

Ce T.P. est consacré à la réalisation de graphiques haute résolution pour des variables quantitatives discrètes. En particulier, on va étudier de nombreuses commandes et options relatives à la procédure `gplot`.

Exercice 04.1

1. Sous UNIX, copier les fichiers suivants :
`~baccini/tpsas/exolim/data/age.don`
`~baccini/tpsas/exolim/data/age.txt`
 Ils sont relatifs à l'exemple fictif présenté dans le volume 1 du cours photocopié de statistique descriptive multidimensionnelle, chapitre 2, paragraphe 1.

2. Sous SAS, faire :

```
data sasuser.age;
infile 'age.don';
input age eff clas cum sexe $;
run;
```

Contrôler.

3. Réalisation d'un **diagramme en bâtons** : copier et exécuter le programme SAS ci-dessous ; commenter.

```
proc gplot data=sasuser.age;
title 'diagramme en batons';
axis1 label=('age') order=(20 to 70 by 10)
      minor=(N=4) length=10 cm;
axis2 label=('effectifs') order=(0 to 7 by 1)
      minor=none offset=(0) length=7 cm;
symbol interpol=needle value=dot;
plot eff*age / haxis=axis1 vaxis=axis2;
run;
quit;
goptions reset=all;
```

Remarques.

- On notera que les options `minor=(N=4)` dans la commande `axis1` et `minor=none` dans la commande `axis2` sont respectivement équivalentes aux options `hminor=4` et `vminor=0` dans la commande `plot` ; toutefois, les deux syntaxes ne sont pas interchangeables.
 - Essayer l'option `offset=(1)`, puis `offset=(3)`, pour comprendre l'influence de cette option.
4. Réalisation d'un **diagramme cumulatif** : même chose avec le programme ci-dessous.

```
proc gplot data=sasuser.age;
footnote 'diagramme cumulatif';
axis1 label=('age') order=(20 to 70 by 10)
      length=10 cm;
axis2 label=('effectifs' justify=right 'cumules')
      order=(0 to 50 by 10) offset=(0) length=6 cm;
symbol interpol=step value=none;
plot cum*age / haxis=axis1 vaxis=axis2;
run;
quit;
goptions reset=all;
```

5. Afin de se familiariser avec les commandes et options de la procédure `gplot`, faire les manipulations suivantes :

- dans le programme réalisant le diagramme en bâtons, remplacer `interpol=needle` par `i=join`;
 - dans le programme réalisant le diagramme cumulatif, remplacer `interpol=step` par `i=stepr`, puis par `i=stepl`, enfin par `i=stepj` (qu'on laissera par la suite);
 - toujours dans le même programme, dans la commande `plot`, ajouter l'option `areas=1`, puis rajouter la commande `pattern value=msolid`, enfin la remplacer par la commande `pattern value=mempty`.
6. Réalisation d'un **diagramme en boîte** (ou boîte-à-moustaches) : même chose avec le programme ci-dessous.
- ```
proc gplot data=sasuser.age;
axis1 label=none length=5cm;
axis2 label=('age') length=8cm
 order=(20 to 70 by 10) minor=(N=4);
symbol interpol=box value=none;
plot age*clas / haxis=axis1 vaxis=axis2;
run;
quit;
goptions reset=all;
```
7. Faire maintenant un programme SAS permettant de réaliser les boîtes parallèles représentant l'âge en fonction du sexe (utiliser l'option `offset` sur l'axe horizontal).

## Exercice 04.2

1. Copier les fichiers suivants :
  - ~baccini/tpsas/exolim/data/temp.don
  - ~baccini/tpsas/exolim/data/temp.txt
2. Déterminer les statistiques élémentaires relatives aux 12 variables considérées à l'aide d'une procédure adéquate de SAS (attention : à l'exception de la première colonne, les données sont quantitatives).
3. On souhaite maintenant représenter des **courbes de températures moyennes** pour diverses villes (par exemple, on peut choisir Ajaccio – ajac –, Strasbourg – stra – et Toulouse – tlse –).
  - (a) *Jusqu'à présent, les valeurs des variables représentées dans les graphiques étaient disposées en colonnes, dans une table SAS. Or, dans l'exemple considéré, les valeurs permettant de représenter la courbe des températures moyennes des différentes villes sont disposées en lignes. D'où la nécessité de mettre d'abord en colonnes ces valeurs. Pour cela, nous devons transposer les données à l'aide de la procédure transpose de SAS.*  
 Pour ce faire, taper les instructions suivantes :
 

```
proc transpose data=sasuser.temp out=sasuser.ttemp;
id ville;
run;
proc print data=sasuser.ttemp;
run;
```
  - (b) *Les mois de l'année devant être rangés par ordre chronologique, il est maintenant nécessaire de créer une variable quantitative contenant leurs numéros, de 01 à 12 (sinon, les mois seront rangés selon l'ordre alphabétique de leurs noms dans la variable `_name_`).*  
 Pour cela, créer, à partir de la variable `_name_`, une variable quantitative, appelée `mois`, prenant les valeurs de 01 à 12 (**attention** : les noms des mois – janv, fev... – sont archivés en minuscule dans SAS). On utilisera la syntaxe suivante (voir cours polycopié, pages 17–18) :

```
data sasuser.mtemp;
set sasuser.ttemp (keep=_name_ ajac tlse stra);
if _name_='janv' then mois=01;
...
if _name_='dec' then mois=12;
run;
```

- (c) Faire enfin le graphique en suivant la démarche ci-dessous :

```
goptions colors=(black);
proc gplot data=sasuser.mtemp;
axis1 label=('mois') order=(1 to 12 by 1)
 offset=(2) length=10cm;
axis2 label=('températures' justify=right 'moyennes')
 order=(0 to 25 by 5) length=8cm;
symbol1 i=join v=triangle;
symbol2 i=join v=square;
symbol3 i=join v=circle;
plot ajac*mois stra*mois tlse*mois / overlay
 haxis=axis1 vaxis=axis2 hminor=0 vminor=4;
run;
quit;
goptions reset=all;
```

Commenter le graphique obtenu. Supprimer la première ligne du programme ci-dessus, relancer et commenter.

4. Faire maintenant la représentation de type nuage de points pour les variables températures moyennes d'Ajaccio (en ordonnée) et de Strasbourg (en abscisse). Représenter ensuite la droite de régression des températures d'Ajaccio sur celles de Strasbourg en utilisant la commande `symbol i=r`. Pour étudier les possibilités de cette commande, remplacer ensuite `i=r` successivement par `i=r1`, par `i=rq` et par `i=rlclm`.

## Compléments sur les procédures graphiques

Nous faisons ici une petite synthèse des principales commandes des procédures `gchart` et `gplot` de SAS (voir encore le chapitre 5 du cours polycopié).

- La commande `title` permet de définir un titre qui apparaîtra en haut du graphique.  
Exemple : `title 'voici un titre'`;
- La commande `footnote` permet, de la même manière, de définir un titre qui apparaîtra, cette fois, en bas du graphique (en caractères nettement plus petits que dans le cas précédent).
- La commande `axis` (`axis1`, `axis2`...) sert à préciser les axes du graphique.  
Ses principaux paramètres sont :
  - `label=( '... ')` pour mettre une légende sur l'axe ;
  - `length=...` pour en spécifier la longueur ;
  - `order=(n1 to n2 by n3)` pour en préciser la graduation ;
  - `offset=(...)` pour décaler le graphique de ses axes ou de son cadre.Pour rendre opérationnelle la commande `axis`, on doit rajouter les options `haxis=axis1` et `vaxis=axis2` dans la commande `plot` de la procédure `gplot`, ou l'option `axis=axis3` dans la commande `vbar` (ou `hbar`) de la procédure `gchart`.
- La commande `symbol` sert à préciser certains éléments du graphique.  
Les deux paramètres principaux en sont :
  - `value=...` (ou `v=...`) pour définir le symbole représentant chaque point ;  
les principales valeurs de ce paramètre sont `none` (pas de dessin des points), `point` (petit point), `dot` (gros point noir), `circle` (cercle vide, donc gros point blanc), `plus`, `X` et `star` pour les signes opératoires correspondants, enfin `square`, `diamond` et `triangle` pour un carré, un losange et un triangle.
  - `interpol=...` (ou `i=...`) pour préciser le tracé entre les différents points du graphique ;  
les principales valeurs en sont `none` (pas d'interpolation), `join` — ou `step` — pour un segment de droite — ou un trait horizontal — entre chaque couple de points consécutifs, `needle` (aiguille) pour un trait vertical entre chaque point et l'axe horizontal, `r` pour le tracé de la droite de régression et `box` pour la réalisation d'un diagramme en boîte (ou de boîtes parallèles).

- La commande `pattern` (motif) permet de définir le motif mis dans chaque partie du graphique.

Sa syntaxe est `pattern value=...` (ou `pattern v=...`). Les principales valeurs du paramètre dépendent, malheureusement, du type de graphique réalisé (mais SAS permet néanmoins une certaine souplesse : les valeurs données dans le premier cas ci-dessous peuvent marcher dans d'autres cas!).

Nous donnons ci-dessous quelques valeurs de ce paramètre, plutôt dans l'optique d'un graphique réalisé en noir et blanc, respectivement dans les 3 cas suivants :

- un diagramme en barre(s) (horizontales ou verticales), réalisé avec la procédure `gchart` ;
- un diagramme en secteurs, réalisé avec la procédure `gchart` ;
- un diagramme réalisé avec la procédure `gplot` ; dans ce dernier cas, le motif concerne l'aire comprise entre la courbe tracée et l'axe horizontal et ne sera réalisé que si l'on a rajouté l'option `areas=1` dans la commande `plot` de la procédure.

Valeurs du paramètre :

pour un motif entièrement noir : `solid`, `psolid` et `msolid` ;

pour un motif entièrement vide (donc blanc) : `empty`, `pempty` et `mempty` ;

pour un motif hachuré selon la diagonale principale : `R3`, `P3N135` et `M3N45` ;

pour un motif hachuré selon la seconde diagonale : `L3`, `P3N45` et `M3N135` ;

pour un motif doublement hachuré : `X3`, `P3X45` et `M3X45`.