

Analyse de données à l'aide du logiciel R

Vincent Vandewalle

Jeudi 10 octobre 2013

Le logiciel R est un logiciel libre comportant la grande majorité des méthodes statistiques actuelles. Il connaît un développement très important depuis les 10 dernières années, cela aussi bien dans le monde universitaire que professionnel. Sa gratuité et la taille de sa communauté d'utilisateurs lui permet de concurrencer les logiciels payants bien établis (SAS, SPAD, ...) dans le monde professionnel. Sa maîtrise est aujourd'hui un atout important pour le statisticien.

Nous allons utiliser R Commander qui est une interface graphique de R de type clique-bouton. Elle permet une utilisation simple de la plupart des méthodes statistiques. Elle permet en outre de se familiariser avec les commandes R puisque le code correspondant aux instructions données s'affiche dans la "Fenêtre de script".

1 Lancement de R Commander

1. Double-cliquer sur l'icône R présente sur votre bureau.
2. Dans la console taper "library(Rcmdr)" l'interface graphique R Commander s'ouvre alors. Autre solution, cliquer sur "Packages", puis "Charger le package..." puis choisir "Rcmdr"

2 Analyse de données météorologiques

Avant de faire des statistiques il faut des **Données**. En pratique on peut :

- Soit les entrer à la main
- Soit les importer à partir d'un fichier Excel, Access, dBase, texte, ...
- Soit charger des données présentes dans un package de R

Ici on opte pour la deuxième option : on charge le fichier ozone.txt.

	maxO3	T9	T12	T15	Ne9	Ne12	Ne15	Vx9	Vx12	Vx15	maxO3v	vent	pluie
20010601	87	15.60	18.50	18.40	4	4	8	0.69	-1.71	-0.69	84	Nord	Sec
20010602	82	17.00	18.40	17.70	5	5	7	-4.33	-4.00	-3.00	87	Nord	Sec
20010603	92	15.30	17.60	19.50	2	5	4	2.95	1.88	0.52	82	Est	Sec
20010604	114	16.20	19.70	22.50	1	1	0	0.98	0.35	-0.17	92	Nord	Sec
20010605	94	17.40	20.50	20.40	8	8	7	-0.50	-2.95	-4.33	114	Ouest	Sec
...

TABLE 1 – Extrait du jeu de données ozone

3 Génération de variables aléatoires

Nous allons maintenant nous focaliser sur l'onglet "Distribution" ce dernier permet les opérations suivantes :

- Calcul des quantiles : "Quantiles d'une distributions ..." (On note u_α le nombre réel tel que $P(X \leq u_\alpha) = \alpha$)
- Calcul de $P(X \leq x)$, $P(X \geq x)$ (et de $P(X = x)$ pour v.a. discrètes uniquement) : "Probabilités ..."
- "Graphe de la distribution ..."
- "Échantillon d'une distribution ..."

A Installation de R puis de R Commander

R et R Commander sont installés sur les machines de l'IUT. Les instructions suivantes vous permettront d'installer R sur votre ordinateur personnel.

R étant un logiciel libre pour pouvez le télécharger et l'installer librement sur votre machine. Voici les principales étapes à suivre pour réaliser cette installation sous Windows (R existe aussi sous MacOS X et Linux) :

1. Aller sur le lien <http://cran.cict.fr/>
2. Cliquer sur Windows puis sur base
3. Cliquer sur download RX.XX.X for windows et choisir l'option enregistrer le fichier
4. Une fois le téléchargement terminé, aller dans le répertoire où le fichier a été sauvegardé et double-cliquer dessus

Pour installer R Commander suivre les instructions suivantes

1. A la fin de l'installation de R, lancer R en double cliquant sur l'icône R qui se trouve sur votre bureau
2. L'installation de R commander se fait dans R, pour cela aller dans le menu "packages", cliquer sur "installer les packages"
3. Choisir un site de téléchargement (france (toulouse), par ex.), puis sélectionner Rcmdr et cliquer sur "ok"
4. Lancer R Commander en cliquant sur "Packages" puis "Charger le package..." puis choisir "Rcmdr"
5. Accepter l'installation des modules complémentaires nécessaires au bon fonctionnement de R Commander

Remarque : l'installation de R Commander peut prendre un peu de temps.

B Installation de Rstudio

L'installation de Rstudio se fait en téléchargeant ce logiciel sur le site <http://www.rstudio.com/>. Il s'agit d'une interface de programmation de R très conviviale.

