

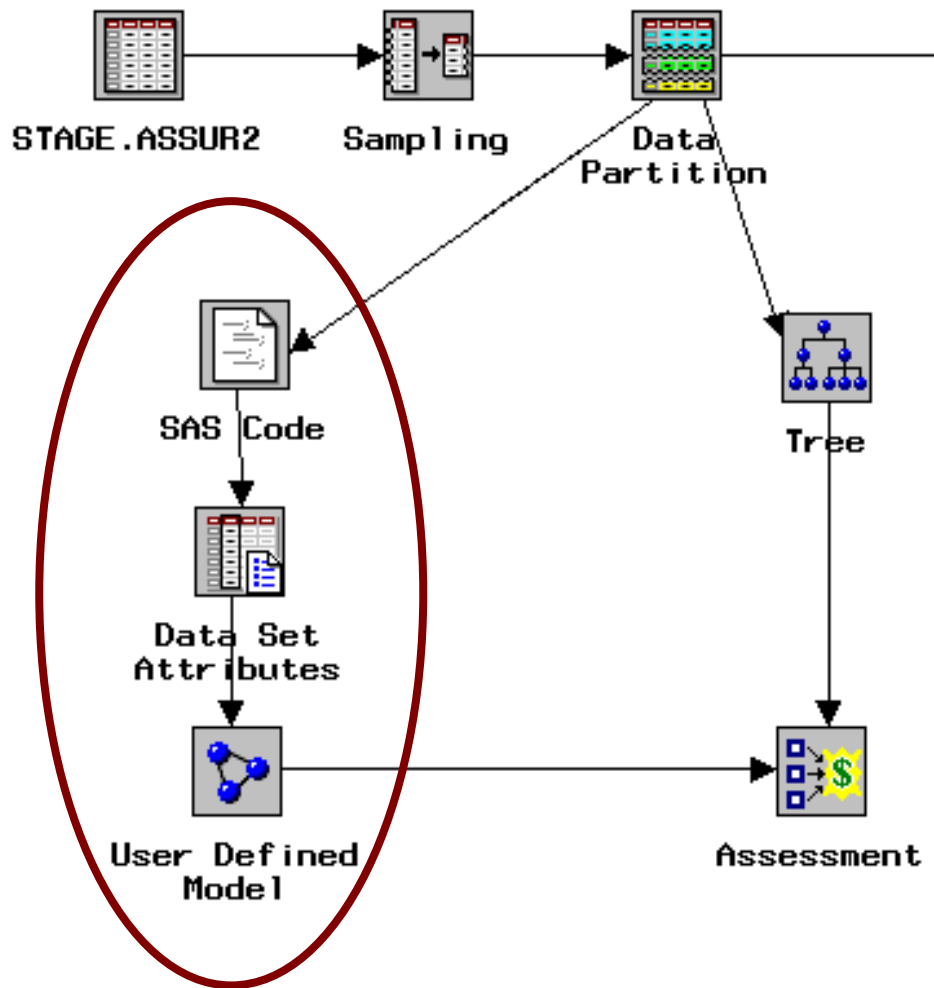
User-Defined Model

Buts :

- ◆ intégrer un modèle développé hors de SEM
- ◆ le comparer aux autres modèles de SEM (arbre, régression, ...)

User-Defined Model

Le diagramme à utiliser



User-Defined Model

Rôle de chaque étape :

◆ SAS CODE

Permet d'insérer le code SAS du modèle « externe ». On peut y écrire n'importe quel programme SAS.

Ce programme doit impérativement produire une table de validation scorée, et si cela est possible, une table de test également scorée.

Il importe donc plus dans ce nœud SAS CODE de fournir un moteur de score appliqué aux différents jeux de données (par exemple avec une étape Data) que de construire un modèle, ce qui peut être effectué en dehors de SEM.

◆ DATA SET ATTRIBUTES

Les deux tables scorées (entraînement et validation) sont proposées « brutes » hors du nœud SAS CODE. Ce deuxième nœud spécifie le rôle des nouvelles variables, en particulier celle qui sert de score proprement dit, ou de décision.

◆ USER-DEFINED MODEL

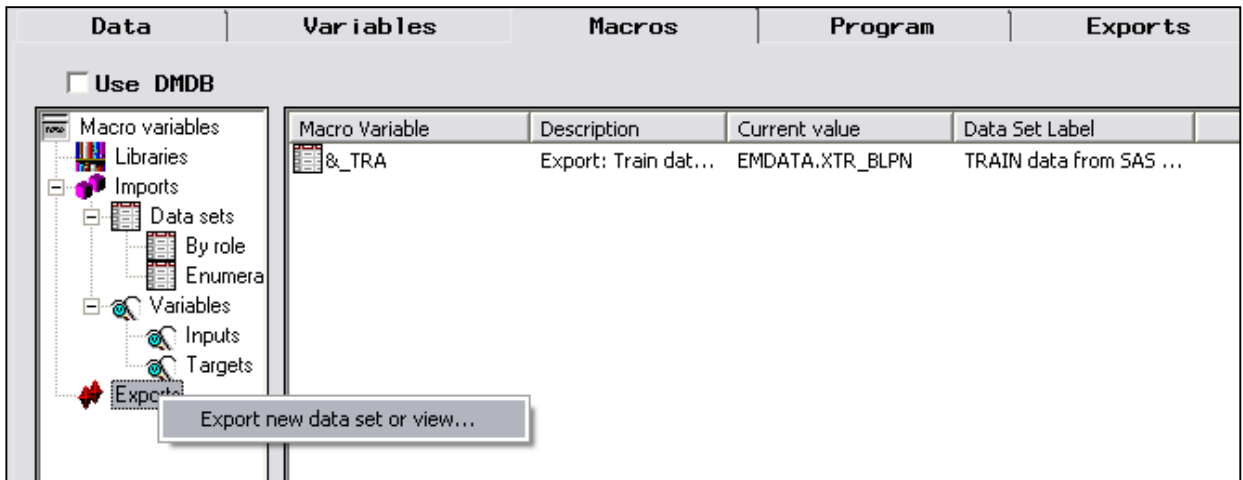
Permet d'introduire dans le circuit des modèles SEM le modèle externe. Calcule entre autres les données de lift.

User-Defined Model

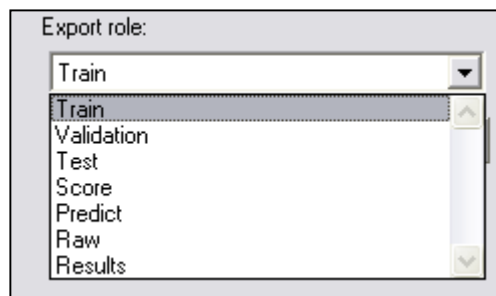
Le nœud SAS CODE

1. Tables en sortie

Selon les cas, il y aura 2 (si pas de test) ou 3 tables en sortie du nœud SAS CODE. Pour les déclarer, dans l'onglet MACRO, cliquer (bouton droit) sur EXPORTS dans l'arborescence à gauche de l'écran.



Le choix qui apparaît alors est EXPORT NEW DATA SET OR VIEW. Il conduit à cet écran :

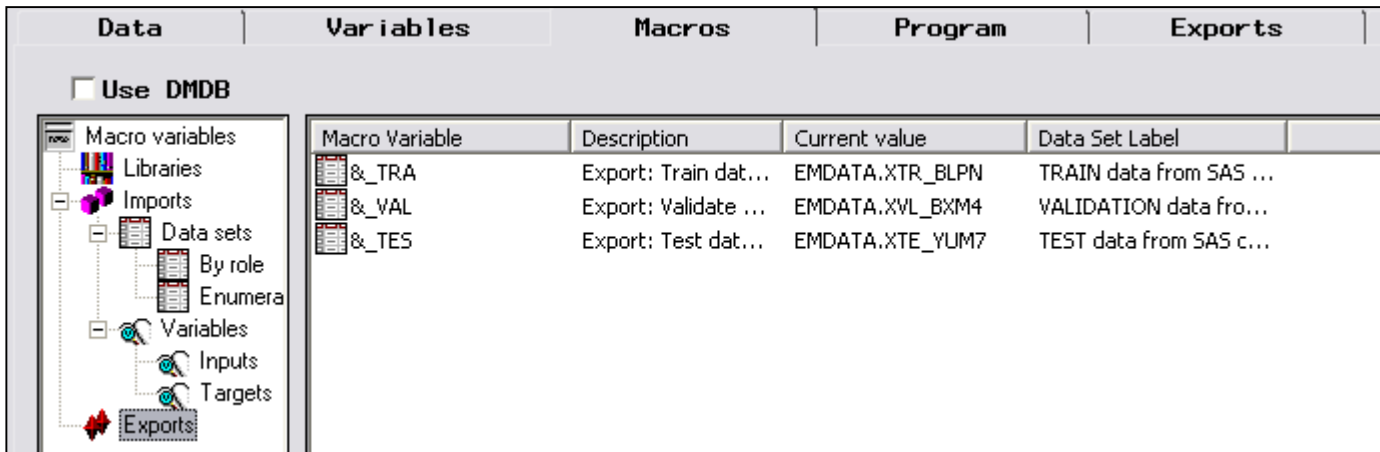


L'opération est répétée pour les trois types de table en sortie (entraînement, validation et test).

User-Defined Model

Le nœud SAS CODE

1. Tables en sortie



Macro Variable	Description	Current value	Data Set Label
&_TRA	Export: Train dat...	EMDATA.XTR_BLPN	TRAIN data from SAS ...
&_VAL	Export: Validate ...	EMDATA.XVL_BXM4	VALIDATION data fro...
&_TES	Export: Test dat...	EMDATA.XTE_YUM7	TEST data from SAS c...

Les tables créées correspondent par défaut aux macro-variables `&_TRA`, `&_VAL` et `&_TES`.

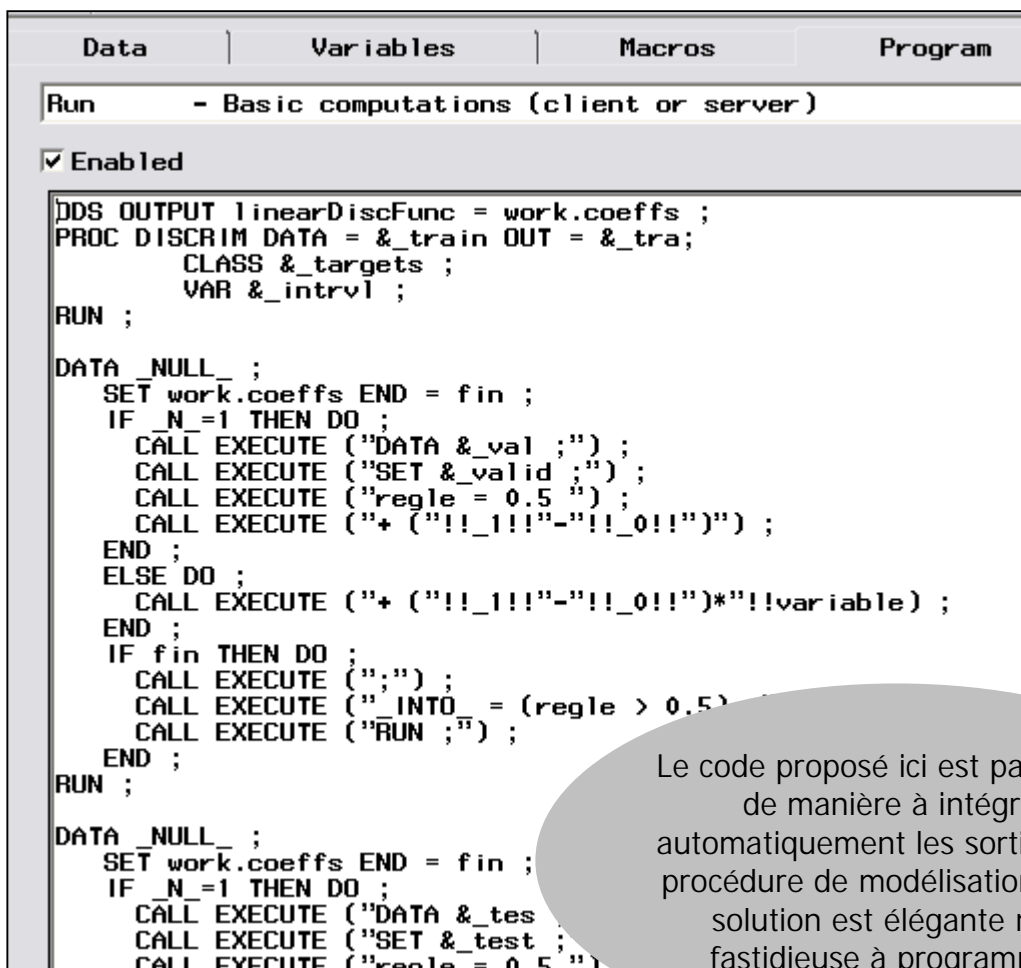
Le moteur de score dans l'onglet PROGRAM fera référence à ces macro-variables dans ses sorties d'étape DATA ou de PROC.

User-Defined Model

Le nœud SAS CODE

2. Code du moteur de score

En général, il s'agit d'une étape Data construite à partir des coefficients du modèle. Pour une analyse discriminante comme pour une régression (linéaire ou logistique), une nouvelle variable est calculée, qui est la prédiction de la valeur de la variable à expliquer (cible).



```
DDS OUTPUT linearDiscFunc = work.coeffs ;
PROC DISCRIM DATA = &_train OUT = &_tra;
    CLASS &_targets ;
    VAR &_intrvl ;
RUN ;

DATA _NULL_ ;
    SET work.coeffs END = fin ;
    IF _N_=1 THEN DO ;
        CALL EXECUTE ("DATA &_val ;");
        CALL EXECUTE ("SET &_valid ;");
        CALL EXECUTE ("regle = 0.5 ");
        CALL EXECUTE ("+ ("!!_1!!"- "!!_0!!")");
    END ;
    ELSE DO ;
        CALL EXECUTE ("+ ("!!_1!!"- "!!_0!!")*"!!variable) ;
    END ;
    IF fin THEN DO ;
        CALL EXECUTE ("");
        CALL EXECUTE ("_INTO_ = (regle > 0.5)");
        CALL EXECUTE ("RUN ;");
    END ;
RUN ;

DATA _NULL_ ;
    SET work.coeffs END = fin ;
    IF _N_=1 THEN DO ;
        CALL EXECUTE ("DATA &_tes");
        CALL EXECUTE ("SET &_test");
        CALL EXECUTE ("regle = 0.5 ");
```

Le code proposé ici est paramétré de manière à intégrer automatiquement les sorties de la procédure de modélisation. Cette solution est élégante mais fastidieuse à programmer.

User-Defined Model

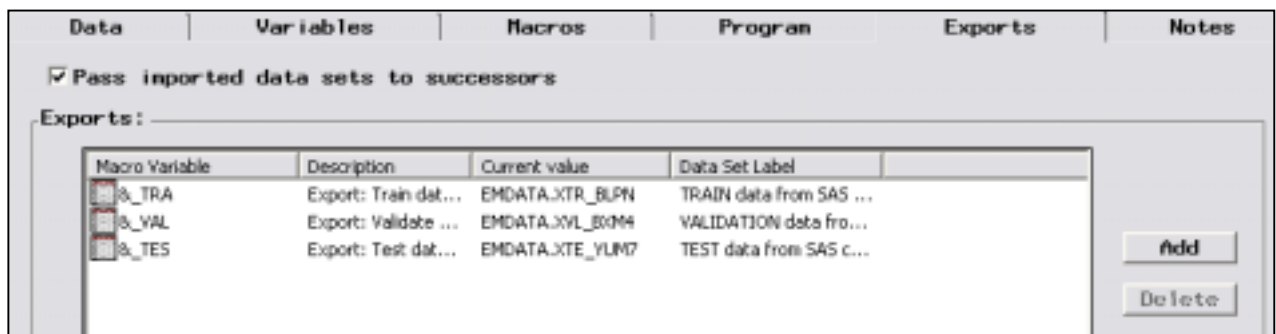
Le nœud SAS CODE

2. Code du moteur de score

Dans le code du moteur de score, il est important de noter le nom de la variable qui contient la prédiction pour l'événement recherché. Ici, la variable `_INTO_` en tient lieu. Dans la sortie d'une régression logistique, on pourra utiliser directement le score.

3. Export du nœud

L'onglet EXPORT doit rappeler les 2 ou 3 tables créées.



User-Defined Model

Le nœud DATA SET ATTRIBUTES

On y sélectionne à son tour chacune des tables créées par le nœud SAS CODE, dans l'onglet DATA.

Name	Output	Export	Role
EMDATA.XTR_BLPN	EMDATA.EXCPRQ9S	Yes	TRAIN
EMDATA.XVL_BXM4	EMDATA.EXTBGODO	Yes	VALIDATE
EMDATA.XTE_YUM7	EMDATA.EXUG4RJO	Yes	TEST
EMDATA.TRN79EDA	EMDATA.EXNDAHSY	No	TRAIN
EMDATA.VALT5Y85	EMDATA.EX089RJO	No	VALIDATE
EMDATA.TST938YM	EMDATA.EXKQTZ8A	No	TEST

Dans l'onglet VARIABLES, il faut alors repérer la variable « prédiction » créée par le moteur de score. Ici, `_INTO_`.

D'un clic du bouton droit sur sa caractéristique NEW MODEL ROLE, on sélectionne SET NEW MODEL ROLE.

Variable	Yes	rejected	rejected	interval
RENTA	Yes	rejected	rejected	interval
GOOD_BAD	Yes	target	target	binary
CRASH	Yes	input	input	binary
_0	Yes	input	input	interval
_1	Yes	input	input	interval
INTO	Yes	input	input	input

- Set New Model Role
- Sort by New Model Role
- Subset by New Model Role
- Find New Model Role
- View Distribution of `_INTO_`
- Reset Role

- input
- target
- rejected
- id
- cost
- freq
- trial
- predict**
- sequence
- residual
- group
- assess
- column
- row
- crossid
- timeid

Ce nouveau rôle est PREDICT.

User-Defined Model

Le nœud DATA SET ATTRIBUTES

On procède de même pour chacune des 2 ou 3 tables créées par le nœud SAS CODE. Au final, pour chacune des tables, l'aspect de l'onglet VARIABLES doit être le suivant...

Name	Keep	Model Role	New Model Role	Measurement
NB_SIN	Yes	input	input	binary
C_PERMIS	Yes	input	input	nominal
MARQUE	Yes	input	input	nominal
USAGE	Yes	input	input	nominal
PROF	Yes	input	input	nominal
PUISS	Yes	input	input	interval
VALEUR	Yes	input	input	interval
GARAGE	Yes	rejected	rejected	nominal
ZONTARIF	Yes	input	input	nominal
FAMIL	Yes	input	input	nominal
PERMIS	Yes	rejected	rejected	nominal
GARFERM	Yes	input	input	binary
PRIME	Yes	rejected	rejected	interval
CONT	Yes	id	id	nominal
NBSIN	Yes	input	input	ordinal
COUT	Yes	rejected	rejected	interval
CRM_N	Yes	input	input	interval
AGE	Yes	input	input	interval
ANC_VEH	Yes	input	input	interval
ANC_CONT	Yes	input	input	interval
RENTA	Yes	rejected	rejected	interval
GOOD_BAD	Yes	target	target	binary
CRASH	Yes	input	input	binary
_0	Yes	input	input	interval
_1	Yes	input	input	interval
INTO	Yes	input	predict	binary

User-Defined Model

Le nœud USER DEFINED MODEL

Son onglet-clé est ici PREDICTED VALUES.

The screenshot shows the 'PREDICTED VALUES' tab of the USER DEFINED MODEL node. The interface includes a 'Target' dropdown set to 'GOOD_BAD', a 'Measurement' field set to 'binary', and an 'Event' field set to '1'. A 'Profile information' button is visible. Below these fields is a table with the following data:

Level	Predicted Variable	Modelling Prior	Label
1		.	Predicted: GOOD_BAD=1
0		.	Predicted: GOOD_BAD=0

On doit y définir quelle est la variable qui représente un score (ou une prédiction) pour les tables proposées en entrée de ce nœud. Un clic (bouton droit) dans la colonne PREDICTED VARIABLE, à hauteur de la ligne appropriée (selon l'événement prédit) permet de sélectionner cette variable.

This screenshot shows the same interface as above, but with a context menu open over the 'PREDICTED VARIABLE' column for the '1' level. The menu options are:

- Select Predicted Variable...
- Clear
- Sort by Predicted Variable
- Find Predicted Variable

User-Defined Model

Le nœud USER DEFINED MODEL

Target: ▼

Measurement: binary

Event: 1

Level	Predicted Variable	Model	Label
1			GOOD_BA
0	<input type="text" value="_INTO_"/>		OD_BA

Sont proposées dans la liste toutes les variables de rôle PREDICT. Ici, une seule variable a été associée à ce rôle.

L'onglet PREDICTED VALUES est alors sous sa forme définitive.

Target: ▼

Measurement: binary

Event: 1

Level	Predicted Variable	Modelling Prior	Label
1	_INTO_	.	Pred
0		.	Pred

Reste à fermer le nœud (la fenêtre du MODEL MANAGER apparaît comme pour les autres modèles de SEM) et à l'exécuter.

Save Model As

Model Name:

Model Description:

User-Defined Model

Le nœud ASSESSMENT

Il considère désormais le modèle externe comme n'importe quel modèle.



Assessment Tool

Models | Options | Reports | Output

Tool	Name	Description	Target	Target Event	By Group I
User Defined	AnaDisc	Analyse discriminante	GOOD_BAD		
Tree	Untitled	Tree	GOOD_BAD	1	

