



Sébastien Ringuedé

SAS[®]

3^e édition

Introduction au décisionnel :
du data management au reporting

Web
ressources

Ressources
en ligne

Couvre le programme de la certification
Base Programming for SAS[®]9

Sébastien Ringuedé

Université d'Orléans, LEO, UMR CNRS 7322

SAS[®]

**Introduction au décisionnel :
du data management au reporting**

3^e édition

PEARSON

Base SAS®, SAS® 6.12, SAS® 8.2, SAS®9, SAS® 9.1.3, SAS® 9.2, SAS® 9.3, SAS® 9.4, SAS® Data Integration Studio, SAS® Integration Technologies, SAS® Enterprise Data Integration Server, SAS® Forecast Server, SAS® Enterprise Guide®, SAS® Enterprise Miner™, SAS OnlineDoc® documentation, SAS/ACCESS®, SAS/CONNECT®, SAS/ETS®, SAS/GRAPH®, SAS/IML®, SAS/IML® Studio, SAS/OR®, SAS/QC®, SAS/STAT® sont des marques déposées de SAS Institute inc., Cary, North Carolina, USA.

Publié par Pearson France

Immeuble Terra Nova II

74, rue de Lagny

93100 Montreuil

www.pearson.fr

Mise en pages : Léa B

ISBN : 978-2-3260-0066-7

© 2014 Pearson France

Tous droits réservés

Tous les autres noms de produits ou de marques cités dans cet ouvrage sont des marques déposées par leurs propriétaires respectifs.

Le nom SAS désignera, en fonction du contexte, le produit, le langage ou la société qui commercialise le progiciel SAS. SAS se prononce « sass », comme un nom, et non comme un sigle.

Aucune représentation ou reproduction, même partielle, autre que celles prévues à l'article L. 122-5 2° et 3° a) du code de la propriété intellectuelle ne peut être faite sans l'autorisation expresse de Pearson France ou, le cas échéant, sans le respect des modalités prévues à l'article L. 122-10 dudit code.

Remerciements

Je tiens à remercier SAS France et plus particulièrement Ariane Liger-Belair, directeur académique SAS France pour le soutien apporté à cet ouvrage.

Le propos de cet ouvrage n'engage que son auteur et en aucun cas SAS Institute s.a.s, filiale de SAS Institute Inc., ni SAS Institute Inc.

Cet ouvrage n'aurait pas pu voir le jour sans l'aide de relecteurs. Mes remerciements vont tout particulièrement à :

Richard Eudes, Analytics Expert – Senior Alliances Manager, SAS France,

Frédérique Léger, Data Manager, Fovea,

Jérémy Noël, Senior Associate Technical Support Engineer, SAS France,

ainsi qu'à Gildas Amegbo, Sylvain Dupuis, Nina Elynkina, Hervé Gnazale, Leila Hani, Clément Juillet, Katia Sborets, Jérôme Schouler et Stéphanie Tring.

Un ouvrage, c'est aussi un éditeur : merci aux équipes de Pearson France et plus particulièrement à Estelle Jelen.

Cet ouvrage doit beaucoup aux étudiants du master ESA (Économétrie et Statistique Appliquée) de l'université d'Orléans qui, par leurs questions et leurs remarques depuis 2004, ont enrichi au fil du temps cet ouvrage qui n'était au départ que le support de mon cours d'introduction à SAS.

www.univ-orleans.fr/deg/masters/ESA/

Les erreurs et imprécisions qui pourront apparaître dans cet ouvrage restent miennes.

Le site compagnon de cet ouvrage est un élément indispensable à votre apprentissage de SAS :

www.sas-sr.com

Merci, une nouvelle fois, à Gilbert Colletaz, Christophe Hurlin et Solange Siegwald pour leur patience...

À Laurence, Léonie, Marceau et Lucien qui ne connaissent de SAS que le temps que j'ai passé à la rédaction des première, deuxième et troisième éditions de cet ouvrage...



Table des matières

Introduction	2
Présentation de l'ouvrage	2
Les nouveautés de la présente édition	3
Progresser dans sa connaissance de SAS	3
La Certification Base Programming for SAS 9	4
Les pictogrammes	5
Comment travailler avec cet ouvrage	6
Disposer du système SAS	7
Les différentes versions de SAS	7
SAS Windows, Unix, z/OS	8
Conventions typographiques	8
Une présentation succincte d'autres modules de SAS	10
1. Faire connaissance avec son environnement SAS	13
1.1 Premiers pas avec SAS	13
1.2 Deux étapes distinctes : DATA et PROC	17
1.3 Les tables SAS	19
1.4 Les bibliothèques	20
1.5 Créer une table SAS	23
1.5.1 <i>Créer une première table SAS</i>	23
1.5.2 <i>L'éditeur SAS : une aide à la programmation</i>	27
1.5.3 <i>Demander l'exécution d'un programme</i>	28
1.5.4 <i>Lire le JOURNAL</i>	29
1.5.5 <i>Examiner la table créée</i>	30
1.6 Les sorties SAS	33
1.6.1 <i>Les destinations de vos sorties</i>	33
1.6.2 <i>OPTIONS et TITLE</i>	36
1.7 Compréhension approfondie du cas simple	37
1.8 Introduction de valeurs manquantes	39
2. Aller plus loin dans la création de tables SAS	43
2.1 La création d'une table SAS à partir d'un fichier de données brutes	43
2.1.1 <i>Principes</i>	44
2.1.2 <i>L'INPUT BUFFER</i>	46
2.2 Autres indicateurs de séparation des champs	48
2.2.1 <i>Champs séparés par un caractère quelconque</i>	48
2.2.2 <i>Champs séparés par un point-virgule</i>	49
2.2.3 <i>Champs séparés par des tabulations</i>	50
2.3 Les enregistrements formatés en colonnes	51
2.3.1 <i>Cas général</i>	51
2.3.2 <i>Le pointeur +X</i>	54
2.3.3 <i>Le pointeur @X</i>	55
2.3.4 <i>Une seconde application de @x : '@texte'</i>	55
2.3.5 <i>Un cas particulier : le dernier champ est de longueur variable</i>	56

2.4	Plusieurs enregistrements pour construire une observation.....	58
2.4.1	<i>Un enregistrement, plusieurs observations : @@ et @</i>	58
2.4.2	<i>Plusieurs enregistrements pour une observation</i>	62
2.5	La gestion des valeurs manquantes.....	66
2.5.1	<i>L'instruction globale MISSING</i>	66
2.5.2	<i>Les valeurs manquantes non notées par un quelconque caractère</i>	67
2.6	Les INFORMAT.....	68
2.6.1	<i>Les INFORMAT des variables caractères</i>	68
2.6.2	<i>Autres INFORMAT caractères</i>	72
2.6.3	<i>Longueur des variables caractères</i>	73
2.6.4	<i>Les INFORMAT des variables numériques</i>	73
2.6.5	<i>Des variables numériques particulières : les dates et heures</i>	77
2.6.6	<i>Créons notre propre INFORMAT !</i>	83
2.7	SAS, Excel, l'importation et l'exportation de tables.....	90
2.7.1	<i>L'assistant d'importation – PROC IMPORT</i>	92
2.7.2	<i>LIBNAME EXCEL et PCFILES</i>	94
2.7.3	<i>Deux solutions alternatives</i>	101
2.7.4	<i>L'interface d'exportation</i>	104
2.7.5	<i>La création de fichiers TXT via FILE et PUT</i>	106
2.8	Le débogage des programmes de création de table.....	108
2.8.1	<i>Regardez et comprenez votre fichier</i>	108
2.8.2	<i>Lisez la fenêtre JOURNAL</i>	109
2.8.3	<i>Examinez votre table</i>	112
2.8.4	<i>Le mode DEBUG</i>	112
3.	Modifier les tables SAS	117
3.1	L'exécution par SAS de vos programmes.....	118
3.1.1	<i>La structuration du PDV</i>	119
3.1.2	<i>Mode de structuration du PDV et variables caractères</i>	121
3.1.3	<i>Détermination du type d'une variable créée par un programme</i>	122
3.1.4	<i>L'attribution des modalités</i>	124
3.1.5	<i>L'instruction OUTPUT</i>	126
3.1.6	<i>Le contrôle des tables de sortie avec l'instruction OUTPUT</i>	128
3.1.7	<i>L'instruction RETURN</i>	129
3.2	Les opérateurs et les fonctions usuelles de construction et de transformation de variables.....	129
3.2.1	<i>Les opérateurs</i>	130
3.2.2	<i>Les fonctions numériques usuelles</i>	131
3.2.3	<i>Les fonctions statistiques</i>	131
3.2.4	<i>Les générateurs de nombres aléatoires</i>	132
3.2.5	<i>Les fonctions propres aux dates et aux heures</i>	134
3.2.6	<i>Autres fonctions utiles</i>	137
3.2.7	<i>Les fonctions sur variables caractères</i>	138
3.2.8	<i>La concaténation de variables caractères</i>	144
3.2.9	<i>La création d'une fonction avec PROC FCMP</i>	146
3.3	Construire une table SAS « allégée ».....	148
3.3.1	<i>Effacer ou conserver uniquement certaines variables</i>	148
3.3.2	<i>Écarter ou ne conserver que certaines observations : IF, WHERE et SELECT</i> ...	149
3.3.3	<i>Sélectionner des observations avec l'option POINT=</i>	154
3.4	Construire de nouvelles variables au sein d'une table SAS.....	155
3.4.1	<i>La fonction INPUT</i>	155
3.4.2	<i>La construction sans condition</i>	156
3.4.3	<i>La construction sous conditions</i>	158

3.5	Les boucles DO	162
3.5.1	<i>Le DO simple</i>	163
3.5.2	<i>Le DO itératif</i>	163
3.5.3	<i>Les boucles DO UNTIL</i>	164
3.5.4	<i>Les boucles DO WHILE</i>	165
3.5.5	<i>Les boucles sur ARRAY</i>	165
3.6	Donner une mémoire à SAS	173
3.6.1	<i>L'instruction RETAIN</i>	173
3.6.2	<i>Les marqueurs FIRST, et LAST</i>	176
3.6.3	<i>Faire référence à des valeurs passées</i>	179
3.7	Trois procédures qui permettent de modifier vos tables	180
3.7.1	<i>PROC SORT</i>	180
3.7.2	<i>PROC RANK</i>	183
3.7.3	<i>PROC TRANSPOSE</i>	186
3.8	L'instruction MODIFY – 1 ^{re} partie	189
4.	Combiner les tables SAS	191
4.1	Concaténer des tables	192
4.1.1	<i>Ajouter des observations avec un SET impliquant deux tables</i>	192
4.1.2	<i>PROC APPEND</i>	195
4.2	Fusionner des tables : MERGE	196
4.2.1	<i>MERGE – Cas standard</i>	197
4.2.2	<i>Clés multiples</i>	198
4.2.3	<i>MERGE ONE TO MANY</i>	199
4.2.4	<i>Un outil pratique avec MERGE : les marqueurs IN</i>	200
4.2.5	<i>Quelques sources de problèmes avec MERGE</i>	201
4.3	Mettre à jour une table : UPDATE	205
4.4	L'instruction MODIFY – 2 ^{de} partie	207
4.4.1	<i>Mises à jour avec MODIFY et BY</i>	207
4.4.2	<i>La variable automatique _JRC_ et la macro %SYSRC</i>	208
4.5	Techniques avancées de fusion de tables	211
4.5.1	<i>Le pouvoir inattendu des FORMAT</i>	211
4.5.2	<i>Les tables de hachage</i>	213
5.	Faciliter la gestion des tables SAS	221
5.1	La procédure PROC CONTENTS	222
5.2	Les LABEL	224
5.2.1	<i>Les LABEL de variables</i>	224
5.2.2	<i>Les LABEL de tables</i>	226
5.3	Les FORMAT	226
5.3.1	<i>L'utilisation de FORMAT prédéfinis sur des variables numériques</i>	226
5.3.2	<i>Quelques FORMAT utiles sur variables numériques</i>	229
5.3.3	<i>Les FORMAT d'affichage des variables de dates et d'heures</i>	230
5.3.4	<i>Les FORMAT d'affichage des variables caractères</i>	235
5.3.5	<i>La création de FORMAT spécifiques</i>	235
5.3.6	<i>La fonction PUT</i>	243
5.4	La longueur des variables numériques	244
5.5	PROC DATASETS	245
5.5.1	<i>Principes</i>	246
5.5.2	<i>Une procédure différente</i>	247
5.5.3	<i>La notion de RUN GROUP</i>	248
5.5.4	<i>Les options de PROC DATASETS</i>	249
5.5.5	<i>L'instruction MODIFY</i>	250
5.5.6	<i>Les instructions de gestion des fichiers SAS</i>	253

5.6	La gestion des tables	254
5.6.1	<i>Quelques bonnes pratiques de gestion des tables.</i>	254
5.6.2	<i>Empêcher SAS de réécrire une table de même nom</i>	256
5.6.3	<i>Les tables en générations.</i>	257
5.7	La gestion des options globales.....	260
5.7.1	<i>Les options de SAS.</i>	260
5.7.2	<i>Les modifications de vos options globales pour une session</i>	263
5.7.3	<i>La modification du fichier de configuration SASV9.CFG.</i>	264
5.7.4	<i>AUTOEXEC.SAS</i>	265
6.	Maîtriser les procédures de base	267
6.1	Quelques éléments communs	267
6.1.1	<i>Rappel : quatre instructions communes</i>	268
6.1.2	<i>TITLE et FOOTNOTE</i>	268
6.1.3	<i>Les mots clés des statistiques descriptives.</i>	269
6.1.4	<i>Les tables CROSS et FOURNISSEUR13</i>	270
6.1.5	<i>L'utilisation des tables nécessaires à la reproduction des exemples</i>	271
6.2	PROC MEANS.....	271
6.2.1	<i>Éléments de syntaxe</i>	271
6.2.2	<i>Applications simples</i>	272
6.2.3	<i>Les tables de résultats.</i>	273
6.2.4	<i>Les variables de classification</i>	276
6.3	PROC PRINT.....	278
6.3.1	<i>Éléments de syntaxe</i>	278
6.3.2	<i>Exemples d'application</i>	279
6.3.3	<i>PROC PRINT et ODS.</i>	283
6.4	PROC REPORT.....	287
6.4.1	<i>Exemples simples</i>	287
6.4.2	<i>L'instruction DEFINE</i>	288
6.4.3	<i>Les variables et leurs usages possibles</i>	289
6.4.4	<i>Les rapports intégrant des informations de type liste et des informations de synthèse</i>	295
6.4.5	<i>Quelques références pour aller plus loin avec PROC REPORT</i>	297
6.4.6	<i>PROC REPORT et ODS.</i>	298
6.5	PROC FREQ.....	299
6.5.1	<i>Éléments de syntaxe</i>	299
6.5.2	<i>Exemples d'application</i>	300
6.5.3	<i>Les tableaux croisés.</i>	300
6.5.4	<i>Les sorties graphiques ODS Statistical Graphics de PROC FREQ</i>	303
6.5.5	<i>Mise en forme des tableaux et tables de résultats</i>	304
6.6	PROC TABULATE	306
6.6.1	<i>Éléments de syntaxe</i>	306
6.6.2	<i>La table FACTURES.</i>	307
6.6.3	<i>Structures lignes-colonnes des tableaux</i>	307
6.6.4	<i>Les variables d'analyse.</i>	310
6.6.5	<i>Choix de la statistique éditée dans le tableau.</i>	312
6.6.6	<i>L'introduction de récapitulatifs dans les tableaux</i>	313
6.6.7	<i>L'introduction de pourcentages dans les tableaux</i>	317
6.6.8	<i>PROC TABULATE et ODS.</i>	320

7. Produire des graphiques	325
7.1 Les procédures ODS Graphics	325
7.1.1 <i>Principes</i>	326
7.1.2 <i>ODS Statistical Graphics</i>	327
7.1.3 <i>Une première approche de PROC SGPLOT</i>	329
7.1.4 <i>Les exercices</i>	333
7.2 Les outils de représentation des variables discrètes.....	334
7.3 Les outils de représentation des distributions de variables continues.....	342
7.3.1 <i>HBOX/VBOX</i>	342
7.3.2 <i>HISTOGRAM et DENSITY</i>	344
7.4 Les outils de représentation des relations entre variables	346
7.4.1 <i>Les instructions basiques</i>	346
7.4.2 <i>BAND</i>	350
7.4.3 <i>Les instructions représentatives d'ajustement</i>	351
7.5 La procédure PROC SGPANEL.....	353
7.6 La personnalisation des graphiques produits par les procédures Statistical Graphics ..	356
7.7 La création de graphiques avec ODS Graphics Designer	358
8. Utiliser ODS	361
8.1 ODS : les principes.....	362
8.1.1 <i>Qu'est-ce qu'une sortie SAS ?</i>	362
8.1.2 <i>Les destinations ODS</i>	362
8.1.3 <i>Ouvrir et fermer une destination ODS</i>	364
8.1.4 <i>Les options et actions des destinations ODS externes à SAS</i>	365
8.1.5 <i>Les styles ODS</i>	367
8.2 ODS OUTPUT et ODS DOCUMENT	368
8.2.1 <i>ODS OUTPUT</i>	368
8.2.2 <i>ODS DOCUMENT</i>	371
8.3 ODS HTML.....	373
8.3.1 <i>La gestion des fichiers des destinations externes</i>	373
8.3.2 <i>ODS HTML – Principes</i>	373
8.3.3 <i>Introduction de tables des matières</i>	375
8.3.4 <i>Paramétrer la table des matières</i>	376
8.3.5 <i>Personnaliser les feuilles de résultats</i>	378
8.3.6 <i>Personnaliser les sorties avec ODS SELECT et EXCLUDE</i>	384
8.3.7 <i>Destinations multiples des sorties</i>	386
8.3.8 <i>ODS HTML5</i>	387
8.4 ODS PDF	388
8.4.1 <i>Principes</i>	388
8.4.2 <i>La table des signets</i>	389
8.4.3 <i>Les outils de mise en pages</i>	390
8.4.4 <i>ODS LAYOUT et ODS REGION</i>	392
8.5 ODS RTF	395
8.6 ODS POWERPOINT	397
8.7 ODS MARKUP	401
8.7.1 <i>Les tagsets HTML</i>	401
8.7.2 <i>Créer des fichiers CSV avec ODS MARKUP</i>	402
8.7.3 <i>Le tagset EXCELXP – Envoyer vos sorties directement dans Microsoft EXCEL</i> ..	404
8.8 PROC TEMPLATE	410
8.8.1 <i>Qu'est-ce qu'un style ODS ?</i>	411
8.8.2 <i>Construire un style ODS – Les outils</i>	413

9. Découvrir PROC SQL	419
9.1 Introduction.....	419
9.1.1 <i>Qu'est ce que SQL ?</i>	419
9.1.2 <i>Que fait-on avec PROC SQL et avec quel type de données ?</i>	420
9.1.3 <i>Un langage adapté aux bases de données relationnelles</i>	421
9.1.4 <i>Pourquoi apprendre PROC SQL ?</i>	422
9.1.5 <i>Terminologie</i>	424
9.2 Générer des rapports simples.....	426
9.2.1 <i>L'instruction SELECT</i>	426
9.2.2 <i>La clause SELECT</i>	427
9.2.3 <i>La clause FROM</i>	431
9.2.4 <i>La clause WHERE</i>	433
9.2.5 <i>La clause ORDER BY</i>	436
9.3 Réaliser des rapports de synthèse.....	437
9.3.1 <i>Le mot clé DISTINCT</i>	437
9.3.2 <i>Les fonctions qui commandent une synthèse</i>	437
9.3.3 <i>Le paramétrage de votre sortie au moyen de la clause HAVING</i>	442
9.4 Les sous-requêtes.....	443
9.4.1 <i>Les différents types de sous-requêtes</i>	443
9.4.2 <i>Les sous-requêtes renvoyant plusieurs valeurs</i>	445
9.5 Les jointures horizontales.....	449
9.5.1 <i>Les jointures CROSS – La production de produits cartésiens</i>	450
9.5.2 <i>Les jointures INNER et OUTER</i>	451
9.5.3 <i>Jointures avec clé répétée</i>	454
9.5.4 <i>Fusionner plus de deux tables au moyen de PROC SQL</i>	456
9.5.5 <i>Les jointures NATURAL</i>	457
9.5.6 <i>Les jointures réflexives</i>	458
9.5.7 <i>Conditions des jointures</i>	460
9.5.8 <i>L'utilisation de vues en ligne</i>	461
9.6 Les jointures verticales.....	463
9.6.1 <i>Les jointures EXCEPT</i>	464
9.6.2 <i>Les jointures INTERSECT</i>	467
9.6.3 <i>Les jointures UNION</i>	469
9.6.4 <i>Les jointures OUTER UNION</i>	470
9.7 Introduction aux vues.....	471
10. Programmer sous SAS avec le langage macro	475
10.1 Les concepts du langage macro.....	476
10.2 Les macro-variables.....	479
10.2.1 <i>Principes des macro-variables</i>	479
10.2.2 <i>Accéder à la valeur d'une macro-variable</i>	480
10.2.3 <i>Les macro-variables automatiques</i>	482
10.2.4 <i>Supprimer des macro-variables</i>	483
10.3 Macro-variables et langage SAS.....	484
10.3.1 <i>Les routines CALL SYMPUT et CALL SYMPUTX</i>	484
10.3.2 <i>La création de macro-variables au moyen de PROC SQL</i>	488
10.3.3 <i>Utiliser vos macro-variables dans une étape DATA</i>	492
10.4 Les macro-fonctions.....	494
10.4.1 <i>La macro-fonction %SYSFUNC</i>	495
10.4.2 <i>Les macro-fonctions permettant l'évaluation</i>	495
10.4.3 <i>Les macro-fonctions de quoting</i>	496
10.4.4 <i>Les outils de manipulation des valeurs de macro-variables</i>	499

10.5	Premiers exemples sans macro-programme	501
10.5.1	<i>Construire des rapports pré-paramétrés avec le macro-langage</i>	502
10.5.2	<i>Économiser vos ressources grâce au macro-langage</i>	503
10.6	Les macro-programmes	505
10.6.1	<i>Principes des macro-programmes</i>	505
10.6.2	<i>Construire ses macro-programmes</i>	506
10.6.3	<i>Environnement local/environnement global</i>	508
10.6.4	<i>Le paramétrage des macro-programmes</i>	509
10.6.5	<i>Les macro-instructions propres aux macro-programmes</i>	511
10.6.6	<i>Exercice récapitulatif</i>	514
10.7	Conserver ses macro-programmes	516
10.7.1	<i>La sauvegarde des macro-programmes dans des fichiers externes</i>	517
10.7.2	<i>L'enregistrement des macro-programmes dans un fichier CATALOG</i>	517
10.7.3	<i>AUTOCALL</i>	518
10.7.4	<i>Les macro-programmes compilés</i>	519
11.	Un cas pratique	521
11.1	Énoncé du problème	521
11.2	Les données	524
11.3	Construire une méta-table	525
11.4	Produire le graphique	529
	Annexe	537
	L'aide SAS	537
	SAS : une communauté active d'utilisateurs	540
	Bibliographie	541
	Index	543
	Liste des messages	550



Préface

Avant de rédiger cette préface, comme des millions de personnes dans le monde désormais, j'ai consacré un certain temps à lire d'innombrables messages électroniques, et autres informations plus ou moins intéressantes, sur les divers réseaux sociaux auxquels je participe, je l'avoue, plus ou moins activement.

Dans cet exercice quotidien, ma longue pratique de la veille et de l'intelligence économique, résolument tournée vers l'action, m'amène, le plus souvent, à hiérarchiser les sujets en fonction de mes intérêts du moment et, compte tenu de ma volonté affirmée de ne pas faire preuve de procrastination pour cette préface, mon attention fut très vite attirée par le titre d'un article : « Data scientist' will be the hottest job of 21st century – but teachers don't know. »¹

Ce dernier exprimait la crainte, parfois justifiée, d'un décalage entre les besoins en formation de cette nouvelle génération de professionnels des solutions analytiques et la réalité des dispositifs d'enseignement. En insistant plus particulièrement sur l'inertie des enseignants en la matière.

Mon point de vue sur le sujet est plus nuancé.

Certes, le sujet précis de l'utilisation extensive des données massives promise par l'irruption des problématiques « *Big Data* » dans les entreprises amène à se poser la question des compétences requises. Quelques rares acteurs en pointe, auxquels SAS a l'insigne honneur d'appartenir, contribuent par anticipation à lever ces incertitudes. Dans cet esprit, l'entreprise organise depuis deux ans un programme immersif de formation (SAS® Spring Campus²) donnant gratuitement, à un nombre limité de jeunes étudiants arrivant sur le marché du travail, les clés pour un démarrage plus rapide dans cet univers des *Big Data*.

Pour autant, ce programme, déjà très apprécié, est aussi le reflet de notre engagement constant depuis plus de 15 ans auprès du monde académique. Cette troisième édition de l'ouvrage que vous avez entre les mains montre à quel point certains de nos enseignants n'ont pas attendu la vague déferlante des *Big Data* pour se préoccuper des connaissances fondamentales requises pour donner du sens aux données.

Sébastien Ringuedé fait partie de ces précurseurs. Dans le cadre du master Économétrie et Statistique appliquée de l'université d'Orléans, dirigé par Gilbert Colletaz, Sébastien a pu concrètement apporter aux étudiants le niveau d'expertise requis par les entreprises en matière de traitement et d'analyse de l'information. À cet effet, depuis plus de 15 ans, tout en utilisant les solutions SAS pour ses recherches en sciences économiques, il prône leur pratique intensive, permettant ainsi aux étudiants d'acquérir aisément les certifications professionnelles attestant de la qualité des compétences acquises. Cet ouvrage est le fruit de sa longue expérience.

Mouloud DEY
Directeur Business solutions et marchés émergents
Mouloud.Dey@sas.com
SAS France

1. <http://siliconrepublic.com/careers/item/35525-data-scientist-will-be/>

2. http://www.sas.com/offices/europe/france/academic/pdf/SAS_Spring_Campus.pdf



Introduction

Présentation de l'ouvrage

Cet ouvrage s'adresse à tout utilisateur des progiciels SAS devant recourir à la programmation Base SAS. S'il a été conçu avant tout à destination des utilisateurs novices, il permettra aux utilisateurs de niveau intermédiaire et expert d'approfondir et d'affiner leurs connaissances sur cet outil particulièrement complet qu'est le système SAS. Le présent ouvrage peut être consulté aussi bien par des professionnels que par des étudiants en statistiques, économie, informatique, mathématiques appliquées, biostatistiques, pharmacie, gestion, marketing, des étudiants préparant par exemple un DUT STID, un master, un diplôme de MIAGE ou encore des étudiants en écoles d'ingénieur et de commerce.

L'objectif de ce livre est de présenter le socle de connaissances fondamentales et transversales nécessaires à la programmation SAS, et commun à tous les utilisateurs quel que soit leur usage de SAS. Il offre une présentation des outils que les utilisateurs de SAS peuvent mobiliser pour construire, modifier, gérer, explorer et présenter leurs données.

Pour cette nouvelle édition, nous avons conservé une organisation très proche de celle adoptée dans les première et seconde éditions. Les chapitres 1 et 2 sont consacrés à la phase de construction des tables SAS. Il s'agit de deux chapitres essentiels puisque nous présentons l'ensemble des outils que vous devrez exploiter pour construire des tables SAS à l'image exacte de vos données. Cependant, même placées dans une table SAS, vos données n'auront jamais directement une forme qui permette leur exploitation : vous aurez très souvent besoin de construire de nouvelles variables ou de sélectionner des observations présentant certaines caractéristiques. Le chapitre 3, après avoir détaillé les caractéristiques du fonctionnement de SAS qui conditionne votre programmation, vous présente les outils (opérateurs, fonctions, instructions et procédures) qui vous permettront de construire les tables contenant les variables requises par votre analyse. Nous examinerons ensuite, dans le chapitre 4, les outils mis à disposition par le langage Base SAS pour combiner vos tables. Il s'agit d'une phase importante, car il vous faudra très souvent mobiliser plusieurs sources pour construire la table dont vous avez besoin.

Le chapitre 5 est le dernier chapitre consacré exclusivement aux tables SAS et à l'étape DATA, étape fondamentale du langage SAS. Nous y détaillons les très nombreux outils que SAS propose pour gérer vos tables. Ces outils concernent aussi bien l'organisation de votre espace de travail avec SAS que la présentation de vos tables, des variables et de leurs modalités. Ils sont indispensables à un travail efficace avec SAS.

Les tables SAS ont bien entendu vocation à être exploitées : puisque notre objet est la présentation des connaissances transverses à l'ensemble des utilisateurs SAS, nous présenterons dans cet ouvrage les procédures de primo-exploitation des tables et les outils qui permettent de retransmettre vos résultats. À ce titre, nous examinons dans le chapitre 6 les procédures qui vous seront nécessaires pour explorer, connaître et comprendre vos données. On ne peut en effet analyser ou présenter des données que si on les connaît bien. Ces outils d'exploration des tables vous seront utiles dans la phase de retransmission de vos résultats : même si votre finalité première est l'analyse économétrique ou statistique, vous aurez très souvent besoin de présenter au moyen de tableaux les données qui servent de base à votre analyse.

Nous traitons dans le chapitre 7 les procédures graphiques ODS Graphics qui peuvent, ici encore, être utiles lors des phases d'exploration et de restitution de vos résultats.

L'*Output Delivery System* (ODS) livré avec SAS est étudié dans le chapitre 8. Grâce à cette fonctionnalité de SAS, vous avez la possibilité de produire en quelques instructions des documents attrayants aux formats HTML, PDF et RTE. C'est aussi *via* ODS que vous pourrez créer des fichiers Excel présentant vos résultats obtenus sous SAS, et les partager avec des personnes ne disposant pas de SAS.

Le chapitre 9 vous propose ensuite une introduction au langage SQL et détaille la procédure PROC SQL que beaucoup d'utilisateurs de SAS mobilisent pour explorer les données contenues dans les bases de données relationnelles.

Le langage macro n'est traité qu'au chapitre 10 car, pour percevoir son intérêt, il faut avoir une certaine expérience de SAS et être conscient que votre travail comprend des phases répétitives ou qu'il suffit de modifier quelques instructions pour qu'un programme dédié au traitement d'une table spécifique puisse aussi analyser une autre table.

Nous présentons dans le chapitre 11 un cas pratique mobilisant les connaissances acquises au cours des dix premiers chapitres.

Les nouveautés de la présente édition

Cette nouvelle édition est le fruit de discussions que nous avons pu avoir aussi bien avec SAS France qu'avec des étudiants, des enseignants ou des utilisateurs professionnels de SAS, lecteurs des deux premières éditions, et nous espérons qu'elle comblera leurs attentes. Il n'est pas possible de répertorier l'ensemble des nouveautés de cette édition, mais citons-en quelques-unes :

- encore plus d'exercices : vous y trouverez plus de 150 exercices contre une centaine dans la deuxième édition – seul le chapitre 8, consacré à ODS, ne propose pas d'exercice ;
- une refonte importante du chapitre 4, consacré aux fusions de tables, et l'introduction des fusions de tables au moyen des tables de hachage ;
- une mise à jour importante du chapitre 7, consacré aux procédures ODS Graphics, et l'introduction dans ce chapitre de la procédure PROC SG PANEL ;
- une introduction, dans le chapitre consacré à ODS, à deux nouvelles destinations proposées par SAS 9.4 : ODS HTML5 et ODS POWERPOINT ;
- un nouveau chapitre consacré à PROC SQL.

Si peu de différences apparaissent entre le plan de la présente édition et celui de la deuxième édition, de nombreuses sections de la présente édition ont fait l'objet d'une réécriture et d'une réorganisation. De nouveaux exemples ont très souvent été introduits.

Progresser dans sa connaissance de SAS

Cet ouvrage ne saurait prétendre à l'exhaustivité mais il aborde les problèmes les plus courants et vous donne les outils qui vous aideront à les résoudre. Si les bases de chaque thème sont traitées le plus complètement possible, des extensions vous seront toujours proposées, et vous trouverez tout au long de ce livre de très nombreux liens vers des documents vous permettant d'approfondir vos connaissances sur des aspects plus avancés. Nous y faisons de très nombreuses références à l'aide SAS, et précisons en annexe comment utiliser les entrées de l'aide citée dans l'ouvrage.

Chaque chapitre commence par expliquer les concepts de base avant d'aborder des notions plus complexes. Ainsi, dans l'exposé d'une procédure, il y aura peu de rapport entre les résultats simples que nous proposerons dans un premier temps et les derniers résultats. Ces derniers résultats ne peuvent cependant pas être compris sans les premières parties dans lesquelles nous exposons les principaux concepts avant de les mettre en pratique.

C'est pour cette même raison que les chapitres consacrés à SQL et au langage macro sont présentés à la fin de l'ouvrage : ils mobilisent en effet l'ensemble des connaissances abordées dans les huit premiers chapitres.

Nous traiterons peu les interfaces proposées par SAS qui permettent en quelques clics de réaliser un graphique, d'importer des données, de créer des tableaux, etc. Ces interfaces, certes conviviales, ont pour principale fonction d'écrire le programme à votre place, mais elles ne vous dispensent absolument pas de connaître les outils de la programmation SAS. En effet, ces interfaces ne peuvent exploiter 100 % des outils de programmation offerts par les procédures qu'elles émulent. Si votre demande s'éloigne des cas standard et si vous n'avez aucune connaissance des outils de la programmation SAS, vous ne pourrez pas comprendre le programme créé par votre interface et le modifier de manière qu'il produise le résultat souhaité. C'est par une compréhension fine de la programmation SAS que vous pourrez appréhender de manière rapide n'importe quelle interface proposée par SAS.

La Certification Base Programming for SAS 9

Cet ouvrage couvre l'intégralité du programme de la Certification Base Programming for SAS 9. Il va même au-delà, en approfondissant certains domaines et en explorant d'autres qui ne sont pas traités par la certification Base programming for SAS 9 – comme la production de graphiques, PROC SQL et le langage macro.

Il nous a semblé important que notre ouvrage embrasse à nouveau le programme de cette certification. Ce programme, établi par SAS en lien avec le monde professionnel, constitue le socle de connaissances indispensables que tout utilisateur de SAS doit maîtriser pour utiliser de manière efficace ce système très complet qu'est SAS.

Singulièrement difficile à obtenir parce qu'elle nécessite une réelle compréhension du fonctionnement de Base SAS, cette certification est reconnue sur le plan international et est particulièrement appréciée des recruteurs. Elle témoigne d'une compétence rare et constitue un vrai « plus » dans un *curriculum vitae*.

L'épreuve, d'une durée de 110 minutes, prend la forme d'un QCM de 64 questions en anglais sur ordinateur. Pour chaque question, 4 réponses sont proposées. Si vous avez au moins 45 bonnes réponses, vous obtenez la certification. Bien entendu, pendant le temps de l'épreuve, vous ne disposez pas de SAS.

Si vous êtes dans le monde professionnel, vous pouvez passer cette certification dans un des centres de tests Pearson VUE. Le centre de formation SAS propose quant à lui une offre associant une journée de préparation à la certification, un manuel d'entraînement et l'examen :

www.sas.com/offices/europe/france/services/training/certification/

Si vous êtes étudiant, le site compagnon de cet ouvrage, en association avec SAS France et son département académique, peut vous aider à passer cette certification :

www.sas.com/offices/europe/france/academic/

Si votre formation est partenaire SAS et si elle présente un nombre suffisant de candidats, une session de certification peut être organisée dans votre établissement et un tarif spécial vous est proposé.

Toujours si vous êtes étudiant, vous pouvez aussi passer cet examen de certification dans le centre de formation SAS de Vincennes. Ici aussi, des conditions financières particulières peuvent vous être proposées.

Dans les deux cas, et uniquement si vous êtes étudiant, vous aurez la possibilité, une dizaine de jours avant le passage de cet examen, d'accéder à un des **maxi-quiz** du site compagnon de cet ouvrage.

Ces maxi-quiz vous permettent d'évaluer votre niveau de connaissance et votre degré de préparation à l'examen. Un maxi-quiz comprend 70 questions en français extraites d'une banque de questions couvrant l'intégralité du programme de cette certification. Vous disposez de deux heures pour répondre à ces questions. Vous n'avez besoin que d'un ordinateur et d'une connexion à Internet pour participer à un maxi-quiz.

Attention, ces quiz sont des outils d'évaluation de vos connaissances et ne doivent pas être considérés comme des outils de formation. Vous ne disposerez pas des réponses aux questions, mais un récapitulatif par grands thèmes vous sera proposé : il vous permettra d'orienter vos dernières révisions et d'optimiser ainsi vos chances d'obtenir cette certification.

L'accès à cette application est soumis à deux conditions : vous devez être étudiant et inscrit à un examen de certification pour accéder à l'un des maxi-quiz.

Depuis sa mise en place en avril 2009, **plus de 300 étudiants** ont obtenu cette certification après être passés par un maxi-quiz : alors, pourquoi pas vous ?

Pour plus de renseignements sur la Certification Base Programming for SAS 9 et les conditions d'accès aux maxi-quiz, vous pouvez consulter la section **Certification** du site compagnon de cet ouvrage. Vous trouverez aussi dans cette section certification un premier mini-quiz de dix questions.

Les pictogrammes

Afin de vous aider dans votre apprentissage de SAS, nous avons défini des pictogrammes particuliers que vous rencontrerez dans le sommaire de chaque chapitre et en marge des intitulés de section :

- © le point évoqué dans la section est au programme de la Certification Base Programming for SAS 9.
- ⊕ vous pouvez aborder ce point au moment où vous le rencontrez. Soit il constitue un thème qui, à notre avis, fait partie des connaissances de base que doit avoir un utilisateur de SAS ; soit il s'agit d'un premier niveau d'approfondissement d'une connaissance de base tout à fait abordable si vous avez compris ce qui précède.
- ⊕² le point abordé dans la section constitue un approfondissement important. Il nécessite, pour être bien compris, la maîtrise de l'ensemble des connaissances © et ⊕ proposées dans cet ouvrage.

Ces approfondissements ne sont pas forcément traités de manière complète, mais vous trouverez toujours des références de documents qui doivent vous permettre d'aller plus loin si le sujet vous intéresse particulièrement.

Dans les sommaires des chapitres, seul le pictogramme © est repris. Il vous indique que la section contient des éléments au programme de la certification SAS Base programming for SAS 9. Dans le chapitre lui-même, si une section principale est marquée d'un pictogramme, toutes les sous-sections sont concernées par celui-ci. Si la section ne présente pas de pictogramme, c'est que les sous-sections sont de divers niveaux : chaque sous-section sera alors associée à un pictogramme. La figure 0-1 reprend et illustre notre propos.

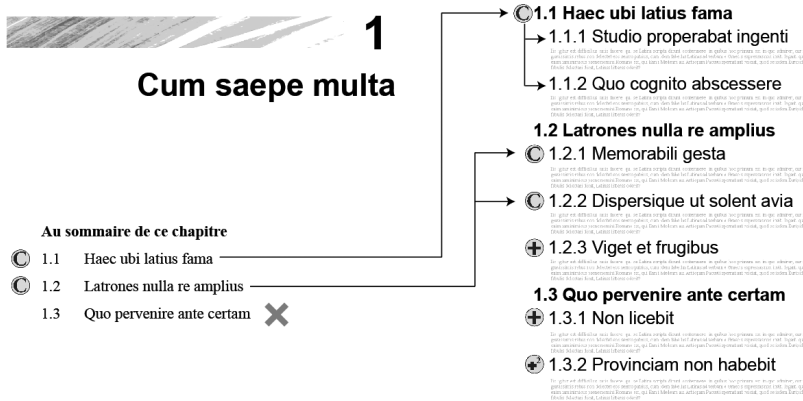


Figure 0-1 • Pictogrammes dans les sommaires de chapitres et dans le corps du texte.

Comment travailler avec cet ouvrage

Le site compagnon de cet ouvrage est le complément indispensable à votre travail d'apprentissage de SAS. Vous retrouverez ce site à l'adresse suivante : www.sas-sr.com

Ce site vous offre :

- un forum de discussions grâce auquel vous pouvez dialoguer directement avec l'auteur ;
- tous les programmes proposés dans l'ouvrage ;
- toutes les sorties générées par ces programmes, y compris celles qui ne sont pas reproduites dans cet ouvrage ;
- tous les fichiers nécessaires à la reproduction des exemples et aux exercices proposés ;
- les solutions aux exercices proposés ;
- des compléments à cet ouvrage ;
- l'ensemble des liens vers des sites Internet proposés dans cet ouvrage ;
- les liens vers l'aide de SAS en ligne pour toutes les entrées citées dans cet ouvrage ;
- l'ensemble des liens vers les documents cités dans l'ouvrage et qui doivent vous permettre d'approfondir certains points.

Tout au long des chapitres, des exercices vous sont soumis. N'attendez pas d'avoir lu l'intégralité d'un chapitre pour les effectuer : ils sont indispensables à la compréhension des notions abordées et constituent généralement des passerelles entre les différents concepts.

Ne consultez les solutions qu'au terme d'une réelle réflexion ! C'est en commettant des erreurs que vous progresserez et que vous affinerez vos techniques de programmation.

Nous vous invitons à reproduire l'intégralité des nombreux exemples de programmation afin de bien comprendre l'impact de chacune des instructions et options utilisées. N'hésitez pas à en retirer certaines et à en introduire d'autres pour bien saisir leur impact.

Même si vous connaissez déjà un peu SAS, commencez votre lecture par le chapitre 1 et consacrez le temps nécessaire à la compréhension de chaque concept exposé. La profonde compréhension des concepts est indispensable à une maîtrise de SAS. Si vous ne comprenez pas les concepts, vous ne pourrez rien apprendre d'un programme offrant plus de cinq instructions. Vous maîtriserez les concepts lorsque, avant même d'exécuter un programme, vous aurez une idée précise de la structure du résultat produit.

Disposer du système SAS

Si vous êtes actuellement dans une université ou une école, il se peut que la formation que vous suivez soit inscrite au programme académique de SAS¹. L'inscription à ce programme permet à votre formation de bénéficier d'un certain nombre de licences SAS à un coût tout à fait intéressant ; elle autorise l'installation sur votre ordinateur personnel de SAS pendant la durée de votre formation initiale (y compris pendant la durée de votre stage), contre une inscription au Club Academic de SAS et la signature d'un contrat de licence. Pour tous renseignements sur le programme académique de SAS, vous pouvez consulter le site suivant :

www.sas.com/offices/europe/france/academic/

Les différentes versions de SAS

Cet ouvrage a été préparé avec les versions 9.2, 9.3 et 9.4 de SAS. Le passage d'une version à une autre implique toujours des ajouts importants. Ceux qui concernent cet ouvrage sont cependant peu nombreux et seront signalés. Parmi les programmes que nous proposons, 90 % peuvent être exécutés sur les versions plus anciennes de SAS (8.2 et 9.1.3).

Le chapitre 7 est le plus dépendant de la version de SAS à votre disposition. Les utilisateurs de versions antérieures à SAS 9.2 ne pourront exécuter aucun des programmes proposés dans ce chapitre. Les procédures ODS Graphics (appelées procédures Statistical Graphics dans SAS 9.2) traitées par le chapitre 7 ne sont en effet livrées avec SAS que depuis la version 9.2. Les programmes proposés dans ce chapitre ont pour objet de vous exposer les outils à votre disposition pour produire des graphiques avec SAS 9.4. Vous trouverez sur le site compagnon de cet ouvrage des programmes adaptés aux versions SAS 9.2 et 9.3.

Quelques programmes du chapitre 8 ne peuvent être exécutés avec SAS 9.1.3 ou SAS 8.2. Les destinations ODS HTML5 et POWERPOINT ne sont proposées que depuis SAS 9.4. Quant au cas pratique exposé au chapitre 11, il ne peut être exécuté qu'avec au minimum SAS 9.2.

Le système SAS est conçu de telle manière qu'un programme qui fonctionne dans une version donnée fonctionnera aussi dans les versions ultérieures de SAS. Certaines fonctions anciennes, remplacées depuis par des fonctions plus puissantes, même si elles ne sont plus documentées, continuent et continueront à pouvoir être utilisées. Le passage d'une version à une autre n'est donc pas synonyme d'un complet réapprentissage de SAS : ce que vous apprenez aujourd'hui sera encore valable demain.

SAS est loué par votre employeur, votre école ou votre université : il n'y a pas de coût supplémentaire demandé au client SAS lorsqu'une nouvelle version apparaît. Seules les contraintes propres aux clients² peuvent expliquer le maintien d'une version relativement ancienne de SAS. Vous comprenez donc que différentes versions de SAS peuvent coexister à un moment donné.

SAS 9.2 a commencé à être déployé et distribué au printemps 2008. Son déploiement complet s'est terminé avec la livraison de la révision 3 (SAS 9.2 TS2M3) en 2010. La distribution de SAS 9.3 a débuté en juillet 2012 et celle de SAS 9.4 en juillet 2013.

1. Environ 150 établissements et 300 diplômés sont actuellement partenaire SAS au travers du programme académique de SAS.
2. Notamment en matière de tests à la suite de la migration des données entre différentes versions.

SAS Windows, Unix, z/OS

Cet ouvrage traitera exclusivement de SAS sous environnement Microsoft Windows Professionnel (de XP à Windows 8)¹. En effet, si vous débutez sous SAS, il y a de très fortes chances pour que la version de SAS à votre disposition soit une version Microsoft Windows. Si, par la suite, vous êtes appelé à travailler dans un autre environnement que Windows, vous découvrirez que les différences entre votre SAS Windows et SAS dans un autre environnement résident dans :

- la façon de démarrer une session SAS ;
- l'adressage des fichiers externes à SAS ;
- l'existence de quelques options spécifiques à votre environnement ;
- la présentation différente des fenêtres du progiciel.

Les programmes écrits au moyen de SAS Windows fonctionneront parfaitement sur d'autres systèmes d'exploitation.

Conventions typographiques

Vous pouvez employer minuscules et majuscules dans les programmes (instructions, options, noms de variables, noms de tables, etc.) que vous écrivez dans SAS. Seules les valeurs prises par les modalités des variables alphanumériques doivent respecter la casse. Nous avons cependant utilisé dans cet ouvrage les majuscules et les minuscules afin de vous aider au mieux à comprendre les nombreux programmes proposés.

Tous les mots clés propres au langage SAS apparaissent en majuscules dans les programmes et dans le corps du texte. Les mots écrits en gras dans les programmes sont les instructions qui commencent et terminent une étape (ici, DATA et RUN). Les constantes numériques peuvent aussi apparaître en gras. À l'intérieur du programme, les instructions, fonctions et options propres au langage SAS apparaissent en majuscules (SET, IF, THEN, DO, RANUNI, etc.). Les FORMAT et les INFORMAT sont aussi présentés en majuscules. Les noms des tables et des variables dans un programme sont systématiquement écrits en minuscules.

```
DATA test2;
  SET test;
  FORMAT date DDMMYY.;
  ATTRIB x FORMAT=8.4
           LABEL="Variable analysée";
  IF x<15 THEN DO;
    x7=RANUNI(234);
    x9=x2+x12;
  END;
RUN;
```

Dans le corps du texte, les noms des instructions, fonctions et options, comme ceux des variables, des tables, des FORMAT, des INFORMAT, etc., seront systématiquement repris en majuscules.

1. Veuillez noter que les versions de SAS sont conçues pour **un fonctionnement optimal sous les éditions professionnelles** du système d'exploitation Windows. Il n'est pas conseillé d'installer SAS sur des ordinateurs disposant d'une édition familiale de Windows.

Les programmes seront indentés afin de vous permettre de mieux suivre et de mieux comprendre leur finalité. Il y a toujours au moins un niveau d'indentation, parfois plusieurs. Dans le cas du programme présenté ici, pour l'instruction ATTRIB, nous indentons une seconde fois pour signifier au lecteur que l'attribut LABEL est un élément de l'instruction ATTRIB. Nous indentons une seconde fois après l'instruction IF pour indiquer au lecteur que les deux instructions permettant de définir les variables X7 et X9 ne s'appliquent pas à toutes les observations (cette construction est en fait conditionnée au respect de la condition énoncée après l'instruction IF). L'instruction CARDS et les données qui la suivent ne seront pas indentées.

Il n'est pas forcément nécessaire d'organiser les programmes de cette manière, mais au fur et à mesure que les programmes que vous rédigerez s'allongeront, vous constaterez qu'une telle organisation facilite grandement la compréhension.

À l'exception de certains programmes très courts, les programmes proposés dans l'ouvrage sont numérotés, ce qui vous permet de les retrouver facilement sur le site compagnon.

Un peu d'histoire

Le logiciel SAS est le fruit du travail d'un groupe de chercheurs de la North Carolina State University qui a commencé en 1966. En 1967, James H. Goodnight, actuel CEO de SAS Institute, rejoint le projet et travaille à la stabilité du logiciel et à la création des procédures. À partir de 1971, le logiciel connaît un grand succès aussi bien auprès de la communauté universitaire qu'auprès des industries pharmaceutique et agricole, en raison notamment de sa capacité à gérer d'une part les données manquantes, d'autre part des tables de très grandes dimensions.

En 1976, l'entreprise SAS Institute est créée notamment par James H. Goodnight et John P. Sall. SAS s'est depuis imposé comme le leader mondial de l'informatique décisionnelle et des solutions de *business analytics*. Il est aujourd'hui présent dans 91 des 100 premières entreprises du classement Fortune Global 500® et sur près de 70 000 sites clients répartis dans 139 pays. SAS Institute a réalisé un chiffre d'affaires de 3,02 milliards de dollars en 2013, dont 25 % ont été réinvestis en recherche et développement.

SAS signifiait à l'origine *Statistical Analysis System*. La signification de l'acronyme a depuis été perdue, marquant ainsi le fait que SAS couvre toute la chaîne du traitement de l'information, du traitement des données jusqu'à la restitution métier, l'analytique étant au cœur de chaque métier.

Les personnes intéressées par l'histoire de SAS pourront consulter la page suivante :

www.sas.com/en_us/company-information.html#history

Le module Base SAS

Le système SAS est composé de différents modules ayant tous une fonctionnalité propre. Le module Base SAS constitue le cœur de votre système SAS – il est nécessaire au fonctionnement des autres produits SAS et c'est ce module que nous étudions dans le présent ouvrage.

Il permet l'exécution des étapes DATA, du langage macro, de PROC SQL, des outils offerts par ODS (*Output Delivery System*), ainsi que des analyses statistiques simples, ou des procédures utilitaires comme PROC CONTENTS, PROC PRINT ou PROC SORT. Depuis SAS 9.3, Base SAS intègre aussi les outils de production de graphiques offerts par les procédures ODS Graphics. Dans SAS 9.2, ces procédures faisaient partie du module SAS/GRAPH.

Le langage SAS est un langage dit de quatrième génération (L4G). Il s'agit, comme nous allons le découvrir au travers de ces pages, d'un langage peu technique, très proche de la syntaxe naturelle.

Une présentation succincte d'autres modules de SAS

Nous n'exposons ci-après que les modules principaux du système SAS – il en existe bien d'autres. Vous trouverez des descriptifs complets des produits et solutions proposés par le système SAS sur le site de SAS Institute :

www.sas.com/en_us/software/all-products.html

Un certain nombre de modules se caractérise par une mise à disposition de procédures spécifiques. La présence du signe «/» indique qu'il s'agit d'un module qui dispose de sa propre programmation. Ces modules peuvent être embarqués dans des interfaces utilisateurs spécifiques émulant du code SAS.

Un premier ensemble est dédié à l'analytique. Il couvre bien plus que la statistique et comprend, entre autres, les modules suivants :

SAS/STAT : outils dédiés à l'analyse statistique (inférentielle, descriptive) et aux calculs probabilistes comme l'analyse de variance, les modèles mixtes, les différentes familles de régressions, l'analyse multivariée, l'analyse bayésienne, la statistique non paramétrique, les modèles de survie... Pour une description exhaustive, vous pouvez consulter la page suivante :

www.sas.com/en_us/software/analytics/stat.html

SAS/ETS (*Econometrics and Times Series*) : outils destinés à l'économétrie des séries temporelles et à la prévision (ARIMA, VARMAX, TIMESERIES...).

www.sas.com/en_us/software/analytics/ets.html

SAS High-Performance Forecasting : module dédié à la prévision.

www.sas.com/en_us/software/high-performance-analytics/forecasting.html

Ce module fait partie d'une architecture plus vaste, **SAS High-Performance Analytics**, dédiée aux besoins d'intégration, de *reporting* et de gestion analytique des données *big data*. Cette architecture comprend d'autres modules comme **SAS Visual Analytics** dédiée à l'exploration des *big data*, **SAS High-Performance Data Mining**, **SAS High-Performance Econometrics**...

www.sas.com/en_us/software/high-performance-analytics.html

SAS/IML (*Interactive Matrix Language*) correspond au langage matriciel. Véritable « laboratoire mathématique de SAS », il permet de concevoir vos propres algorithmes selon une approche statistiques-mathématiques et d'exploiter vos calculs vectoriels et matriciels en recourant à de nombreuses fonctions et routines (optimisation non linéaire, filtre de Kalman, transformées de Fourier...) pour les appliquer directement en production sur vos données. IML est un langage de prototypage qui vous permet donc de concevoir vos propres algorithmes, selon une approche plus proche de l'algèbre linéaire. IML est également le langage de choix pour communiquer avec d'autres environnements (Matlab, R...)

www.sas.com/en_us/software/analytics/iml.html

SAS/OR (*Operational Research*) regroupe des procédures avancées dédiées à la recherche opérationnelle traitant de l'optimisation mathématique (programmation linéaire, non linéaire, variables entières et/ou mixtes), de métaheuristique (algorithmes génétiques) comme de la résolution d'optimisation par la théorie des graphes.

www.sas.com/en_us/software/analytics/sas-or.html

Un deuxième ensemble de modules est dédié à l'accès aux données et à leur gestion. Il comprend les modules suivants :

SAS/ACCESS permet d'accéder aux données produites par d'autres logiciels, de les exploiter dans SAS et même de créer à partir de SAS des fichiers dans des formats externes. Ses

différents modules permettent une intégration à différents Systèmes de Gestion de Base de Données Relationnelles (SGBDR) ou NoSQL du marché et à certains progiciels de gestion intégrée (PGI ou ERP – *Enterprise Resource Planning*). Voici une liste d'exemples non exhaustive : SAS/ACCESS Interface to SAP BW, to Teradata, to DB2, to MySQL, to Hadoop, to ODBC, to OLE DB, to ORACLE, to PC Files (Microsoft Excel, Microsoft Access, Lotus 1-2-3, DBF), to PeopleSoft, etc.

www.sas.com/en_us/software/data-management/access.html

SAS/CONNECT permet de connecter des ordinateurs sur lesquels est installé SAS. Vous partagerez au moyen de cette application vos tables de données ; des programmes écrits sur un ordinateur pourront être exécutés sur un second. Les applications de *Grid Computing* de SAS nécessitent SAS/CONNECT.

http://support.sas.com/software/products/connect_share/index.html

D'autres modules mettent à votre disposition des interfaces graphiques :

SAS Enterprise Guide permet de manipuler des données, de faire des analyses statistiques et des rapports. Cette application développée en .NET est donc une application client uniquement Windows. Enterprise Guide génère du code SAS, que ce soit du Base SAS, SAS/STAT, SAS/QC..., qui est exécuté par un moteur Base SAS. Ainsi, l'ensemble des procédures d'analyses statistiques est utilisable en mode production : SAS Enterprise Guide permet d'effectuer ici la majorité des tâches qu'un programmeur SAS peut exécuter, qu'il s'agisse de *data management* ou d'analyses statistiques en mode « production ».

www.sas.com/en_us/software/enterprise-guide.html

SAS Enterprise Miner est l'interface de SAS dédiée au *data mining*. SAS considère le *data mining* comme une discipline à part entière, et notons ici que les algorithmes embarqués sont bien différents de ceux plus classiquement utilisés en statistiques. Cette interface permet une industrialisation de l'exploitation intensive de gigantesques bases de données. Il s'agit donc d'une interface experte, qui va bien au-delà de la statistique, aux croisements du *statistical learning* et du *machine learning*. Les procédures embarquées sous SAS EM sont spécifiques à la discipline (DMREG : *Data Mining Regressions*), ce qui les distingue d'une approche statistique « classique ».

www.sas.com/en_us/software/analytics/enterprise-miner.html

SAS Data Integration Studio est une interface cliente Java de la plate-forme décisionnelle SAS dédiée à l'ETL (*Extract, Transform, Load*). Elle fait partie des packages « SAS Data Integration Server » et « SAS Enterprise Data Integration Server ».

<http://support.sas.com/software/products/etls/>

SAS Forecast Server est une application d'aide à la production de prévisions.

www.sas.com/en_us/software/analytics/forecastserver.html

SAS/IML Studio est une interface experte avancée pour le statisticien maîtrisant l'exploration graphique comme la programmation SAS. Elle permet de recourir à la statistique exploratoire tout en se concentrant sur le développement de nouveaux algorithmes (approche « statistiques-mathématiques ») et nécessite la maîtrise de la programmation SAS/STAT et SAS/IML.

<http://support.sas.com/rnd/app/studio/studio.html>

Enfin, citons :

SAS NLS (*National Language Support*), qui n'est pas un module en tant que tel mais une application qui permet à SAS de fonctionner au mieux sur des systèmes non américains.

SAS®

3^e édition

Introduction au décisionnel : du data management au reporting

Sébastien Ringuédé

Sébastien Ringuédé est maître de conférences en sciences économiques à l'université d'Orléans. Il utilise SAS depuis plus de 20 ans dans le cadre des recherches qu'il effectue au sein du Laboratoire d'Économie d'Orléans (LEO – UMR CNRS 7322). Il enseigne SAS aux étudiants du master ESA (Économétrie et Statistique Appliquée) depuis sa création.

Ce livre présente le socle de connaissances communes à tous les utilisateurs de SAS, le progiciel d'informatique décisionnelle le plus utilisé au monde. Il traite tout particulièrement les fonctionnalités de Base SAS, module au cœur du système SAS. Pédagogique et complet, il peut servir aussi bien de guide d'initiation pour les utilisateurs débutants que d'ouvrage de référence pour les utilisateurs plus expérimentés. Il couvre par ailleurs l'intégralité du programme de la certification Base Programming for SAS®9.

Parmi les sujets développés :

- la création, la manipulation et la gestion des tables de données
- les procédures d'exploration des données : construction de tableaux, de rapports, de graphiques au moyen des procédures ODS Graphics
- la production de documents au format HTML, PowerPoint, RTF, XLS ou PDF avec ODS
- la procédure PROC SQL et le langage SQL de SAS
- le langage macro spécifique à SAS

Pour vous aider à bien assimiler tous les concepts, le livre comprend près de 500 programmes d'exemples, plus de 150 exercices d'application et des liens vers une centaine d'articles en ligne.

Enrichie de près d'une centaine de pages, cette nouvelle édition propose des mises à jour importantes sur :

- les passerelles entre SAS et Excel
- les nouveautés des versions 9.3 et 9.4
- les fusions de tables au moyen du langage SAS, des tables de hachage et de PROC SQL
- la production de graphiques au moyen de PROC SGPlot et PROC SGPANEL

Le livre, qui porte essentiellement sur les versions 9.x de SAS, est compatible avec la version 8.

Public : universités, écoles d'ingénieurs, écoles de commerce ; professionnels

Cours : introduction à SAS, SAS

Niveau : débutants / intermédiaires

Web
ressources

Sur le site www.sas-sr.com :

- le code des programmes et les sorties générées
- les tables et fichiers nécessaires aux exemples et aux exercices
- les solutions des exercices et des exercices supplémentaires
- un forum de discussion
- des ressources sur la certification Base Programming for SAS®9
- de nombreux autres suppléments

ISBN : 978-2-3260-0066-7

