

TP1

Introduction au logiciel

Statistical Analysis System (SAS)

Le but de ce premier TP est la prise en main du logiciel **SAS** à travers l'étude des ondes électriques cérébrales. **SAS** est un logiciel de gestion de données et d'analyse statistique, pour lequel vous trouverez de l'aide

- dans la documentation en ligne <http://v8doc.sas.com/sashtml/>;
- dans le polycopié 108 de l'IFSIC
- dans l'aide intégrée du logiciel.

Un programme **SAS** est une suite *d'instructions* SAS saisie au sein d'une fenêtre d'édition de la appartenant à l'une des 2 étapes suivantes :

- **Étape DATA** : c'est dans cette étape qu'ont lieu les créations ou mises à jour des tables SAS contenant les données.
- **Étape PROC** : c'est lors cette étape que les données peuvent être analysées.

Lors de ces étapes, prenez l'habitude d'insérer des commentaires en écrivant votre texte entre /* et */. Cela améliore la lisibilité du code et vous aidera beaucoup lorsque vous le relirez.

1 Étape DATA

1.1 Création de table dans la fenêtre d'édition

jour	moment	temps(en heure)
lundi	matin	4
mardi	journée	6
mercredi	journée	8
vendredi	après-midi	4

La définition de la table ci-dessus, représentant un emploi du temps, peut être faite par la séquence d'instructions suivante :

```
/* Déclaration de la table */
DATA nom_de_la_table ;

/* Déclaration des variables observées */
INPUT jour$ moment$ heure;

/* Entrée des données (en ligne) */
CARDS ;
lundi matin 4
mardi journée 6
mercredi journée 8
vendredi après-midi 4
;
```

1.2 Chargement de données à partir d'un fichier

Si un fichier nommé `table.txt` contient une table d'observation des variables `v` et donne les observations pour les caractéristiques `a`, `b`, `c`, la séquence d'instructions suivante permet de créer une table, au sens SAS du terme, à partir des données dudit fichier.

```
DATA nom_de_la_table;  
INFILE 'nom_du_fichier';  
INPUT v$ a b c$ ;
```

Remarquez que pour déclarer des variables alphanumériques (ici `v` et `c`), il faut les faire suivre du symbole `$` dans l'instruction `INPUT`.

Par défaut, les tables sont créées temporairement. Pour les stocker de façon permanente, il faut créer un répertoire de sauvegarde à l'aide de l'instruction suivante :

```
LIBNAME rep 'C:\mon_rep_SAS\' ;
```

2 Étape PROC

Lorsque les données ont été chargées dans une table, il est possible de les visualiser, les transformer, les analyser... Ceci se fait en utilisant des procédures. La syntaxe d'un appel de procédure est la suivante :

```
PROC nom_de_la_procédure ;  
liste_options ;  
RUN ;
```

2.1 Visualisation

La procédure `PROC PRINT` permet de visualiser une table. Chaque appel de procédure doit être suivi de la commande `RUN` et le résultat de l'appel est visible après soumission dans la fenêtre *Output* de l'éditeur. Ainsi, la séquence de commandes suivante affiche la table `exemple1`.

```
PROC PRINT DATA=exemple1 ;  
RUN ;
```

En cas d'absence de l'option `DATA`, la dernière table éditée sera affichée par défaut.

2.2 Tri

La procédure `PROC SORT` permet de trier des données suivant les valeurs d'une ou plusieurs variables.

3 Instructions

3.1 Mise en page

L'instruction suivante ajoute un titre.

```
TITLE 'mon_titre' ;
```

L'instruction suivante permet d'expliciter les variables d'une table, l'affichage nécessite l'option LABEL.

```
PROC PRINT LABEL ;
LABEL var1='autre_nom_pour_var1' var2='autre_nom_pour_var2' ;
RUN ;
```

3.2 Sélection

L'instruction SET permet de sélectionner des tables déjà définies, ainsi que de fusionner horizontalement deux tables.

3.3 Fusion

L'instruction MERGE regroupe les observations provenant de plusieurs tables SAS en une nouvelle table.

4 Questions

[Question 1.] *Définissez un répertoire permanent dans lequel seront stockées vos tables SAS.*

[Question 2.] *Entrez la table tab1 contenant une classification des caractéristiques électriques de trois ondes cérébrales. Puis, visualisez cette table.*

Type d'onde	Fréquence Min en Hz	Fréquence Max en Hz	Amplitude Min en μV	Amplitude Max en μV
alpha	8	13	5	100
beta	18	30	2	20
gamma	30	50	2	10

[Question 3.] *Donnez un titre à la table et visualisez les modifications.*

[Question 4.] *Créez une seconde table tab2 contenant des observations complémentaires à la table précédente.*

Type d'onde	Localisation	État	Normalité
alpha	OP	Éveil	N
beta	PF	Éveil	N
gamma	PF	Éveil	N

[**Question 5.**] Définissez une troisième table `tab3` obtenue par fusion horizontale des deux tables précédentes. Visualisez cette nouvelle table.

[**Question 6.**] Récupérez et visualisez la table d'observation définie dans le fichier `wave.dat`, cette table possède les mêmes variables que la table `tab3` (mais pas les mêmes données) et est décrite ci-dessous. Fusionnez cette table avec `tab3` pour définir la table `tab4`.

Note : le fichier `wave.dat` contient une classification des différentes ondes électriques cérébrales (5 observations) rencontrées en Electro Encéphalo Graphie (EEG). On y trouve les variables suivantes :

- le type d'onde
- fréquence Min (en Hz)
- fréquence Max (en Hz)
- amplitude Min (en μV)
- amplitude Max (en μV)
- localisation cérébrale de ces ondes
 - OP Occipital-Pariétal
 - PF Précentral-Frontal
 - V Variable
 - FT Frontal-Temporal
 - AT Antérieur-Temporal
 - P Précentral.
- état d'apparition (éveil ou endormi) ;
- normalité de ces ondes :
 - N Normal
 - A Anormal
 - ? inconnu

[**Question 7.**] Donnez un titre à la table `tab4` et des noms plus explicites aux variables de cette même table.

[**Question 8.**] Utilisez l'instruction `MEAN` pour ajouter à la table `tab4` les variables contenant les fréquences moyennes et les amplitudes moyennes pour chaque type d'onde.

[**Question 9.**] Triez et affichez la table `tab4` suivant les fréquences moyennes croissantes, puis suivant les amplitudes moyennes décroissantes groupées en fonction de la variable `Normalité`.

[**Question 10.**] Créez une table ne contenant que les variables `fréquence moyenne` et `amplitude moyenne`. Représentez graphiquement cette table. Pour cela vous utiliserez la procédure `PROC CHART`.