

Université de Nantes – U.F.R. des Sciences
Master professionnel II : Ingénierie mathématique

Présentation du logiciel SAS

Frédéric Lavancier

Le logiciel possède une aide très fournie. Par ailleurs une documentation en ligne (en anglais) se trouve à l'adresse suivante :

<http://support.sas.com/documentation/93/index.html>

Une documentation assez complète (en français) peut également être trouvée à l'adresse suivante : <http://www.math.univ-toulouse.fr/~besse/pub/saspdf.pdf>

1 Prise de Contact

A l'ouverture du logiciel, il est proposé d'essayer le "Guide d'initiation" : c'est un guide interactif qui permet de cerner les principes de base du logiciel. Il est vivement conseillé d'en profiter : après le choix "New SAS programmer (quick-start guide)", faire au moins le tour des quatre parties de la première section "Getting to know SAS" (chez soi).

L'interface de base se compose des fenêtres suivantes :

- "Explorateur" : affiche l'arborescence des bibliothèques contenant les tables SAS ainsi que celle de l'ordinateur;
- "Résultats" : affiche l'arborescence des résultats (sorties de procédures, graphiques,...) pour un accès rapide dans la fenêtre "Sortie";
- "Editeur" : éditeur de texte où l'on écrit les programmes SAS avant de les lancer;
- "Journal" : affiche le résumé des programmes lancés et les messages d'erreurs éventuels;
- "Sortie" : affiche les résultats des programmes (sauf les graphiques qui s'ouvrent dans une fenêtre spécifique).

Le fonctionnement de base de SAS consiste en la modification et/ou l'analyse de tableaux de données grâce à des procédures que l'on écrit dans l'éditeur avant de les lancer. Les procédures permettent de modifier ou de créer des tableaux, d'en faire des analyses statistiques, des représentations graphiques, etc. Certaines procédures peuvent être lancées directement en clique-bouton à l'aide du menu déroulant.

2 Les bibliothèques et les tableaux de données

2.1 Les bibliothèques

Les bibliothèques (ou librairies) contiennent les tableaux. Par défaut on trouve les librairies:

- "Maps" : des données cartographiques (pour faire des fonds de cartes),
- "Sasuser" : bibliothèque permanente, on peut y stocker des tables pour les sessions suivantes,
- "Sashelp" : des tables illustrées dans l'aide de SAS,
- "Work" : la librairie par défaut, c'est ici que les nouvelles tables sont placées si une librairie n'est pas précisée.

On peut créer sa propre librairie en l'associant à un répertoire personnel qui contient des tableaux SAS : soit en cliquant sur l'icône "nouveau" lorsqu'on est dans l'explorateur, soit en ligne de commande dans l'éditeur avec la commande LIBNAME. Par exemple

```
LIBNAME TP-SAS '/mes documents/tp';
```

crée la librairie TP-SAS associée au répertoire /mes documents/tp.

Dans les programmes on accède à la table "matable" de la librairie "TP-SAS" par "TP-SAS.matable". Par défaut la librairie est "work" de telle sorte que les commandes "work.matable" et "matable" sont équivalentes.

2.2 Les tableaux

En double-cliquant sur les tableaux présents dans les librairies, on les ouvre dans une fenêtre nommée VIEWTABLE. Cela permet de les visualiser mais pas de les modifier. La modification de tableaux se fait essentiellement par des lignes de codes dans l'éditeur à l'aide de l'étape DATA (voir plus loin).

Il est aussi possible d'éditer les tableaux pour des modifications mineures en cliquant dans Solution/Analyse/Analyse interactive des données.

Exercice :

Créer une librairie TP-SAS dans laquelle vous importerez, dans une table que vous nommerez "us", le fichier US_Air_Transportation.txt (cf partie suivante). Visualiser cette table et modifier quelques cellules à l'aide de l'Analyse interactive des données.

3 Manipulation des données : Etape DATA

3.1 Importer des données

Cela se fait à l'aide du menu Fichier/Importer des données. Plusieurs formats sont présents pour l'importation. Pour les fichiers .txt, il se peut que le séparateur soit mal géré, dans ce cas on peut tenter de le définir à l'aide du menu User-defined formats, sinon il faut importer la table à l'aide d'une ligne de commande dans l'éditeur en précisant les

options (cf les manuels de SAS). Pour les tables qui s'importent mal dans SAS, le plus simple est de les convertir au préalable à l'aide d'autres logiciels (comme un tableur).

Exercice :

Importer la série "voitures.txt" dans la librairie "work". On appellera "voit" cette table par la suite.

3.2 Modifier et créer un tableau

Cela se fait dans l'éditeur grâce à l'étape DATA. La syntaxe générale d'une étape DATA est la suivante (les commentaires dans les programmes SAS sont encadrés par /* et */):

```
DATA TP-SAS.matable; /*création de la table "matable" dans la librairie TP-SAS*/
set TP-SAS.us; /*on importe la table "us" déjà existante dans TP-SAS*/
... ; /*on code les différentes opérations souhaitées*/
run;
```

On peut alors lancer cette étape DATA en sélectionnant le bloc depuis DATA jusque run puis en cliquant sur "soumettre" (le petit bonhomme qui court) ou en appuyant sur F8.

Quelques exemples de manipulation (on travaille dans la librairie "work") :

- Si on veut ajouter à la table "voit" contenant la série "ventes" une variable "diff12" égale à la différence à l'horizon 12 des ventes et une variable "logvente" égale au logarithme des ventes :

```
DATA voit; /*on crée la table "voit", cela écrasera la précédente*/
set voit; /*on y copie l'ancienne table "voit"*/
ventes12 = lag12(ventes); /*lag12 décale la série de 12 pas*/
diff12 = ventes12-ventes;
drop ventes12; /*on supprime la variable intermédiaire ventes12 */
logvente = log(ventes);
run;
```

- Pour créer une table voit2 ne contenant que les observations 100 à 200 :

```
DATA voit;
set voit;
id = _N_; /*on ajoute la variable "id" qui numérote les lignes*/
run;
DATA voit2;
set voit;
where id<=200 and id>=100; /*restriction pour la création de voit2*/
run;
```

- Pour ajouter à la table "voit" une variable log12 valant $\log(\text{diff12})$ lorsque $\text{diff12} > 0$, et valant 0 sinon :

```
DATA voit;
set voit;
if diff12>0 then log12 = log(diff12);
else log12=0;
run;
```

- Pour ajouter un bruit blanc gaussien de moyenne égale à 10 et de variance égale à 2 à la série "ventes" :

```
DATA voit;
set voit;
ventes2 = ventes+sqrt(2)*rannor(0)+10;/*rannor(0) simule une N(0,1)*/
run;
```

4 Les procédures

L'essentiel de la programmation sous SAS se fait à l'aide de procédures : elles permettent des manipulations de données (PROC SORT, PROC SQL, ...), la représentation de graphiques (PROC GPLOT, PROC GCHART,...), le calcul de statistiques de base (PROC MEANS, PROC FREQ, ...), ou des analyses statistiques plus poussées telles que des modèles de régression, des modèles ARMA, de la classification (PROC REG, PROC ARIMA, PROC CLUSTER,...).

La syntaxe générale d'une procédure est la suivante (par exemple pour PROC MEANS):

```
PROC MEANS data=TP-SAS.voit; /*on précise à quelle table appliquer la proc*/
var ventes; /*la ou les variables de la table à traiter*/
... ; /*éventuellement d'autres commandes*/
run;
```

Les options nécessaires à la procédure ainsi que ses commandes (comme `var` ici) sont différentes selon les procédures. Il est conseillé de consulter l'aide de SAS pour connaître la syntaxe de la procédure utilisée (par exemple en tapant son nom dans "index" et en sélectionnant "syntax" dans les choix proposés).

Quelques procédures :

- PROC CONTENTS : résume le contenu d'un tableau;
- PROC SORT : permet de trier un tableau de données selon une ou plusieurs variables;
- PROC SQL : permet de faire des requêtes au langage SQL (peut se substituer à l'étape DATA dans certains cas);
- PROC MEANS ou PROC UNIVARIATE : analyse descriptive simple de variables quantitatives;

- PROC FREQ : analyse descriptive simple de variables qualitatives;
- PROC GCHART : pour faire des graphiques descriptifs tels que des histogrammes, des camemberts, etc;
- PROC GPLOT : pour faire différentes représentations graphiques (nuage de points, séries,...);
- PROC CORR : calcul de corrélations;
- PROC TTEST : pour faire des tests de comparaison entre échantillons;
- PROC PRINCOMP ou PROC FACTOR : permet d'effectuer des ACP;
- PROC CORRESP : permet d'effectuer des AFC;
- PROC CLUSTER : permet d'effectuer des CAH;
- PROC REG : permet d'estimer un modèle de régression linéaire avec les tests associés;
- PROC GLM : pour estimer un modèle linéaire généralisé;
- PROC LOGISTIC : permet d'effectuer une régression logistique;
- PROC DISCRIMINATE : analyse discriminante (classification);
- PROC ARIMA : pour estimer des modèles ARMA ou SARIMA.

Certaines analyses peuvent être réalisées en clique-bouton à l'aide du menu "Solutions/Analyse" (représentations graphiques, modélisation SARIMA,...) mais le choix des options est moins complet que via la PROC correspondante.

Exemples d'utilisation de certaines procédures effectuées sur la table "voit" :

```

/*Résumé de la table voit (située dans la librairie work)*/
proc contents data=voit;
run;

/*Statistiques de base sur la variable "ventes"*/
proc univariate data=voit;
var ventes;
run;

/*Un histogramme en 3D des valeurs de "ventes"*/
proc gchart data=voit;
vbar3D ventes /percent; /*percent est une option, elle se met après / */
run;

/*On initialise les paramètres graphiques*/
goption reset=global;

```

```

/*Représentation de la série "ventes"*/
proc gplot data=voit;
plot ventes*id;
symbol1 interpol=join; /*permet de lier les points en une ligne*/
title "Ventes mensuelles de voitures";
run;

```

Exercice :

Importer le fichier `US_Air_Transportation.txt` dans la table "us" de la librairie `work`. Ce fichier concerne l'industrie du transport aérien aux Etats-Unis. Il contient les variables suivantes :

- K : production du facteur capital;
- L : production du facteur travail;
- W : indice des prix du facteur travail;
- R : indice des prix du facteur capital;
- Y : production.

L'objectif est de modéliser le ratio L/K en fonction du ration W/R et de la production Y .

1. Résumer le contenu de la table "us". (→ `proc contents`)
2. Créer les variables L/K , W/R puis $\log(L/K)$, $\log(W/R)$ et $\log(Y)$. (→ `data ...`)
3. Faire une analyse descriptive des variables. (→ `proc means`)
4. Représenter les nuages de points de L/K en fonction de W/R , et de $\log(L/K)$ en fonction de $\log(W/R)$. (→ `proc gplot`)
5. Effectuer la régression de $\log(L/K)$ en fonction de $\log(W/R)$ et de $\log(Y)$. (→ `proc reg`)
6. Conclure sur le lien obtenu (il s'agit du modèle de Cobb-Douglas).

Exercice :

Effectuer une analyse en composantes principales sur le fichier `eaux-2008.txt` (→ `proc princomp`) et représenter les variables et les individus sur les deux premiers axes.