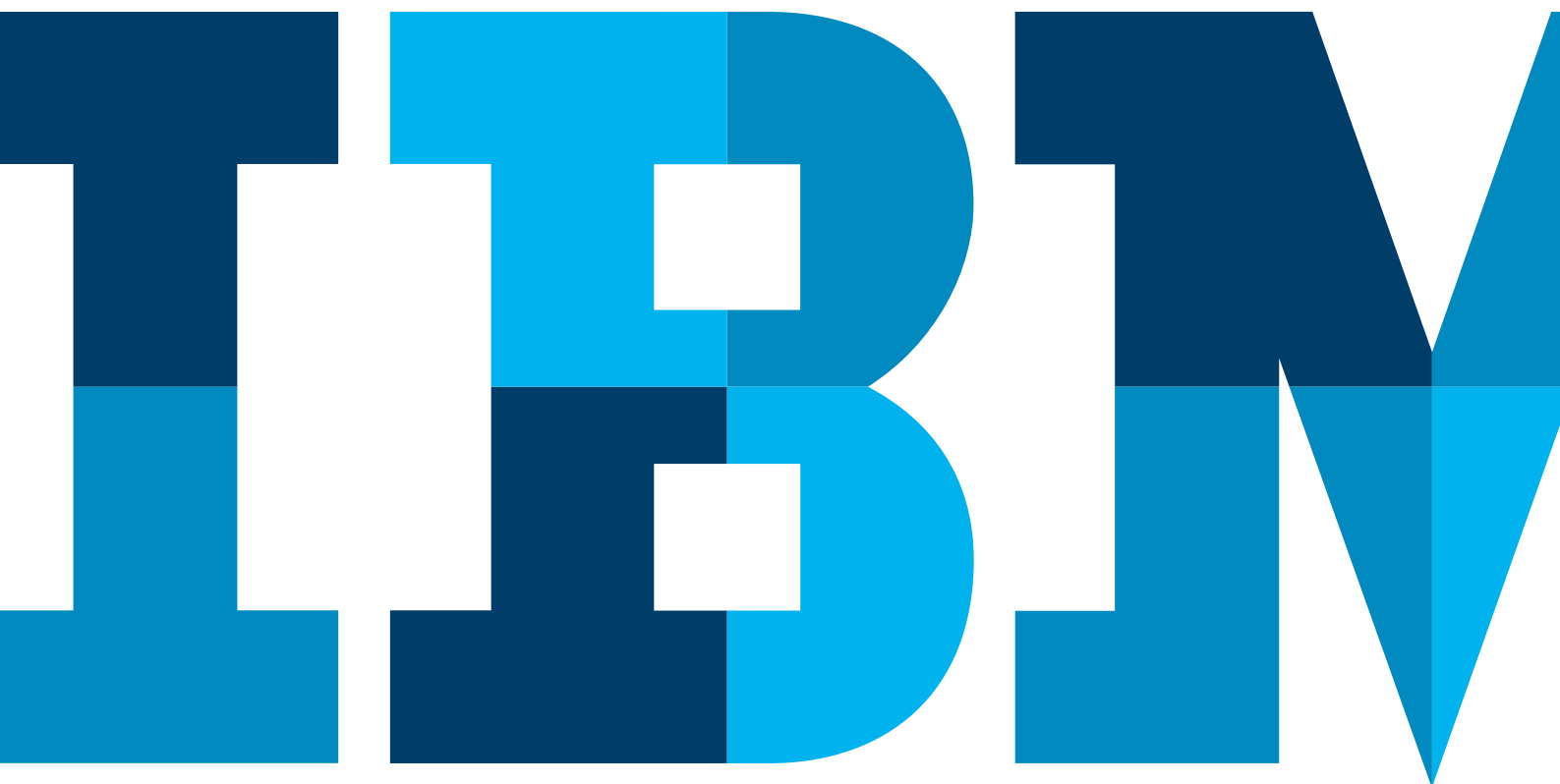


# La puissance d'IBM SPSS Statistics et de R combinés

A large, stylized graphic of the letters 'IBM' in a bold, sans-serif font. The letters are composed of horizontal bars in two shades of blue: a dark blue and a lighter blue. The 'I' is dark blue on top and light blue on the bottom. The 'B' is light blue on top and dark blue on the bottom. The 'M' is dark blue on top and light blue on the bottom. The letters are set against a white background.

## Sommaire

- 2 Résumé
- 2 Pourquoi intégrer SPSS Statistics et R ?
- 4 Intégration de R à IBM SPSS Statistics
- 4 Bundles d'extension : Utiliser des programmes R créés par d'autres
- 7 Écrire vos propres programmes
- 10 Conclusion

## Résumé

Cet article présente les fonctions et capacités résultant de l'intégration d'IBM SPSS Statistics et de R. Ainsi, les utilisateurs R bénéficient des avantages propres à IBM SPSS Statistics en termes de gestion des données, de convivialité et de qualité des sorties pour les présentations. Pour leur part, les utilisateurs SPSS Statistics peuvent accéder à une collection riche et toujours évolutive de bibliothèques d'analyses statistiques et de graphiques qui leur permettent de dégager des connaissances plus approfondies de leurs données. En utilisant conjointement IBM SPSS Statistics et R, vous tirez le meilleur parti des deux produits.

## Pourquoi intégrer SPSS Statistics et R ?

IBM SPSS Statistics compte parmi les toutes premières solutions logicielles statistiques à l'échelle mondiale, avec des modèles prédictifs et des analyses avancées pour vous aider à résoudre les problèmes touchant à vos activités et à vos recherches. Pour bon nombre d'entreprises, de laboratoires de recherche et de statisticiens, IBM SPSS Statistics constitue le standard « de facto » en matière d'analyse statistique. Les organisations utilisent SPSS Statistics pour :

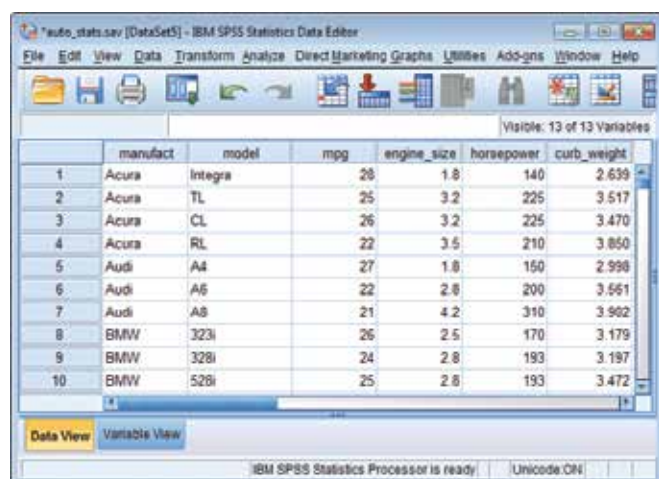
- Comprendre les données.
- Analyser les tendances.
- Prévoir et planifier.
- Valider les hypothèses.
- Tirer des conclusions précises.

Les développements et tests de SPSS Statistics n'ont jamais cessé depuis l'apparition du produit, en 1968. Depuis, de nombreuses formes d'analyse statistique ont été intégrées au logiciel. Parallèlement, les algorithmes qui exécutent les équations ont été testés à la fois par des développeurs et des utilisateurs en milieu universitaire, en laboratoire et dans pratiquement tout type d'entreprise. Les utilisateurs peuvent donc être assurés que le logiciel a été minutieusement testé et que la fiabilité de ses résultats a été attestée.

L'environnement SPSS Statistics facilite et accélère l'affichage, la gestion et l'analyse de tout ensemble de données, ce qui inclut – entre autres – les données d'enquêtes, les bases de données d'entreprises et les données téléchargées depuis le Web. Les procédures statistiques et affichages avancés vous permettent de disposer d'une plateforme à la fois robuste, conviviale et intégrée pour la compréhension de vos données et pour la résolution des problèmes complexes touchant à vos activités et à vos recherches.

IBM SPSS Statistics vous aide à maîtriser *tous* les aspects du processus analytique, depuis la préparation des données jusqu'à la gestion de ces données en passant par l'analyse et la génération de rapports. La solution met à disposition des fonctionnalités personnalisées ainsi que des interfaces adaptables aux différents niveaux de compétences et de responsabilités fonctionnelles. Elle permet également aux utilisateurs de communiquer facilement les résultats en créant des graphiques haute résolution et des rapports directement exploitables pour les présentations.

Prenons l'exemple de Robert qui souhaite analyser des données concernant la consommation en carburant de différents types de voitures. La Figure 1 montre une fraction des données telles qu'elles apparaissent dans l'éditeur de données de SPSS Statistics.



	manufact	model	mpg	engine_size	horsepower	curb_weight
1	Acura	Integra	28	1.8	140	2.639
2	Acura	TL	25	3.2	225	3.517
3	Acura	CL	26	3.2	225	3.470
4	Acura	RL	22	3.5	210	3.860
5	Audi	A4	27	1.8	150	2.998
6	Audi	A6	22	2.8	200	3.561
7	Audi	A8	21	4.2	310	3.902
8	BMW	323i	26	2.5	170	3.179
9	BMW	328i	24	2.8	193	3.197
10	BMW	528i	25	2.8	193	3.472

Figure 1 : L'éditeur de données de SPSS Statistics affiche les informations sur la consommation pour différentes marques et différents modèles.

Pour cette analyse, Robert peut commencer par exécuter la procédure Descriptives afin de se faire une idée de la distribution des données relatives à la consommation de carburant. Il va exécuter cette procédure à partir de la boîte de dialogue Descriptives (Figure 2).

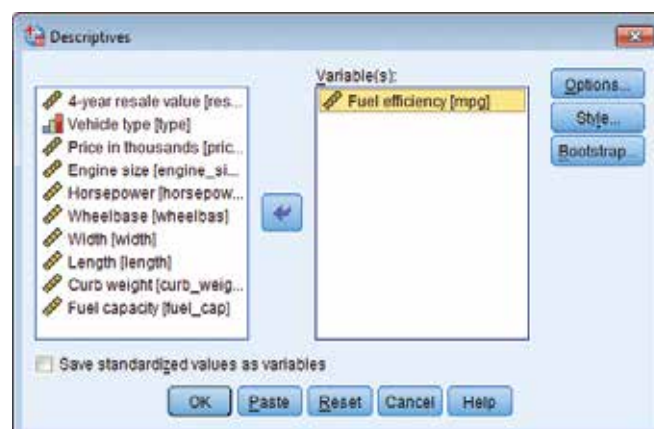


Figure 2 : Boîte de dialogue Descriptives.

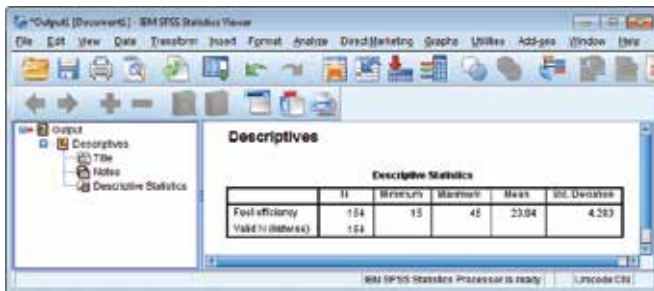


Figure 3 : Sortie de la boîte de dialogue Descriptives dans le visualiseur de SPSS Statistics.

La sortie (ici, une sortie tabulaire) apparaît dans le visualiseur de SPSS Statistics (Figure 3).

R est un langage de programmation et un environnement logiciel en code source ouvert pour les calculs statistiques et les graphiques ([www.r-project.org](http://www.r-project.org)). Le langage R est devenu très populaire auprès des statisticiens et des Data Miners pour le développement de logiciels statistiques, et il est largement utilisé pour l'analyse de données avancée. Ce langage offre un vaste assortiment de techniques statistiques et graphiques avancées, ceci tout en étant hautement extensible. Il est disponible en tant que logiciel en libre accès selon les termes de la [licence publique générale GNU de la Free Software Foundation](http://www.gnu.org/licenses/gpl.html). Il s'exécute sous Windows et MacOS ainsi que sur une large gamme de plateformes UNIX et de systèmes similaires (dont FreeBSD et Linux). R peut être facilement étendu avec différents packages.

Les utilisateurs R débattent couramment sur des sujets tels que l'évolutivité ou les problèmes de base que pose la gestion des données et des sorties, comme par exemple la connexion aux bases de données, l'amélioration de la qualité des sorties et le partage des algorithmes R avec d'autres utilisateurs ne maîtrisant pas le langage R. De leur côté, les utilisateurs SPSS Statistics souhaiteront éventuellement pouvoir utiliser certaines fonctions R dont la solution ne dispose pas, sans pour cela devoir apprendre le langage R – après tout, la programmation en R n'est pas à la portée de tout le monde.

L'intégration de R à SPSS Statistics se justifie donc clairement. Les forces combinées des deux produits aident les deux groupes d'utilisateurs à répondre à leurs besoins respectifs. SPSS Statistics est une plateforme pratique à partir de laquelle les utilisateurs R peuvent gérer de vastes jeux de données et obtenir des graphiques et autres types de sortie d'une grande qualité. La facilité d'emploi constitue l'un des autres avantages de SPSS Statistics, de même que la possibilité de distribuer des packages R intégrés à toutes sortes d'utilisateurs non familiarisés avec R. En outre, cette intégration permet aux utilisateurs SPSS Statistics d'accéder facilement à près de 4000 fonctions statistiques en code source ouvert.

### Intégration de R à IBM SPSS Statistics

Pour pouvoir utiliser les fonctions de programmation R avec SPSS Statistics, il vous faut le module d'extension SPSS Statistics-Integration Plug-In for R. Ce module est disponible gratuitement et s'installe en tant que partie intégrante de SPSS Statistics-Essentials for R. Il vous est nécessaire si vous voulez utiliser des extensions écrites aussi bien par vous-même que par d'autres utilisateurs. SPSS Statistics-Integration Plug-In for R fait partie d'une famille de plug-ins d'intégration disponibles gratuitement. Cette famille comprend des plug-ins pour Python, Java et .NET. R communique avec SPSS Statistics par le biais d'API incluses dans le plug-in, et l'intégration nécessite l'écriture de code encapsuleur R. SPSS Statistics-Integration Plug-in for R étend le langage de syntaxe de commande avec toutes les capacités du langage de programmation R. Ce plug-in permet également d'accéder à un environnement de développement R intégré qui facilite le développement, les tests et le débogage des programmes R devant être utilisés avec SPSS Statistics. Il est disponible pour Windows, Linux, Mac OS et SPSS Statistics Server.

Après installation de SPSS Statistics-Integration Plug-In for R, vous pouvez décider soit d'utiliser des programmes R déjà écrits, soit d'écrire vos propres programmes.

### Bundles d'extension : Utiliser des programmes R créés par d'autres

L'intégration de SPSS Statistics et de R vous permet d'exploiter des programmes R que d'autres personnes ont écrits et mis à disposition sous forme de « bundles d'extension ». Les algorithmes précodés dispensent d'une programmation R intensive, ce qui est particulièrement utile si vous êtes pressé par le temps ou manquez de compétences dans cette discipline. Ces programmes R sont déployés sous forme de bundles d'extension.

Dès lors qu'un bundle d'extension est installé, sa boîte de dialogue est accessible à partir des menus SPSS Statistics, et la commande d'extension peut être exécutée de la même façon que toute commande intégrée. Le programme R fonctionne comme s'il s'agissait d'une boîte de dialogue native et d'une commande de syntaxe.

Afin de mieux comprendre ce qu'est un bundle d'extension, nous allons reprendre l'exemple de cet utilisateur qui analyse des données sur la consommation de carburant. Nous supposons que Robert souhaite maintenant analyser la consommation au regard de la cylindrée du moteur, du nombre de chevaux et de la masse en état de marche du véhicule, sachant toutefois qu'il veut aller au-delà de la régression linéaire standard et réaliser l'analyse avec la régression quantile. La régression quantile est mise à disposition en tant que bundle d'extension dans le plug-in d'intégration pour R, et Robert peut utiliser cette régression pour comprendre la distribution de la consommation par rapport aux différents prédicteurs. Pour les lecteurs plus techniques, nous dirons que la régression quantile estime un ou plusieurs quantiles conditionnels ( $0 \leq q < 1$ ) pour un modèle linéaire, alors que la régression ordinaire se limite à estimer la moyenne conditionnelle.

La boîte de dialogue pour la régression quantile se présente exactement comme n'importe quelle autre boîte de dialogue native de SPSS Statistics (Figure 4).

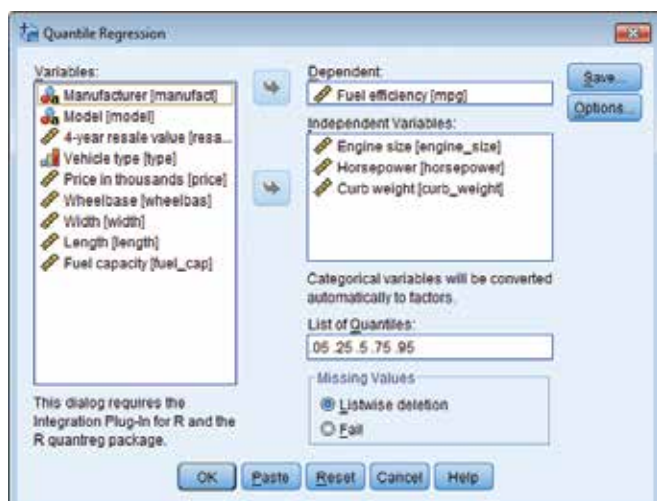


Figure 4 : Boîte de dialogue Régression quantile.

Il suffit à Robert de spécifier les variables dépendantes et indépendantes, d'entrer la liste des quantiles, puis de cliquer sur OK pour lancer l'analyse. SPSS Statistics appelle R et exécute l'analyse en utilisant le package R `quantreg`, mais tout cela est transparent pour Robert. Les résultats de l'analyse au niveau R sont ensuite présentés sous forme tabulaire et graphique dans le visualiseur de SPSS Statistics. Comme pour toute analyse de régression type, les résultats incluent une sortie tabulaire des coefficients de régression. Ici, un tableau des coefficients de régression distinct est créé pour chacun des quantiles spécifiés (voir Figure 5 avec affichage du tableau pour le quantile 0.05). Le tableau contient les résultats qui ont été obtenus suite à exécution de l'analyse dans l'environnement R, puis qui ont été extraits de cet environnement pour affichage dans le visualiseur de SPSS Statistics.

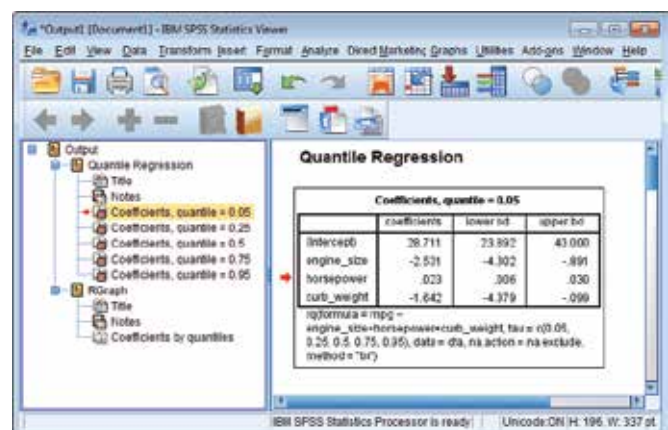


Figure 5 : Tableau de coefficients de régression.



## Écrire vos propres programmes

Le recours aux bundles d'extension n'est qu'une façon d'utiliser R dans SPSS Statistics. Vous pouvez écrire votre propre programme R et l'intégrer à SPSS Statistics à différents niveaux en procédant comme suit :

- Créez une boîte de dialogue personnalisée qui va générer la syntaxe pour une commande d'extension R ou du code R explicite.
- Créez une commande d'extension implémentée en R.
- Exécutez le code R directement à partir de SPSS Statistics.

Dans ces structures, vous avez accès à la fois au langage de programmation R et aux fonctions spécifiques à SPSS Statistics, disponibles dans le pack R Integration Package for SPSS Statistics. Vous pouvez également écrire des fonctions R qui utilisent des fonctionnalités SPSS Statistics depuis l'environnement R avec retour des résultats vers cet environnement.

Vous pouvez utiliser ces fonctions pour :

- Lire les données d'observation de l'ensemble de données actif dans l'environnement R.
- Extraire des informations sur les données de l'ensemble de données actif.
- Extraire les résultats en sortie des commandes de syntaxe.
- Écrire des résultats R (depuis un nouvel ensemble de données vers un tableau croisé dynamique et vers des graphiques) pour affichage dans SPSS Statistics.

### Créer une boîte de dialogue personnalisée qui va générer la syntaxe pour une commande d'extension R ou du code R explicite

Le générateur de boîtes de dialogue personnalisées vous permet de créer une interface utilisateur qui génère une syntaxe de commande pour une commande d'extension implémentée en R. Vous pouvez ensuite afficher la sortie (Figures 7-9) par exécution de la boîte de dialogue dans le visualiseur de SPSS Statistics. Il est également possible d'imbriquer un programme R directement dans une boîte de dialogue personnalisée.

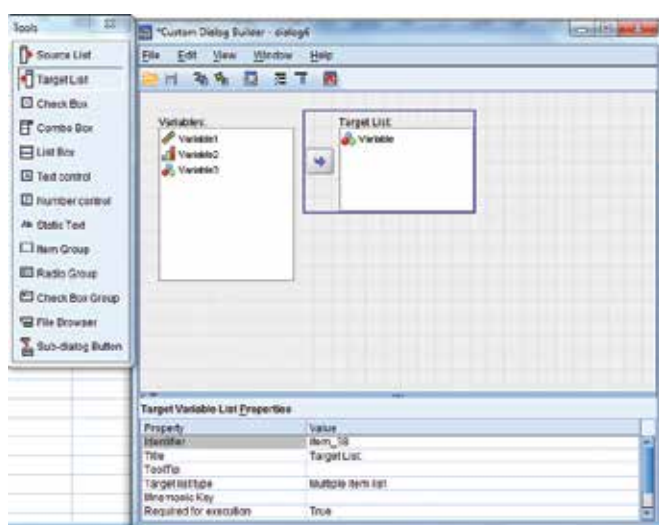


Figure 8 : Exemple de présentation d'une boîte de dialogue personnalisée.

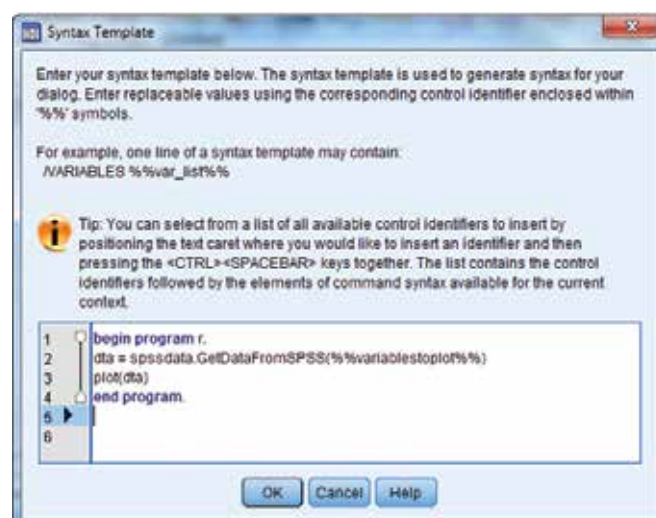


Figure 9 : Modèle de syntaxe pour code R explicite.

### Créer une commande d'extension implémentée en R

Une commande d'extension est une commande SPSS Statistics qui est implémentée en R, Python ou Java. Vous pouvez intégrer un programme R dans SPSS Statistics en créant une commande d'extension qui implémente le programme R. Il vous sera particulièrement utile d'intégrer un algorithme R à SPSS Statistics si vous avez besoin d'une fonction statistique avancée sans disposer de l'expertise ou du temps nécessaire pour créer ce type de programme. Dans un tel scénario, le code pourra être écrit et distribué sous forme de commande d'extension par un groupe de méthodologie chargé de créer des algorithmes R pour les fonctions statistiques très demandées. Cette commande pourra alors être utilisée tout comme s'il s'agissait d'une commande SPSS Statistics intégrée.

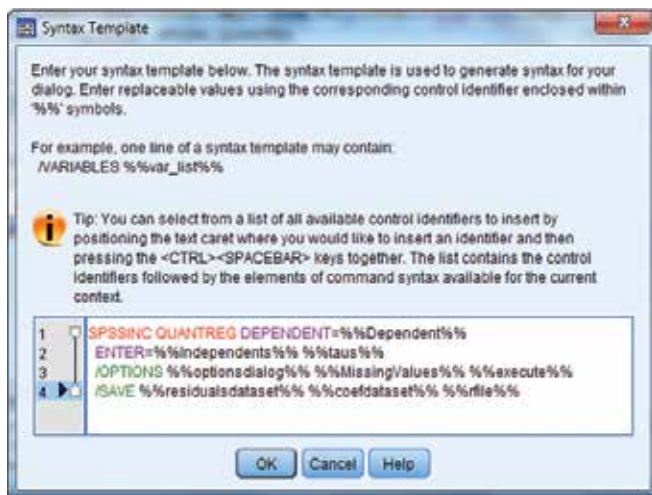


Figure 10 : Modèle de syntaxe pour une commande d'extension.

Pour créer une commande d'extension :

- Écrivez le programme comme vous le feriez pour une fonction R.
- Définissez la syntaxe SPSS Statistics pour la commande d'extension dans un fichier XML qui spécifie le nom de la commande, les sous-commandes et les mots clés (Figure 11).
- Déclarez la syntaxe dans une « fonction d'exécution R » et appelez la fonction (Figure 12).

L'entrée utilisateur est automatiquement validée et mappée aux variables R puis dirigée vers la fonction d'implémentation.

Le code R appelle les API dans le plug-in d'intégration pour l'interaction de R avec SPSS Statistics. Vous pouvez demander une traduction pour les textes du code R devant apparaître en sortie (libellés de tableau croisé dynamique ou messages d'erreur, par exemple).



Figure 11 : Syntaxe pour la commande d'extension.



```

62 Run<-function(args){
63   args <- args[1:]
64   oobj<-spsspkg SyntaxTemplateList(
65     spsspkg Template("DEPENDENT", subco="", Htype="existingvarlist", var="y", islst=FALSE),
66     spsspkg Template("ENTER", subco="", Htype="existingvarlist", var="x", islst=TRUE),
67     spsspkg Template("MISSING", subco="OPTIONS", Htype="lst", var="missing",
68       valist = list("sdvcs", "lsl"))
69   )
70
71   # ("HELP" %in% attr(args,"names"))
72   writeLines(helpbed)
73   else
74     res <- spsspkg processomo(oobj,args,"cbarr")
75 }

```

Figure 12 : Déclaration de la syntaxe et appel de la fonction.

### Exécution du code R directement à partir de SPSS Statistics

Si R vous est familier, vous pouvez exécuter votre propre code R dans SPSS Statistics. Pour cela, vous devez placer le code dans un bloc BEGIN PROGRAM R – END PROGRAM de la syntaxe de commande SPSS Statistics. Quand vous soumettez la syntaxe, le code entré dans le bloc est exécuté en R. Ce code est généralement une combinaison de code R ordinaire et appelle des fonctions R spécifiques SPSS (fournies avec le module Integration Plug-in for R) qui permettent à R d'interagir avec SPSS Statistics.

La Figure 13 montre un exemple de programme R utilisé pour exécuter une régression.

Les fonctions qui facilitent l'utilisation de fonctions de programmation R avec la syntaxe de commande SPSS incluent :

- `spssdata.GetDataFromSPSS` (Extrait des données de l'ensemble de données actif).
- `spssdictionary.GetDictionaryFromSPSS` (Extrait des informations de dictionnaire de variables de l'ensemble de données actif).
- `spsspivortable.Display` (Restitue une sortie tabulaire de R sous forme de tableau croisé dynamique pouvant être affiché dans le visualiseur d'IBM SPSS Statistics ou pouvant être écrit dans un fichier externe avec le système de gestion des sorties de SPSS Statistics. Les tableaux produits avec cette fonction se présentent exactement comme des tableaux croisés dynamiques générés par le code SPSS natif).

1	begin program r.	2,3	y, x are dependent and independent variables
2	y = "salary"	5	combine into one list and get data, converting categorical variables to factors
3	x = "salbegin jobcat prevexp"	7	equation needs to look like y~x1+x2*x3 - convert blank(s) to + - "~ +" is a regular expression
4		8,9	add dep var and combine into string
5	dta = spssdata.GetDataFromSPSS(c(y, x),		
6	missingValueToNA=TRUE, factorMode="levels")		
7	oplus = paste("~", "+", x)		
8	eqn = c(y, "~", oplus)		
9	eqn = paste(eqn, collapse="")		
10	print(eqn)		
11	res = lm(as.formula(eqn), data=dta, na.action=na.omit)	11	run the equation -convert string to formula object -identify data source -omit missing values -create result object "res" -notice the factor
12	print(res)		
13	end program.		

Figure 13 : Un programme R pour l'exécution d'une régression.

Tous les artefacts peuvent être facilement regroupés pour créer un bundle d'extension qui simplifiera la distribution et l'installation. Le bundle d'extension contient :

- Un fichier de package de boîte de dialogue personnalisée (.spd) qui spécifie la boîte de dialogue personnalisée.
- Un fichier XML qui spécifie la syntaxe de la commande d'extension.
- Le ou les fichiers de code d'implémentation écrits en R.

Les autres utilisateurs qui ont installé Essentials for R peuvent simplement passer par les menus SPSS Statistics pour installer le bundle d'extension. Ils peuvent ensuite utiliser la boîte de dialogue et la commande d'extension de la même manière qu'une boîte de dialogue ou une commande native. Si approprié, pour permettre la collaboration et le partage des connaissances, des extensions basées R d'appoint peuvent être hébergées sur le site Web de la communauté SPSS.

## Conclusion

SPSS Statistics et R peuvent l'un et l'autre se prévaloir d'atouts qui ont été testés au cours du temps et qui sont très favorablement accueillis dans la communauté statistique. En outre, ces atouts se complètent mutuellement pour créer un ensemble de fonctions encore plus puissant et qui bénéficie à la communauté statistique dans son intégralité.

Les utilisateurs R peuvent ainsi accéder à des capacités supérieures pour la gestion des données, ce qui leur permet de gérer des jeux de données beaucoup plus volumineux. Par ailleurs, avec le système de gestion des sorties de SPSS Statistics, les utilisateurs R disposent d'un ensemble enrichi de fonctions de sortie pour les graphiques et les tableaux croisés dynamiques, d'où une expérience utilisateur de meilleure qualité. Enfin, SPSS Statistics constitue un vecteur de déploiement idéal pour la distribution des packages R vers une large palette d'utilisateurs.

Les utilisateurs SPSS Statistics ont accès à un bien plus grand nombre de fonctions statistiques, ce qui leur permet d'exécuter des analyses sophistiquées sans l'obstacle que constitue l'apprentissage d'un langage de programmation aussi complexe que R. Les avantages à utiliser R et SPSS Statistics conjointement sont donc nombreux et méritent d'être envisagés.

---

## Ressources

Les ressources suivantes pourront aider les utilisateurs qui souhaitent exploiter R dans SPSS Statistics :

- Des instructions pour accéder à Essentials for R sont incluses dans le système d'aide de SPSS Statistics, sous *Integration Plug-in for R Help > How to get the IBM SPSS Statistics – Integration Plug-in for R*.
- Vous pouvez disposer de tutoriels à partir de *Help > Working with R*.
- Une documentation complète pour le module Integration Plug-in for R est disponible dans le système d'aide de SPSS Statistics, sous « Integration Plug-in for R Help ».
- Des questions sur l'utilisation de R dans SPSS Statistics peuvent être postées dans le cadre du forum sur la programmabilité R qui est hébergé par le site de la communauté SPSS sur developerWorks ([ibm.com/developerworks/spssdevcentral](http://ibm.com/developerworks/spssdevcentral)).
- Vous trouverez des informations détaillées sur la création de commandes d'extension dans l'article « Writing IBM SPSS Statistics Extension Commands » ainsi que dans le chapitre « Extension Commands in Programming and Data Management for IBM SPSS Statistics », tous deux disponibles auprès de la communauté SPSS.
- L'article « Extension Bundles from IBM SPSS », disponible auprès de la communauté SPSS, donne la liste des bundles d'extension que vous pouvez télécharger à partir de cette communauté. La liste inclut les bundles d'extension qui implémentent les commandes d'extension Python et ceux qui implémentent les commandes d'extension R.

---

## À propos d'IBM Business Analytics

Les logiciels IBM Business Analytics permettent aux organisations de disposer d'informations orientées données, ce qui les aide à travailler de manière plus intelligente et à surpasser leurs concurrents. Très complet, le portefeuille de produits inclut des solutions d'aide à la décision (Business Intelligence), d'analyse prédictive, de gestion des décisions, de pilotage de la performance et de gestion des risques.

Les solutions Business Analytics aident les entreprises à identifier et à visualiser les tendances et modèles qui relèvent d'activités telles que l'analyse client et qui peuvent avoir un effet déterminant sur la performance métier. Ces entreprises peuvent ainsi comparer des scénarios, anticiper les menaces et opportunités potentielles, mieux planifier, budgétiser et prévoir leurs ressources, équilibrer les risques par rapport aux retours sur investissements prévus, et assurer le respect des réglementations. En élargissant l'utilisation des analyses, les organisations peuvent adapter leurs décisions tactiques et stratégiques de façon à atteindre leurs objectifs. Pour plus d'informations, voir [ibm.com/business-analytics/fr](http://ibm.com/business-analytics/fr)

## Je souhaite être contacté(e)

Pour être contacté ou pour poser une question, accédez au site [ibm.com/business-analytics/fr](http://ibm.com/business-analytics/fr). Un représentant IBM vous répondra sous deux jours ouvrés.



---

**Compagnie IBM France**

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

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

**ibm.com**

IBM, le logo IBM, SPSS et ibm.com sont des marques d'International Business Machines Corp. déposées dans de nombreuses juridictions réparties dans le monde entier. Les autres noms de produit et de service peuvent être des marques d'IBM ou d'autres sociétés. Une liste actualisée de toutes les marques d'IBM est disponible sur la page Web « Copyright and trademark information » à l'adresse suivante : [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

Linux est une marque de Linus Torvalds aux États-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux États-Unis et/ou dans certains autres pays.

Java ainsi que tous les logos et toutes les marques incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

UNIX est une marque enregistrée de The Open Group aux États-Unis et/ou dans certains autres pays.

Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier par lui-même le fonctionnement des produits ou logiciels non-IBM avec les produits ou logiciels IBM.

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

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

© Copyright IBM Corporation 2015



Pensez à recycler ce document