

TP 2. Manipulation de tableaux

1. AJOUT ET SUPPRESSION DE LIGNES OU DE COLONNES

1. Copier les fichiers ‘eurostat1’ ‘eurostat2’ et ‘eurostat_BDI’ sur votre bibliothèque de travail de SAS (vous trouverez ces données à l’adresse suivante : <http://perso-math.univ-mlv.fr/users/hebiri.mohamed/> à la rubrique ‘Enseignements’).

2. On peut créer une table depuis une autre avec un ordre DATA.

```
DATA nom de la nouvelle table;  
SET table;  
RUN;
```

Copier la table eurostat1 en une table temporaire nommée ‘copie_eurostat’.

3. Inclure dans une nouvelle table nommée ‘eurostat’ les deux tables ‘eurostat1’ et ‘eurostat2’. L’ordre dans lesquelles on les pose est-il important ?

4. On peut ne retenir que certaines des variables ; il y a deux manières possibles :

```
DATA nom de la nouvelle table (KEEP=var1 var2 ...);  
SET table;  
RUN;
```

ou bien

```
DATA nom de la nouvelle table;  
SET table (KEEP=var1 var2 ...);  
RUN;
```

Créer une table ‘pays’ avec la liste des pays européens et leur population seulement. Faire la même chose avec la commande `DROP= var1 var2` qui au contraire de `KEEP` supprime des colonnes.

5. Avec la commande `WHERE` on ne peut retenir que certaines lignes. Essayer :

```
DATA ma_lib.gros_pays;  
SET ma_lib.pays (KEEP=var1 var2 ...  
                WHERE= (pop >50000000) ) ;  
RUN;
```

6. Depuis la table ‘eurostat1’, créer une table ‘temp’ contenant seulement les pays dont la population est supérieure à 10 millions et l’espérance de vie supérieure à 75 ans.

7. Depuis la table 'eurostat1', créer une table 'temp2' avec les pays dont le nom commence par la lettre 'G'. Trier la table obtenue par population croissante.

8. La table 'eurostat_BDI' contient des données supplémentaires sur les pays. On peut la fusionner avec la table 'eurostat'. On choisit de fusionner selon la colonne des noms de pays. Pour cela il faut que les deux tables soit triées suivant cette variable. Ensuite on utilise l'instruction `MERGE` :

```
DATA ma_lib.eurostat_fusion;
MERGE ma_lib.eurostat
      ma_lib.eurostat_BDI;
BY country;
RUN;
```

Comment SAS a-t-il géré les données manquantes ?

Est-il possible de fusionner selon une autre variable ?

9. On peut créer des nouvelles variables quand on dépose une table :

```
DATA temp3;
SET ma_lib.eurostat_BDI;
\* on met la formule de la nouvelle variable *\
croiss_nat=birth-death;
RUN;
```

Créer aussi une variable qui contient la valeur absolue du taux de croissance.

10. Dans la définition de nouvelles variables on peut mettre des conditions :

```
IF condition THEN DO;
  nouvelle_var=formule;
END;
ELSE DO;
  nouvelle_var= autre formule;
END;
```

Créer une table qui contient les données de 'eurostat1' et une nouvelle variable qui vaut 1 si le pays a plus de 30 millions d'habitants et zéro sinon.

11. En fusionnant deux tables on peut ne retenir que les individus qui apparaissent dans les deux tables :

```
DATA temp;
MERGE ma_lib.eurostat (IN=dans1)
      ma_lib.eurostat_BDI (IN=dans2);
BY country;
IF dans1 AND dans2 then OUTPUT;
RUN;
```

Si on ne veut retenir les individus exclusivement issue de la première table, comment modifier le code ?

12. On peut utiliser `OUTPUT` pour séparer une table en deux :

```
DATA table1 table2;
SET table;
IF .... THEN DO;
OUTPUT table1;
END;
ELSE DO;
OUTPUT TABLE2;
RUN;
```

Séparer la table 'pays' en deux tables, une avec les pays de population supérieur à 50 millions et l'autre avec le reste.

2. UN PEU DE GRAPHS

1. On peut utiliser le module `SAS/INSIGHT` (c'est l'onglet solution/analyse/analyse interactive des données). En utilisant les données poids et taille dans la table TP1 du précédent TP, tracer un graphe du poids en fonction de la taille.

2. On peut tracer des nuages de points et courbes :

```
PROC Gplot DATA= table;
PLOT vertical*horizontal;
RUN;QUIT;
```

Tracer le poids en fonction de la taille, en utilisant les données du tp précédent.

On peut précéder la procédure d'une option globale :

```
SYMBOL1 INTERPOL= NONE (JOIN ou SPLINE) COLOR=BLUE (RED....)
VALUE= NONE (CIRCLE, TRIANGLE,...);
```

et après la procédure, on remet a zero les options par :

```
GOPTIONS RESET=ALL; /* ou RESET=SYMBOL */
```

Refaire le graphe précédent en joignant les points et en colorant le graphe.

12. Essayer une commande `BY` dans la procédure `Gplot` pour séparer les individus selon leur genre.
12. Il y a plein d'options graphiques, par exemple. pour produire des graphes interactifs pour mettre dans un page web. Essayer l'option (à mettre avant la procédure de dessin)
`GOPTIONS DEVICE=JAVA` (ou bien `ACTIVEEX, JPEG`); Cette option se supprime par :
`GOPTIONS RESET=ALL; /* ou RESET=DEVICE */`