

Initiation au logiciel SAS

PROC UNIVARIATE

Les indicateurs statistiques élémentaires

Plus complète que les procédures MEANS, la procédure UNIVARIATE produit les indicateurs statistiques traditionnels et analyse de manière approfondie la distribution d'une série de variables numériques. La procédure UNIVARIATE reprend ainsi toutes les options et toutes les instructions de la procédure MEANS.

PROC UNIVARIATE DATA=nomtab1 optnum;

VAR var1 var2 var3 var4 var5 var6 ;

CLASS var2 ...;

WEIGHT var3;

ID var4;

FREQ var5 ... ;

BY var7 ...;

OUTPUT OUT=nomtab2 optvar=lvar ;

RUN;

PLOT : demande l'impression d'un graphique.

VAR fixe la liste des variables retenues (ici, var1, var2, var3, var4, var5 et var6). En l'absence de cette instruction, , toutes les variables sont retenues.

CLASS réalise les calculs par classe de variables (ici, var1, ...) triées a priori.

WEIGHT crée la variable (ici, var5) servant de facteur de pondération aux autres variables

ID retient une variable (ici, var6) comme identificateur (en l'absence de cette instruction, l'identificateur est le numéro de ligne `_N_` attribué automatiquement à chaque observation.

FREQ estime en pourcentages, simples et cumulés, les variables citées (ici, var1)

BY réalise les calculs par classe de variables données a posteriori (ici, var7,...).

PROC PLOT

Production de graphiques : Y fonction de X, Y fonction de X et Z. Les coordonnées de chaque point correspondent aux valeurs des variables pour chaque observation; on peut identifier (avec des limites), les points figurant sur les graphiques.

PROC PLOT DATA = Tableau SAS Options;

PLOT demande / options;

BY variables;

RUN;

QUELQUES OPTIONS DE LA PROCEDURE

- **DATA**= nom du tableau de données.
- **UNIFORM** : pour conserver une échelle identique si on utilise un BY, les graphiques par strates sont donc directement comparables.
- **NOLEGEND** : pour supprimer la légende en tête de graphique
- **VPERCENT** = Valeur
- **HPERCENT** = Valeur : pour éditer plusieurs graphiques par page

L'INSTRUCTION PLOT

- Un graphique se demande par **variable verticale * variable horizontale**

OPTIONS DE L'INSTRUCITON PLOT

- **OVERLAY** : Les graphiques de l'instruction sont superposés

PLOT X*Y A*B/OVERLAY;

- **VAXIS** = liste de valeurs: Contrôle l'impression de l'échelle

PLOT Y*X='*/ VAXIS=0 TO 20 BY 2;

PLOT Y*X='*/ VAXIS= 10 20 30 40; On peut ainsi faire un échelle logarithmique.

- **HAXIS** = liste de valeurs

- **CONTOUR** = n : Permet de faire des graphiques avec n niveaux d'intensité.

PLOT A*B = X / CONTOUR =n;

A sera en ordonnée , B en abscisse , X est une variable numérique que SAS découpe en n tranches. Selon la tranche de X pour un couple de valeurs (A,B), le point sera imprimé de façon plus ou moins sombre. C'est une façon de faire des graphiques en trois dimensions.

PROCEDURE CHART

Outil de statistique descriptive permettant de visualiser les variables étudiées par des graphiques: histogrammes à barres verticale (VBAR), horizontales (HBAR), graphiques circulaires, en camembert (PIE), en étoile (STAR), en blocs (BLOCK).

```
PROC CHART DATA = nom;  
BY variables de stratification;  
FORMAT variable format variable format...;  
HBAR variables / options;  
VBAR variables / options;  
PIE variables / options;  
STAR variables / options;  
BLOCK variables / options;  
RUN;
```

DATA = nom peut être omis s'il s'agit du dernier tableau créé.

HBAR, ..., **BLOCK** commandent les graphiques et peuvent être employés en nombre quelconque.

FORMAT précise les formats à utiliser dans les impressions; ce sont des formats SAS ou bien des formats créés par une **PROC FORMAT**. On peut indiquer plusieurs variables pour **HBAR**,

BLOCK. Pour chacune des variables ainsi mentionnées, SAS établit le nombre de tranches, branches, ou blocs à construire. Il y en a autant que de modalités, et pour les variables numériques,

SAS opère un découpage en tranches. Cependant, les options permettent d'en contrôler le nombre, le champ, la dimension. **BY** permet d'obtenir des analyses séparées par groupes d'observations des variables **BY**.

OPTIONS COMMUNES AUX INSTRUCTIONS HBAR, VBAR, PIE, STAR, BLOCK

MISSING : Les données manquantes entrent dans l'histogramme, sinon elles sont ignorées.

DISCRETE : Précise que la variable numérique indiquée est discrète, et que chaque valeur est une modalité, sinon SAS découpe en tranches. 44

FREQ = variable : Nomme une variable dans le tableau permettant de pondérer les observations.

SUMVAR = variable : Précise une variable du tableau d'entrée dont on veut le résumé numérique, qui servira à déterminer la taille des barres, blocs, tranches... Le type de résumé est précisé par l'option **TYPE**.

TYPE = : Permet de spécifier ce que les unités d'histogramme représentent.

TYPE = FREQ : Fréquence d'une modalité, valeur ou intervalle; valeur par défaut.

TYPE = CFREQ : Fréquence cumulée

TYPE = PERCENT ou PCT : Pourcentage

TYPE = CPERCENT ou CPCT : Pourcentage cumulé

TYPE = SUM : Somme de la variable précisée dans SUMVAR; type par défaut de SUMVAR

TYPE = MEAN : Moyenne de la variable précisée dans SUMVAR

MIDPOINTS = Liste de valeurs : chaque barre ou tranche représente une "plage" dont le milieu est précisé dans MIDPOINTS; on peut ainsi imposer un découpage à SAS, se programmer en échelle logarithmique....

OPTIONS PARTICULIERES A INSTRUCTIONS HBAR, VBAR, BLOCK

- **GROUP** = variable : Produit dans le graphique des histogrammes pour chaque modalité de la variable GROUP, y compris pour les valeurs manquantes.

- **SUBGROUP** = variable : La contribution de chaque modalité de la variable SUBGROUP est représentée dans les barres ou blocs d'histogramme.

Exemple 1

Source de données : Tables *SAShelp (prdsal3,retail)*

```
proc chart;  
  vbar var1 / subgroup=var2 sumvar=var3 discrete; run;
```

Exemple 2

```
data shirts;  
  input Size $ @@;  
  datalines;  
medium    large  
large     large  
large     medium  
medium    small  
small     medium  
medium    large  
small     medium  
large     large  
large     small  
medium    medium  
medium    medium  
medium    large  
small     small  
;  
  
proc chart data=shirts;  
  vbar size;  
  
  title 'Number of Each Shirt Size Sold';  
run;
```

N.B. Les étapes DATA ne sont pas explicitées.

SAS GRAPH

Le module SAS GRAPH permet de tracer des graphes dont la qualité est bien supérieure à celle des procédures précédentes (PLOT, CHART, ...), à conditions de bien maîtriser les options des procédures permettant de modifier tous les aspects des graphiques : dimensions, couleurs, titres, légendes, étiquettes, axes, échelles, polices et dimensions des caractères....

Les procédures de base (GCHART, GPLOT) possèdent la même syntaxe que les procédures correspondantes en basse résolution mais sont complétées par des options spécifiques.

D'autres procédures permettent de tracer :

- des fonds de cartes (GMAP) ;
- des surfaces ou nuages de points en trois dimensions (G3D).

Les graphiques ainsi créés peuvent être stockés dans des catalogues graphiques que l'on pourra consulter à nouveau en utilisant la procédure GREPLAY.

Tout texte ou figure géométrique complémentaires peuvent être rajoutés sur un graphe en les décrivant dans une table SAS spéciale dite d'annotation : **ANNOTATE DATA SET**.

Instructions générales pour la construction des graphiques

Tous les graphiques, quelles que soient les variables représentées et donc les procédures utilisées ont en commun certaines instructions.

Structure du programme :

y : axe des ordonnées , x : axe des abscisses , n : renvoi à un numéro de symbole

```
goptions ... ; /*orientation du graphique, bordures, dimensions, couleurs, police
de caractère*/
title... ; /*édition de titres dans les graphiques*/
footnote... ; /*notes de bas de graphique*/
legend1... ; /*légendes*/
axis1... ; /*options sur les axes*/
axisi... ;
symbol1... ; /*uniquement dans gplot*/
pattern... ; /*instructions sur les surfaces*/
data annotate ; /*pour créer les instructions qui permettent d'écrire à
l'intérieur du graphique*/
;
run ;
proc gplot data=
  plot y*x=n/<options>; run ;
run;
```