

PROC FREQ

TRIS A PLAT ET TRIS CROISES DE VARIABLES

(EFFECTIFS ET POURCENTAGES)

La procédure FREQ effectue et affiche le produit des tris à plat et des tris croisés de variables appartenant à une même table (Data) Sas de données. Comme l'indique le tableau ci-dessous, la commande PROC FREQ, suivie éventuellement de l'option DATA= puis des options optnum séparées chacune par un caractère blanc, effectue un tri à plat des variables listées après l'instruction TABLES (ici, *var1*, *var2*, ...) quand celles-ci sont séparées par le caractère blanc.

```
PROC FREQ DATA=nomtab1 optnum;  
  TABLES var1 var2 ... / opt ;  
  WEIGHT var4;  
  FREQ var5 ... ;  
  BY var6 ...;
```

De manière similaire, la commande PROC FREQ, suivie éventuellement de l'option DATA= puis des options optnum séparées chacune par un caractère blanc, effectue un tri croisé entre les variables listées après l'instruction TABLES lorsque celles-ci sont séparées par le caractère blanc. Le tableau ci-dessous demande ainsi la réalisation et l'affichage du tri croisé entre la variable *var1* et la variable *var2*, puis du tri croisé entre la variable *var1* et la variable *var3* (¹).

```
PROC FREQ DATA=nomtab1 optnum;  
  TABLES var1*var2 var1*var3 ... / opt;  
  WEIGHT var4;  
  FREQ var5;  
  BY var6 ... ;
```

Les instructions TABLES, WEIGHT, FREQ et BY, séparées chacune par un point-virgule, permettent simplement d'affiner les conditions du tri à plat, ou du tri croisé, demandé.

¹ Les tris croisés sont souvent facilités en utilisant des parenthèses ou des tirets. A titre d'exemple, l'instruction TABLES A*(B C) où la variable A est distribuée à la fois sur la variable B et sur la variable C, est équivalente à l'instruction TABLES A*B A*C. De même, l'instruction TABLES A—D est équivalente à TABLES A B C D.

Panorama des options disponibles

Lorsque l'analyse statistique ne porte pas sur la dernière table (Data) Sas mémorisée, la commande PROC FREQ 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, après l'éventuelle option DATA=nomtab1, permettent de préciser les conditions techniques de l'analyse,

Option	Utilité
ORDER= DATA, FREQ, INTERNAL ou FORMATTED	spécifie l'ordre d'affichage des données non manquantes. Si ORDER=DATA, les données sont affichées dans l'ordre de leur table (Data) Sas d'entrée ; si ORDER=INTERNAL, les données sont affichées dans l'ordre de leur valeur interne; si ORDER=FREQ, les données sont affichées par ordre d'effectif décroissant ; si ORDER=FORMATTED, les données sont fournies dans l'ordre de leur valeur externe formatée (par défaut, ORDER=INTERNAL).
FORMCHAR (1, 2, 7) = ' ...'	indique les caractères qui seront utilisés pour construire les lignes verticales (par défaut, le caractère), les lignes horizontales (par défaut, le caractère -) et les intersections entre les lignes horizontales et verticales (par défaut, le caractère +) des tables de contingence affichées (par exemple, FORMCHAR (1, 2, 7) = ' +') .
PAGE=	impose qu'une seule table de contingence soit affichée par page.

Panorama des instructions disponibles

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

Instruction	Utilité
TABLES	précise le type de tri demandé et les variables utilisées. Quand les variables utilisées sont toutes séparées par un caractère blanc, le tri à plat demandé produit en liste les effectifs, simples et cumulés, ainsi que les pourcentages, simples et cumulés, des différentes modalités prises par chaque variable. Lorsque les variables utilisées sont séparées chacune par les caractères blanc et *, le tri croisé demandé produit, par défaut dans une table de contingence, les effectifs, simples et cumulés, ainsi que les pourcentages, simples et cumulés, des différentes modalités prises par chaque croisement. En l'absence de l'instruction TABLES, la commande PROC FREQ réalise un tri à plat, présenté sous forme de liste, de toutes les variables présentes dans la table (Data) Sas lue en entrée.
WEIGHT	créé la variable (ici, var4) servant de facteur de pondération aux autres variables
FREQ	estime en pourcentages, simples et cumulés, les variables (ici, var5, ...)
BY	réalise les calculs par classe de variables données a posteriori (ici, var6, ...).

L'instruction TABLES peut être suivie, en fin de ligne après un slash (/), des options générales `opt` figurant dans le tableau ci-dessous,

Option	Utilité
MISSING	requiert l'insertion des valeurs manquantes dans les calculs des pourcentages, simples et cumulés, et des autres statistiques
LIST	requiert l'affichage des résultats sous forme de liste quelle que soit la dimension du tableau de données lu (par défaut, les résultats sont affichés sous forme de tableau de contingence). L'effectif attendu sous l'hypothèse d'indépendance des données ne peut être toutefois affiché même si l'option EXPECTED a été retenue.
OUT=	donne le nom de la table (Data) de sortie (ici, <i>nomtab2</i>)

des options `opt` d'affichage de diverses informations figurant dans le tableau ci-dessous

Option	Utilité
EXPECTED	requiert l'affichage, dans chaque case du tableau de sortie, de l'effectif attendu sous l'hypothèse d'indépendance des données
DEVIATION	requiert l'affichage, dans chaque case du tableau de sortie, de l'écart entre l'effectif calculé et l'effectif attendu
CELLCHI2	requiert l'affichage du χ^2 dans chaque case du tableau de sortie
CHISQ	édite diverses statistiques et tests liés au χ^2 : χ^2 de Pearson, ratio de vraisemblance du χ^2 , χ^2 de Mantel-Haenszel, tests de Fisher, ...)
CUMCOL	requiert l'affichage des pourcentages cumulés dans la (les) colonne(s).
EXACT	requiert l'utilisation de tests de Fisher exacts lorsque la dimension des tables d'entrée est supérieure à deux
MISSPRINT	requiert l'affichage de l'effectif des données manquantes
SPARSE	requiert l'affichage de toutes les combinaisons possibles entre les valeurs des différentes variables, même si ces combinaisons n'existent pas réellement dans la table d'entrée Sas (les combinaisons sont alors affichées en format liste)

et des options `opt` d'interdiction de publication de certaines informations figurant dans le tableau ci-dessous.

Option	Utilité
NO_FREQ	interdit l'affichage des effectifs dans la table de sortie (Data) Sas
NO_PERCENT	interdit l'affichage des pourcentages dans un tableau de sortie simple, et l'affichage des pourcentages, simples et cumulés, dans un tableau de contingence
NOROW	interdit l'affichage des pourcentages par ligne dans le tableau de contingence
NOCOL	interdit l'affichage des pourcentages par colonne dans le tableau de contingence
NOCUM	interdit l'affichage des effectifs et des pourcentages cumulés
NOPRINT	interdit l'affichage des tables (Data) Sas de sortie en autorisant toutefois l'affichage des statistiques demandées à l'aide des options CHISQ, MEASURES, CMH et ALL.