

PROC CORRESP

ANALYSE FACTORIELLE PAR CORRESPONDANCE

Dans une optique exploratoire, l'analyse factorielle par correspondance procède, sur la base de la distance du Chi-deux, à une étude typologique des informations contenues dans un tableau de contingence. Comme l'indique le tableau ci-dessous, la commande PROC CORRESP peut être suivie de l'option DATA= puis des options optnum séparées chacune par un caractère blanc. Les instructions VAR, CLASS, WEIGHT, FREQ, ID, CLASS, BY et OUTPUT OUT, séparées chacune par un point-virgule, affinent l'analyse.

```
PROC CORRESP DATA=nomtab1 optnum;  
  VAR var1 var2 var3 var4 var5 var6;  
  CLASS var1 ...;  
  PRIOR var2 ...;  
  WEIGHT var3;  
  FREQ var4 ... ;  
  ID var5;  
  TESTCLASS var1 ...;  
  TESTFREQ var4 ...;  
  TESTID var5;  
  BY var6 ...;
```

Panorama des options disponibles

Lorsque l'analyse statistique ne porte pas sur la dernière table (Data) Sas mémorisée, la commande PROC CORRESP 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 (Data) Sas mémorisée.

Les options optnum figurant dans les tableaux ci-après, suivant éventuellement DATA=nomtab1, permettent à la fois de préciser les conditions techniques de l'analyse,

Option	Utilité
CANONICAL	réalise une analyse discriminante canonique
METHOD=NORMAL	effectue une analyse discriminante basée sur l'hypothèse d'une distribution normale des variables (option par défaut)
METHOD=NPAR	effectue une analyse discriminante basée sur une méthode non paramétrique
ANOVA	teste l'hypothèse d'égalité des moyennes des classes en mode univarié
MANOVA	teste l'hypothèse d'égalité des moyennes des classes en mode multivarié
ALL	retient toutes les options possibles

les tables Sas (Data) à retenir,

Option	Utilité
OUT=nomtab2	donne le nom de la table (Data) de sortie (ici, <i>nomtab2</i>)
OUTCROSS=nomtab3	donne le nom de la table (Data) de sortie (ici, <i>nomtab3</i>) contenant les données de base, les probabilités ainsi que les classifications réalisées
OUTSTAT=nomtab4	donne le nom du fichier de sortie (ici, <i>nomtab4</i>) contenant les différentes statistiques relatives aux variables sélectionnées
TESTDATA=nomtab5	donne le nom de la table (Data) de sortie (ici, <i>nomtab5</i>) contenant le classement des variables (utile seulement si on utilise l'instruction TESTCLASS, TESTFREQ ou TESTID)

ainsi que les matrices de corrélation ou les matrices de variance-covariance désirées des classes construites lors de l'analyse.

Option	Utilité
BCORR	crée la matrice des corrélations inter-classes
PCORR	crée la matrice des corrélations intra-classes
WCORR	crée les matrices de corrélations intra-classes pour chaque classe
BCOV	crée la matrice de variances-covariances inter-classes
PCOV	crée la matrice de variances-covariances intra-classes
WCOV	crée les matrices de variances -covariances intra-classes pour chaque classe

Panorama des instructions disponibles

De nombreuses instructions, séparées chacune par un point-virgule, peuvent figurer après la commande PROC CORRESP comme l'indique le tableau ci-dessous qui précise l'utilité de chacune de ces instructions.

Instruction	Utilité
VAR	fixe la liste des variables retenues (<i>var1, var2, var3, var4, var5, var6</i>). En l'absence de cette instruction, toutes les variables sont retenues
CLASS	réalise les calculs par classe de variables (ici, <i>var1, ...</i>) données a priori (si les résultats de l'analyse sont identiques à ceux obtenus avec l'instruction BY, ils sont toutefois présentés sous une forme légèrement différente)
PRIOR	détermine la probabilité associée à la variable indiquée (ici, <i>var2, ...</i>)
WEIGHT	crée la variable (ici, <i>var3</i>) servant de facteur de pondération aux autres variables
FREQ	estime en pourcentages, simples et cumulés, les variables (ici, <i>var4, ...</i>)
ID	retient une variable (ici, <i>var5</i>) comme identificateur (par défaut, l'identificateur est le numéro de lignes <i>N</i> attribué à chaque observation)
TESTCLASS	vérifie que les variables citées (ici, <i>var1</i>) sont bien classées
TESTFREQ	vérifie que les fréquences associées aux variables indiquées (ici, <i>var4, ...</i>) sont bonnes
TESTID	vérifie que l'identificateur (ici, <i>var 5</i>) est bon
BY	réalise les calculs par classe de variables données a posteriori (ici, <i>var6, ...</i>).

N.B. : L'analyse factorielle par correspondance est parfois suivie par une classification réalisée au terme de l'analyse (à l'aide des procédures PROC CLUSTER et PROC TREE).