



Analyse de Données

Synthèse des Principales Notions à Retenir

Sylvain Barthélémy

Masters 1 IEF et FE - Université de Rennes 1

2019/2020

Rudiments du langage R

- Calculs arithmétiques simples
- Afficher une variable, effectuer des opérations, arrondir,...
- Faire une boucle « for », ajouter des conditions « if »,...
- La notion de « seed » et la génération de variables aléatoires
- La notion de data.frame et les séries temporelles: ts et xts
- Tester la classe d'une variable
- Compter les lignes, les colonnes, les éléments,...
- Supprimer les données manquantes
- Afficher les noms de lignes ou de colonnes
- Renommer les colonnes
- Ajouter des lignes ou des colonnes et utiliser la fonction merge

Manipuler des données sous R

- Faire une régression simple (MCO) et lire les résultats
- Installer puis charger un pack.
- Ecrire une fonction
- Utiliser les fonctions apply
- Lire des données dans un fichier CSV ou Excel (exemple du pack xlsx)
- Ecrire des données dans un fichier CSV
- Transformer ces données en séries temporelles
- Calculer des taux de croissance, des différences, ajouter des retards,...
- Afficher des données sur un graphique
- Visualiser un graphique global des données « deux à deux » (scatter graph multiple)

Traiter des données économiques et financières sous R

- Faire une régression linéaire sur une période, ou sur des périodes différentes (lm, dynlm, window,...) et afficher les prévisions
- Utilisation du pack quantmod pour les données financières
- Utilisation du pack PerformanceAnalytics pour calculer des performances boursières
- Utiliser les fonctions rollapply sur des données xts
- Afficher une donnée xts sur un mois, une année, un jour
- Lire des données sur une API (ex: Banque Mondiale)
- Représenter visuellement ces données, les synthétiser et analyser les corrélations
- Mise en place de fonctions pour traiter une succession d'opérations répétitives (performances sur plusieurs périodes, lecture et stocke de données sur une API, ...)

Les Analyses en Composantes Principales

- Comprendre le sens et l'intérêt d'une ACP (dans quel cas l'utiliser ?)
- Analyser les corrélations entre plusieurs indicateurs
- Effectuer une Analyse en Composantes Principales sur R (FactoMineR)
- Critères de sélection du nombre d'axes (critère du coude, critère de Kaiser,...)
- Représenter graphiquement, puis analyser et comprendre les résultats d'une ACP (interprétation des axes, nombre d'axes, position des individus, etc...)
- Supprimer ou ajouter des individus lors d'une ACP
- Utiliser des variables illustratives dans les ACP (quali ou quanti)
- Comprendre les principales différences entre une analyse en composantes principales et une analyse des correspondances multiples (ACM sous FactoMineR)

Introduction aux Méthodes de « Clustering »

- Bien comprendre l'intérêt des méthodes dites de « clustering »
- Centrer et réduire les variables d'une data.frame
- Comprendre la notion de distance et calculer les distance entre les variables d'une data.frame.
- Effectuer une classification ascendante hiérarchique
- Afficher et comprendre les résultats et les groupes obtenus
- Effectuer une classification avec la méthode K-Means
- Comparer les résultats entre les groupes obtenus à l'aide de deux méthodes de classification différentes.
- Effectuer une classification ascendante hiérarchique à la suite d'une PCA

Les analyses supervisées

- Bien distinguer la différence entre les méthode d'analyse de données « supervisées » et « non supervisées »
- Afficher les corrélations deux à deux et colorer la variable cible
- Afficher une matrice de confusion (groupes « prévus » et groupes « observés »)
- Comprendre l'importance de la notion de « matrice de confusion » dans le cas de l'analyse supervisée
- Introduction au calcul de la sensibilité et de la spécificité. Pourquoi est-ce important ?

Pour me Contacter

Sylvain BARTHELEMY

Directeur Général de TAC ECONOMICS

sylvain.barthelemy@taceconomics.com

(perso: sylbarth@gmail.com)

Vous pouvez télécharger le code sur cours sur

<https://github.com/sylbarth/m1mbfa>