

PROC PRINCOMP

ANALYSE EN COMPOSANTES PRINCIPALES

Dans une optique le plus souvent exploratoire, l'analyse en composantes principales (A.C.P.) recherche, sur la base de la distance euclidienne, les axes factoriels doués d'une inertie maximale dans un tableau de données composés d'individus en ligne et de variables en colonne. Comme l'indique le tableau ci-dessous, la commande PROC PRINCOMP 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 PRINCOMP 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 PRINCOMP 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.

Les options optnum, figurant après l'éventuelle option DATA=nomtab1, permettent à la fois d'afficher les matrices de corrélation ou les matrices de variances-covariances désirées des classes construites lors de l'analyse

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

et de préciser les conditions techniques de l'analyse dans le tableau ci-dessous.

Option	Utilité
COVARIANCE	calcule les composantes principales à partir de la matrice variances-covariances (à défaut, les composantes principales sont calculées sur la base de la matrice des corrélations).
N=n	fixe le nombre de composantes principales (de vecteurs propres et de valeurs propres) à déterminer (par défaut, le nombre de variables)
NOINT	indique que la matrice de corrélation ne doit pas être corrigée par la moyenne
OUT=nomtab2	donne le nom de la table (Data) de sortie (ici, <i>nomtab2</i>) contenant les données et les composantes principales.
OUTSTAT=nomtab3	donne le nom du fichier de sortie (ici, <i>nomtab3</i>) contenant les différentes statistiques relatives aux variables sélectionnées
PREFIX='CP'	associe un préfixe (ici, CP) aux différentes composantes principales (par défaut, PRIN) trouvées (la première composante principale aura donc ici pour nom CP1, la seconde CP2, la troisième CP3 ...)
STANDARD	standardise la matrice des composantes principales (les variances des composantes principales sont alors égales à l'unité)
STANDARD	standardise la matrice des composantes principales (les variances des composantes principales sont alors égales à l'unité)
VARDEF=N, DF, WGT ou WDF	spécifie le type de dénominateur à utiliser dans le calcul de la variance (et de covariance): VARDEF=N dans le calcul de la variance théorique, VARDEF=DF dans le calcul de la variance empirique, VARDEF=WGT dans le calcul de la variance pondérée, VARDEF=WDF dans le calcul de la variance pondérée minorée

Panorama des instructions disponibles

De nombreuses instructions, séparées chacune par un point-virgule, peuvent figurer après la commande PROC PRINCOMP 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>var5</i>) est bon
BY	réalise les calculs par classe de variables données a posteriori (ici, <i>var6, ...</i>).

N.B. : L'analyse en composantes principales (A.C.P.) est parfois suivie par une classification réalisée au terme de l'analyse (à l'aide des procédures PROC CLUSTER et PROC TREE).