

# TP2 : Statistiques Descriptives

Yannig Goude : yannig.goude@math.u-psud.fr

Projet Machine Learning & prévision, M2 MDA & StatML

## Exercice 1 : analyse descriptive à l'aide des fonctions de base de R

Importer les données `data_conso_hebdo0.txt` et les stocker dans une `data.frame`. Ces données correspondent à la consommation électrique Française moyenne hebdomadaire (variable `Load`) observée de 1996 à 2008.

1. décrire les données : nom des variables, caractéristiques, type (utiliser par exemple la fonction `summary`)
2. faire le graphique de la consommation électrique en fonction du temps (utiliser la fonction `plot`). Soigner la présentation graphique : nom des axes, lisibilité...Qu'observez vous ?
3. estimer et représenter sur ce graphique la tendance de la consommation électrique.
4. superposer sur un même graphique la consommation électrique et la température pour la première année. Ajouter une légende. Qu'en déduisez vous ?
5. Effectuer le nuage de point de la consommation électrique en fonction de la température. Qu'en conclure ?
6. De même effectuer les nuages de points de la consommation électrique en fonction des autres variables explicatives. Interpréter.
7. Calculer la consommation moyenne par mois. Représenter cette moyenne par un graphique en bâton (fonction `barplot`). Faire la même chose par an, puis croiser par an et par mois, on préférera ici un graphique sous forme de courbe (utiliser la fonction `matplot`). Soigner la présentation graphique : nom des axes, légende, couleur...
8. Calculer et analyser les auto-corrélations et auto-corrélations partielles (fonctions `acf` et `pacf`).

## Exercice 2 : représentation de séries temporelles avec le package `dygraphs`

1. charger les packages `xts` et `dygraph` ainsi que les données de l'exercice 1.

2. créer une variable `Date` au format date à l'aide des fonctions `as.POSIXct` et `strptime` et des variables `Year,Month , Day`.
3. utiliser la fonction `xts` pour convertir les variables `Load` et `Temp` au format `xts` : extensible time-series. Représenter à l'aide de la fonction `dygraph` les séries temporelles obtenues séparément puis simultanément. Zoomer sur les périodes d'intérêt. Qu'observe t-on ?
4. effectuer les mêmes traitements sur les données de consommation et de température standardisées.
5. à l'aide de la fonction `dyRangeSelector` parcourir l'ensemble des séries.

### Exercice 3 : ggplot2

1. charger les packages `ggplot2` et `ggthemes` ainsi que les données de l'exercice 1.
2. à l'aide de la fonction `ggplot`, représenter la courbe de la consommation électrique en fonction du temps. Superposer les points de mesure. Modifier les thèmes et les couleurs de la courbe.
3. représenter le nuage de point de la consommation en fonction de la température. Représenter les gradients de température en faisant varier la couleur des points : du rouge (chaud) au bleu (froid).
4. représenter le nuage de point de la consommation en fonction de la température. Superposer les estimations de la fonction  $Y = f(T)$  obtenus par régression linéaire puis polynomiale.

### Exercice 4 :

Effectuer une analyse descriptive des données de consommation électrique suivantes : <https://opendata-iles-ponant.edf.fr/pages/home/>. A l'aide du package `riem`, charger les données météo associées et effectuer le même type d'analyse que dans l'exercice 1.