLEM, 112
- ensembles à réponses multiples
 - dans les expressions CLEM, 112, 116
- ensembles de dichotomies multiples
 - dans les expressions CLEM, 112
- ensembles de règles
 - évaluation, 51
- entiers, 122–123
- espace de travail, 15
- espace de travail de flux
 - Paramètres, 54
- espaces
 - suppression dans les chaînes, 107, 139
- états
 - chargement, 83
 - enregistrement, 80–81
- Étiquettes
 - affichage, 54
 - value, 210
 - variable, 210
- étiquettes de version, objet du IBM SPSS Collaboration and Deployment Services Repository, 186
- étiquettes, objet du IBM SPSS Collaboration and Deployment Services Repository, 186
- exécution des flux, 68
- exemples
 - Aperçu, 4
- Exemples
 - Guide des applications, 2
- exemples d'application, 2
- exportation
 - PMML, 209, 212

- Exportation
 - descriptions de flux, 68
- expressions, 122
- expressions CLEM
 - Performances, 254
 - recherche et remplacement de texte, 118
- extraction d'objets dans le IBM SPSS Collaboration and Deployment Services Repository, 175

- factor, 265
- Fenêtre Arbre interactif
 - accessibilité, 265
- fenêtre principale, 15
- fichier de dictionnaire, 265
- fichiers de données de texte
 - encoding, 268
- fichiers de flux de sauvegarde
 - restauration, 80
- Fichiers de résultat
 - enregistrement, 82
- fichiers journaux
 - affichage du code SQL généré, 240
- fichiers PowerPoint, 216
- Fichiers SAS
 - encoding, 268
- fichiers Statistics
 - encoding, 268
- flux, 8, 15
 - ajout aux projets, 216–217
 - ajout de commentaires, 70
 - ajout de noeuds, 40, 43
 - annotation, 70, 78
 - chargement, 83
 - connexion de noeuds, 40
 - contournement de noeuds, 42
 - création, 38
 - désactivation de noeuds, 43
 - enregistrement, 80
 - Exécution, 68
 - fichiers de sauvegarde, 80
 - options, 51
 - options de déploiement, 188
 - renommer, 65, 78
 - stockage dans le IBM SPSS Collaboration and Deployment Services Repository, 172
- flux de données
 - création, 38
- fonction @BLANK, 97, 130, 156
- fonction @DIFF, 149, 151
- fonction (exponentielle) de puissance, 135
- fonction @FIELD, 97, 157
- fonction @FIELDS_BETWEEN, 97, 110, 157
- fonction @FIELDS_MATCHING, 97, 110, 157
- fonction @INDEX, 149, 151
- fonction @LAST_NON_BLANK, 149, 151, 156
- fonction @MAX, 149, 151
- fonction @MEAN, 149, 151
- fonction @MIN, 149, 151
- fonction @MULTI_RESPONSE_SET, 112, 157
- fonction @NULL, 97, 130, 156
- fonction @OFFSET, 149, 151
 - remarques sur les performances, 254
- fonction @PARTITION_FIELD, 157
- fonction @PREDICTED, 157
- fonction @SDEV, 149, 151
- fonction @SINCE, 149, 151
- fonction @SUM, 149, 151
- fonction @TARGET, 157
- fonction @TESTING_PARTITION, 157
- fonction @THIS, 149, 151
- fonction @TODAY, 145
- fonction @TRAINING_PARTITION, 157
- fonction @VALIDATION_PARTITION, 157
- fonction abs, 135
- fonction aléatoire, 139
- fonction allbutfirst, 139
- fonction allbutlast, 139
- fonction alphabefore, 139
- fonction arccos, 136
- fonction arccosh, 136
- fonction arcsin, 136
- fonction arcsinh, 136
- fonction arctan, 136
- fonction arctan2, 136
- fonction arctanh, 136
- fonction cdf_chisq, 137
- fonction cdf_f, 137
- fonction cdf_normal, 137
- fonction cdf_t, 137
- fonction cos, 136
- fonction cosh, 136
- fonction count_equal, 110, 132
- fonction count_greater_than, 110, 132
- fonction count_less_than, 110, 132
- fonction count_non_nulls, 132
- fonction count_not_equal, 110, 132
- fonction count_nulls, 97, 110, 132
- fonction count_substring, 139
- fonction date_before, 132
- fonction datetime_date, 131
- fonction DIFF, 151
- fonction div, 135
- fonction endstring, 139
- fonction exponentielle, 135
- fonction first_index, 112, 132
- fonction first_non_null, 112, 132
- fonction first_non_null_index, 112, 132
- fonction fracof, 135
- fonction hasendstring, 139
- fonction hasmidstring, 139
- fonction hasstartstring, 139
- fonction hassubstring, 139
- fonction INDEX, 151
- fonction integer_bitcount, 137

- fonction integer_leastbit, 137
- fonction integer_length, 137
- fonction intof, 135
- fonction is_date, 130
- fonction is_datetime, 130
- fonction is_integer, 130
- fonction is_number, 130
- fonction is_real, 130
- fonction is_string, 130
- fonction is_time, 130
- fonction is_timestamp, 130
- fonction isalphacode, 139
- fonction isendstring, 139
- fonction islowercode, 139
- fonction ismidstring, 139
- fonction isnumbercode, 139
- fonction isstartstring, 139
- fonction issubstring, 139
- fonction issubstring_count, 139
- fonction issubstring_lim, 139
- fonction isuppercode, 139
- fonction last_index, 112, 132
- fonction LAST_NON_BLANK, 151
- fonction last_non_null, 112, 132
- fonction last_non_null_index, 112, 132
- fonction length, 139
- fonction locchar, 139
- fonction locchar_back, 139
- fonction log, 135
- fonction log10, 135
- fonction lowertoupper, 139
- fonction matches, 139
- fonction max, 132
- fonction MAX, 151
- fonction max_index, 112, 132
- fonction max_n, 110, 132
- fonction MEAN, 149, 151
- fonction mean_n, 110, 135
- fonction member, 132
- fonction min, 132
- fonction MIN, 151
- fonction min_index, 112, 132
- fonction min_n, 110, 132
- fonction mod, 135
- fonction negate, 135
- fonction OFFSET, 151
- fonction oneof, 139
- fonction pi, 136
- fonction random0, 139
- fonction rem, 135
- fonction replace, 139
- fonction replicate, 139
- fonction round, 135
- fonction SDEV, 151
- fonction sdev_n, 110, 135
- fonction sign, 135
- fonction sin, 136
- fonction SINCE, 151
- fonction sinh, 136
- fonction skipchar, 139
- fonction skipchar_back, 139
- fonction soundex, 144
- fonction soundex_difference, 144
- fonction sqrt, 135
- fonction startstring, 139
- fonction stripchar, 139
- fonction strmember, 139
- fonction subsers, 139
- fonction substring, 139
- fonction substring_between, 139
- fonction SUM, 151
- fonction sum_n, 110, 135
- fonction tan, 136
- fonction tanh, 136
- fonction testbit, 137
- fonction THIS, 151
- fonction time_before, 132
- fonction to_date, 131, 145
- fonction to_dateline, 145
- fonction to_datetime, 131
- fonction to_integer, 131
- fonction to_number, 131
- fonction to_real, 131
- fonction to_string, 131
- fonction to_time, 131, 145
- fonction to_timestamp, 131, 145
- fonction trim, 139
- fonction trim_start, 139
- fonction trimend, 139
- fonction undef, 156
- fonction unicode_char, 139
- fonction unicode_value, 139
- fonction uppertolower, 139
- fonction value_at, 112, 132
- fonctions, 125–126, 130, 149
 - @BLANK, 96
 - dans les expressions CLEM, 115
 - Exemples, 102
 - @FIELD, 113, 157
 - @GLOBAL_MAX, 155
 - @GLOBAL_MEAN, 155
 - @GLOBAL_MIN, 155
 - @GLOBAL_SDEV, 155
 - @GLOBAL_SUM, 155
 - @PARTITION, 157
 - @PREDICTED, 113, 157
 - @TARGET, 113, 157
 - traitement des valeurs manquantes, 97
- fonctions CLEM
 - Conversion, 131
 - liste des fonctions disponibles, 128
- fonctions datetime
 - datetime_date, 145
 - datetime_day, 145

- datetime_day_name, 145
- datetime_day_short_name, 145
- datetime_hour, 145
- datetime_in_seconds, 145
- datetime_minute, 145
- datetime_month, 145
- datetime_month_name, 145
- datetime_month_short_name, 145
- datetime_now datetime_second, 145
- datetime_time, 145
- datetime_timestamp, 145
- datetime_weekday, 145
- datetime_year, 145
- fonctions de comparaison, 132
- fonctions de conversion, 131
- fonctions de date, 125–126
 - date_before, 132, 145
 - date_days_difference, 145
 - date_in_days, 145
 - date_in_months, 145
 - date_in_weeks, 145
 - date_in_years, 145
 - date_months_difference, 145
 - date_weeks_difference, 145
 - date_years_difference, 145
 - fonction @TODAY, 145
- Fonctions de distribution, 137
- fonctions de probabilité, 137
- fonctions d'heure, 125–126
 - time_before, 132, 145
 - time_hours_difference, 145
 - time_in_hours, 145
 - time_in_mins, 145
 - time_in_secs, 145
 - time_mins_difference, 145
 - time_secs_difference, 145
- fonctions d'heure et de date, 125–126
- fonctions d'informations, 130
- fonctions globales, 155
- fonctions if, then, else, 134
- Fonctions logiques, 134
- fonctions numériques, 135
- fonctions séquentielles, 149, 151
- fonctions spéciales, 157
- Fonctions sur chaînes, 139
- fonctions sur les bits, 137
- fonctions trigonométriques, 136
- format d'affichage monétaire, 53
- Formats d'affichage
 - Décimales, 53
 - Monnaie, 53
 - nombres, 53
 - scientifique, 53
 - symbole de regroupement, 53
- formats d'affichage des nombres, 53
- Formats de date, 52, 125–126
- formats d'heure, 52, 125–126
- Générateur de formules, 260
 - accès, 114
 - Aperçu, 113
 - recherche et remplacement de texte, 118
 - Utilisation, 114
- génération de scripts, 24, 99
 - recherche et remplacement de texte, 118
- génération SQL, 237
 - consignation, 240
 - prévisualisation, 240
- gestion des blancs
 - fonctions CLEM, 156
- gestionnaires, 16
- graphiques
 - ajout aux projets, 216
 - Enregistrement des résultats, 82
- IBM InfoSphere Warehouse (ISW)
 - exportation en PMML, 212
- IBM SPSS Collaboration and Deployment Services, 161
- IBM SPSS Collaboration and Deployment Services Enterprise View, 161, 188
- IBM SPSS Collaboration and Deployment Services Repository, 159, 161
 - accès, 164
 - connexion, 162–163
 - connexion unique, 162
 - dossiers, 180, 182
 - extraction d'objets, 175
 - propriétés de l'objet, 184
 - recherche, 177
 - stockage d'objets, 166
 - suppression d'objets et de versions, 181
 - transfert de projets, 219
 - verrouillage et déverrouillage d'objets, 181
- IBM SPSS Modeler, 1, 14
 - Aperçu, 8, 230
 - astuces et raccourcis, 89
 - démarrage, 8
 - démarrage à partir de la ligne de commande, 9
 - documentation, 2
 - fonctions d'accessibilité, 255
 - options, 230
- IBM SPSS Modeler Advantage, 161, 187
- IBM SPSS Modeler Server
 - ID utilisateur, 9
 - mot de passe, 9
 - nom de domaine (Windows), 9
 - nom d'hôte, 9, 11
 - numéro de port, 9, 11
- IBM SPSS Text Analytics, 2
- Icônes
 - définition des options, 54
- ID utilisateur
 - IBM SPSS Modeler Server, 9
- importation
 - PMML, 210, 212

- Impression, 23
 - flux, 45
- Info-bulles
 - annotation de noeuds, 78
- introduction
 - IBM SPSS Modeler, 8, 230
- Introduction, 122
- invites, exécution, 60

- Java, 267
- JAWS, 255, 265, 267

- keywords
 - annotation de noeuds, 78

- langue
 - options, 230
- Lecteurs d'écran, 257, 260, 265
 - exemple, 263–264
- ligne de commande
 - démarrage de IBM SPSS Modeler, 9
- liste de tous les commentaires sur un flux, 76
- listes, 122, 124

- mappage de champs, 84
- mappage des données, 87
- marques commerciales, 270
- mémoire
 - gestion, 230–231
- messages
 - affichage du code SQL généré, 240
- messages d'erreur, 56
- méthode de différence
 - conversion des confiances en scores de propension, 203
 - modèles R. neurones (Réseau de neurones), 203
- méthode SoftMax
 - conversion des confiances en scores de propension, 203
 - modèles R. neurones (Réseau de neurones), 203
- méthodologie CRISP-DM, 28–29
- mise en mémoire cache des noeuds
 - activation, 47, 250
- Modèles, 69, 85
 - ajout aux projets, 216
 - Exportation, 240
 - rafraîchissement, 194
 - remplacement, 234
 - stockage dans le IBM SPSS Collaboration and Deployment Services Repository, 174
- modèles d'arbre décision
 - scores de propension, 203
- modèles de rafraîchissement, 194
- modèles de régression logistique
 - scores de propension, 203
- modèles de réseau de neurones
 - scores de propension, 203
- modèles de valeur, 203
 - modèles d'ensemble de règles.
 - scores de propension, 203
 - modèles d'offre, 203
 - modèles IBM SPSS Statistics, 37
 - modèles PMML
 - régression linéaire, 241
 - Régression logistique, 241
 - modèles Statistics, 37
 - modélisation
 - branche, 70
 - mot de passe
 - IBM SPSS Modeler Server, 9

- navigation
 - raccourcis clavier, 257
- noeud
 - ajout aux projets, 216
- noeud Agréger
 - Performances, 253
- noeud Audit données
 - utilisation au cours de l'exploration, 25
- Noeud Audit données
 - utilisation dans le Data mining, 27
- noeud Bâtir règle
 - chargement, 83
- noeud Discrétiser
 - Performances, 252
- noeud Distinguer
 - Performances, 253
- Noeud Enterprise View, 188
- noeud Evaluation
 - Performances, 254
- noeud fichier cache
 - chargement, 83
- noeud Fusionner
 - Performances, 253
- noeud K-Means
 - grands ensembles, 51
 - Performances, 254
- noeud Kohonen
 - grands ensembles, 51
 - Performances, 254
- noeud Qualité
 - Valeurs manquantes, 95
- noeud R. neurones
 - Performances, 254
- noeud R. neurones (Réseau de neurones)
 - grands ensembles, 51
- noeud Remplacer
 - Valeurs manquantes, 96
- Noeud Sélection de fonction
 - Valeurs manquantes, 95
- noeud Trier
 - Performances, 252
- noeud Typer
 - Performances, 253
 - Valeurs manquantes, 96

- noeuds, 8
 - activation, 43
 - affichage sur la palette, 244
 - ajout, 40, 43
 - ajout aux projets, 217
 - ajout de commentaires sur, 70
 - annotation, 70, 78
 - aperçu des données, 50
 - chargement, 83
 - connexion dans un flux, 40
 - contournement dans un flux, 42
 - création de sous-palettes personnalisées, 247
 - création d'une palette personnalisée, 244
 - définition des options, 45
 - désactivation, 43, 45
 - désactivation dans un flux, 43
 - duplication, 45
 - enregistrement, 80
 - Introduction, 39
 - modification, 45
 - ordre, 249
 - Performances, 252, 254
 - personnalisation d'un onglet palette, 248
 - recherche, 63
 - stockage dans le IBM SPSS Collaboration and Deployment Services Repository, 173
 - Suppression, 40
 - suppression de connexions, 44
 - suppression de la palette, 244
 - verrouillage, 50
- noeuds de modélisation, 30, 39
 - Performances, 254
 - personnalisation d'un onglet palette de modélisation, 248
- noeuds de sortie, 39
- noeuds d'exécution, 39
 - Performances, 252
- noeuds source, 39
 - mappage des données, 85
 - rafraîchissement, 51
- noeuds terminaux, 39
- noeuds violets, 237
- nom de domaine (Windows)
 - IBM SPSS Modeler Server, 9
- nom d'hôte
 - IBM SPSS Modeler Server, 9, 11
- nombres, 109, 123
- noms de flux, 78
- noms de noeud, 78
- Notation scientifique
 - Format d'affichage, 53
- notifications
 - définition des options, 232
- Nouvelles fonctionnalités, 6–7
- nuggets, 69
 - défini, 17
- nuggets de modèle, 69
- numéro de port
 - IBM SPSS Modeler Server, 9, 11
- objets
 - propriétés, 223
- objets de sortie
 - stockage dans le IBM SPSS Collaboration and Deployment Services Repository, 173
- observation, 25
- opérateur and, 134
- opérateur Différent de, 132
- opérateur Egal à, 132
- opérateur Inférieur à, 132
- opérateur not, 134
- opérateur or, 134
- opérateur Supérieur à, 132
- opérateurs
 - jonction de chaînes, 131
- optimisation, 237
- options, 230
 - affichage, 235
 - PMML, 240
 - pour IBM SPSS Modeler, 230
 - propriétés du flux, 51, 54, 56
 - utilisateur, 232
- options de déploiement
 - scénarios, 188
- options utilisateur, 232
- outil de mappage de données, 84–85
- Ouverture
 - états, 83
 - flux, 83
 - Modèles, 83
 - noeuds, 83
 - projets, 216
 - sortie, 83
- palette de modèles, 174
- palette de modèles générés, 16
- palettes, 15
 - personnalisation, 242
- paramètres
 - création de modèles, 192
 - dans les expressions CLEM, 116
 - flux, 58, 60–61, 106
 - invites d'exécution, 60
 - scoring, 192
 - session, 58, 60–61, 106
 - type, 61
 - utilisation dans des scénarios, 192
- paramètres de flux, 58, 60–61, 106
- paramètres de session, 58, 60–61, 106
- paramètres régionaux
 - options, 230
- passage de jours, 53
- Performances
 - des noeuds d'exécution, 252

- expressions CLEM, 254
- mise en mémoire cache des noeuds, 47, 250
- noeuds de modélisation, 254
- personnalisation d'un onglet palette, 248
- plusieurs sessions IBM SPSS Modeler, 14
- PMML
 - exportation de modèles, 209, 212
 - importation de modèles, 210, 212
 - options d'exportation, 240
- point, 51
- Polices, 235
- PredictiveMarketing, 201
- priorité, 126
- priorité des opérateurs, 126
- prise en charge d'Unicode, 268
- probabilités
 - applications prédictives, 203
- programmation visuelle, 14
- projets, 18, 214
 - ajout d'objets, 217
 - annotation, 221
 - création, 216–217
 - dans le IBM SPSS Collaboration and Deployment Services Repository, 219
 - définition de propriétés, 220
 - définition d'un dossier par défaut, 215
 - fermeture, 224
 - générations de rapports, 224
 - mode Classes, 216
 - Mode CRISP-DM, 215
 - propriétés de l'objet, 223
 - propriétés du dossier, 222
 - stockage dans le IBM SPSS Collaboration and Deployment Services Repository, 172
- propriétés
 - dossier de projet, 222
 - flux de données, 51
 - phases dans le rapport, 224
- propriétés de l'objet, IBM SPSS Collaboration and Deployment Services Repository, 184
- raccourcis
 - clavier, 22, 257, 260–262
 - utilisation générale, 89
- raccourcis clavier, 257, 260–262
- radians
 - unités de mesure, 52
- rafraîchir
 - noeuds source, 51
- rafraîchissement de modèle, 188
- rapports
 - ajout aux projets, 216
 - définition de propriétés, 224
- recherche
 - de noeuds dans un flux, 63
- recherche de connexions dans COP, 12
- recherche de texte, 118
- rechercher les objets dans le IBM SPSS Collaboration and Deployment Services Repository, 177
- redimensionnement, 21
- réduction, 21
- réécriture des flux
 - activation, 237
- réels, 122–123
- régression, 265
- régression linéaire
 - exportation au format PMML, 241
- Régression logistique, 265
 - exportation au format PMML, 241
- remplacement de modèles, 234
- remplacement de texte, 118
- renommer
 - flux, 65
 - noeuds, 78
- répercussions, 237
- répertoire
 - default, 231
- répertoire temporaire, 13
- scénarios, 187
 - défini, 161
 - options de déploiement, 188
- scores de propension
 - applications prédictives, 203
- scoring
 - branche, 70, 192–193, 198
- sélection de la palette de noeud, 244
- serveur
 - ajout de connexions, 11
 - connexion, 9
 - recherche de serveurs dans COP, 12
 - répertoire par défaut, 231
- sortie, 16
- sortie HTML
 - Lecteur d'écran, 265
- souris
 - utilisation dans IBM SPSS Modeler, 22, 40
- sous-palette
 - affichage dans l'onglet palette, 246
 - création, 247
 - suppression de l'onglet palette, 246
- SPSS Modeler Server, 1
- stockage des objets dans le IBM SPSS Collaboration and Deployment Services Repository, 166
- Super noeud
 - paramètres, 58, 106
- symbole de regroupement
 - formats d'affichage des nombres, 53
- symbole décimal, 51
 - formats d'affichage des nombres, 53
- système
 - options, 230

tableaux

- ajout aux projets, 216

Tableaux, 260

- Enregistrement des résultats, 82

Tableaux de bord

- Enregistrement des résultats, 82

- génération, 224

- touches de raccourci, 22

traitement parallèle

- activation, 237

- type de déploiement, 188

Types de données, 104

- dans les paramètres, 61

- valeurs, 104

Valeurs

- affichage à partir d'un audit de données, 116

- ajout dans des expressions CLEM, 116

- valeurs de date/heure, 109

valeurs globales

- dans les expressions CLEM, 116

valeurs manquantes

- remplacement, 93

- traitement, 93

Valeurs manquantes, 94–95, 108

- dans les enregistrements, 95

- expressions CLEM, 97

- valeurs manquantes système, 93

- valeurs manquantes utilisateur, 93

- valeurs nulles, 93, 108

Variables, 25

- vérification des expressions CLEM, 118

- verrouillage des noeuds , 50

- verrouillage d'objets du IBM SPSS Collaboration and

- Deployment Services Repository, 181

- virgule, 51

- zoom, 19