

Statistiques avec SAS : Sujet du TD N°1

Jean-Sébastien Roy (js@jeannot.org)

1 Bibliothèques, procédures élémentaires

A l'aide de la commande `libname`, définir une bibliothèque de tables pointant vers le répertoire `Tables`. Lister les tables de cette bibliothèque à l'aide de la procédure `datasets`.

Lister ensuite le contenu de la table `q0` à l'aide de la procédure `print`. Visualiser aussi son contenu dans l'environnement SAS.

Réaliser ensuite des statistiques univariées sur la variable `q0` de cette table à l'aide de la procédure `univariate`.

Tracer cette même variable en fonction de la variable `dtu` à l'aide de la procédure `gplot`.

2 Etape data élémentaire et agrégation de données

Créer une nouvelle table nommée `q0_1` à partir de la table `q0` possédant une nouvelle variable `date` correspondant à la partie date de la variable `dtu`. La fonction `datepart` permet d'extraire la partie date. Lui attribuer un format adéquat.

Créer une nouvelle table nommée `q0_date` à partir de la table précédente dont les observations sont les couples (`date`, moyenne de `q0` pour cette date). Utiliser pour cela la procédure `summary`.

Tracer en fonction de `date` la variable `q0` obtenue.

Créer une nouvelle table nommée `q0_date1` à partir de la table précédente en rajoutant une variable `wday` indiquant le numéro du jour de la semaine. Utiliser pour cela la fonction `weekday`.

Refaire le graphique précédent en éliminant les samedi et les dimanches.

3 Lecture d'un fichier

Créer une table `ca10` à partir du fichier `Angleterre` (dans le répertoire `infiles`). Le fichier est organisé en colonnes comme suit :

- `année` sur deux chiffres (colonnes 11 et 12). Si l'année est < 60 alors il s'agit de 2000+année sinon 1900+année ;
- `mois` sur deux chiffres (colonnes 13 et 14) ;
- `jour` sur deux chiffres (colonnes 15 et 16) ;
- `demi` variable sur un chiffre (colonne 17), valant alternativement 1 ou 2 : chaque ligne est répétée deux fois ;
- `cday` type de jour sur un chiffre (colonne 20).

Créer au moment de la lecture une variable `date` à partir de `année`, `mois` et `jour`, et lui donner un format adéquat. Utiliser pour cela la fonction `mdy`.

La variable `cday` est définie comme suit :

- 0** Mardi-Vendredi normal ;
- 1** Samedi normal ;
- 2** Dimanche normal ;
- 3** Lundi normal ;
- 4** Jour férié ;
- 5** Lendemain de jour férié ;
- 7** Veille de jour férié ;
- 9** Invalide.

Créer une variable `wday` contenant le numéro du jour de la semaine (1 = Dimanche). Utiliser pour cela la fonction `weekday`. A partir des variables disponibles, créer une variable `tday` définie comme suit :

- O** Jour normal ;
- L** Lundi ou lendemain de jour férié ;
- V** Vendredi ou veille de jour férié ;
- S** Samedi ;
- D** Dimanche ou jour férié.

4 Jointure de deux tables

Rajouter à la table `q0_date` l'information contenue dans la table `cal`. On utilisera pour cela une étape `DATA` et la commande `merge`.

Tracer alors la puissance moyenne par jour, en ne sélectionnant que les jours normaux (`tday='0'`).