

PROC CLUSTER ET PROC TREE

CLASSIFICATION ASCENDANTE HIERARCHIQUE

Lorsque l'analyse statistique ne porte pas sur la dernière table (Data) Sas mémorisée, la procédure de classification ascendante hiérarchique PROC CLUSTER doit être suivie de l'option DATA=nomtab1 où nomtab1 est le nom du tableau d'entrée (Data) Sas contenant les données à étudier. Si l'option DATA=nomtab1 est absente, l'analyse porte alors sur la dernière table Sas mémorisée.

Avec ou sans option DATA, la procédure PROC CLUSTER doit être obligatoirement suivie de l'option d'agrégation des données METHOD= suivie de la méthode sélectionnée, AVERAGE ou CENTROID. L'option METHOD=AVERAGE retient la distance moyenne comme mesure de la distance entre deux groupes d'observations alors que l'option METHOD=CENTROID prend la distance entre les centres de gravité comme mesure de la distance entre deux groupes d'observations.

La procédure CLUSTER est traditionnellement suivie de la procédure TREE dessinant l'arbre (le dendrogramme) de la classification hiérarchique. Le tableau de sortie créé par la procédure PROC CLUSTER étant repris, comme tableau d'entrée, par la procédure PROC TREE, on peut passer de la procédure CLUSTER à la procédure TREE sans aucune difficulté comme l'illustre le programme de classification ci-dessous.

```
PROC CLUSTER DATA=nomtab1 METHOD=Average;  
  VAR V1-V5 ;  
  PROC TREE ;  
  RUN ;
```

On peut aussi réaliser une classification hiérarchique PROC CLUSTER, suivie de la procédure PROC TREE), dans l'espace des facteurs donnés au terme d'une analyse factorielle par correspondance réalisée avec PROC CORRESP comme l'indique le programme ci-dessous.

```
PROC CORRESP DATA=nomtab1 DIM=3 OUT=toto NOPRINT ;  
  VAR cac dj ftm can net ;  
  ID ident ;  
  PROC CLUSTER METHOD=CENTROID ;  
  VAR DIM1-DIM3 ;  
  ID ident ;  
  PROC TREE ;  
  RUN ;
```

