

TP 1 – séries temporelles

Outils d'analyse graphique et lissage exponentiel

Indications

Le logiciel R propose des fonctions standard *ts*, *plot.ts*, *acf* et *window* nécessaires pour générer, représenter et manipuler des séries chronologiques. Leur utilisation est disponible dans l'aide de l'outil.

Les méthodes de lissages exponentiels sont disponibles sous R, à l'aide de la fonction *HoltWinters*.

Pour une série temporelle *X*, elle s'utilise comme suit :

- Pour un lissage exponentiel simple :
`result <- HoltWinters(X, alpha= α , beta=FALSE, gamma=FALSE)`
- Pour un lissage de Holt-Winters sans composante saisonnière :
`result <- HoltWinters(X, alpha= α , beta= β , gamma=FALSE)`
- Pour un lissage Holt-Winters additif :
`result <- HoltWinters(X, alpha= α , beta= β , gamma= γ , seasonal="add")`
- Pour un lissage Holt-Winters multiplicatif :
`result <- HoltWinters(X, alpha= α , beta= β , gamma= γ , seasonal="mul")`

Le fichier de données *varicelle.dat* représente le nombre de cas de varicelle relevé à New York entre janvier 1931 et juin 1972. Chaque ligne de données représente une année soit douze valeurs, la dernière ligne contenant uniquement six.

1 Introduction aux outils d'analyses

1. Charger le jeu de données, créer un objet de type série temporelle.
2. Représenter graphiquement la série, effectuer une analyse qualitative, c'est-à-dire repérer d'éventuelles tendance et/ou saisonnalité (un phénomène qui semble se produire de façon séquentielle).
3. Tracer sur un même graphique les évolutions mensuelles du nombre de cas de varicelles pour chaque année.
4. Tracer sur un même graphique l'évolution annuelle du nombre de cas de varicelles.
5. Les deux questions précédentes permettent-elles d'améliorer vos conclusions de la question 2 ?
6. Construire par année un tableau donnant la moyenne, l'écart-type, le minimum, le maximum et la somme totale du nombre de cas de varicelles. A partir de ce jeu de données :
 - (a) Tracer sur des graphiques différents, la courbe par année de chacune des statistiques
 - (b) Faire la statistique descriptive de chaque colonne et interpréter convenablement les résultats obtenus.

7. Afin de choisir entre un modèle additif et un modèle multiplicatif, nous utilisons une approche par régression linéaire simple. Construire un modèle de régression de l'écart-type annuel en fonction de la moyenne annuelle.
 - (a) Analyser la pertinence globale du modèle et proposer un test permettant d'évaluer la pertinence du modèle. On prendra un risque de première espèce à 5%.
 - (b) Analyser la normalité des résidus du modèle.
 - (c) Tester la significativité de la pente, et décider entre un modèle additif ou multiplicatif.
8. Tracer les fonctions d'auto-corrélation et auto-corrélation partielle de la série.

2 Lissage exponentiel

2.1 A partir du fichier de données varicelles.dat

1. Construire le modèle de lissage saisonnier multiplicatif de Holt-Winter. Calculer l'erreur quadratique, l'erreur quadratique moyenne, et tester la prédiction des données de 1970 à 1972 en utilisant celles de 1931 à 1969.
2. Représenter sur un même graphique les données mensuelles observées et prévues de 1970 à 1972. Qu'en pensez-vous ?
3. reprendre les questions 1 et 2 en utilisant respectivement les lissages exponentiels simple et double.
4. Quelle méthode de lissage retenez-vous ? justifier.

2.2 Données de la concentration en co2 du volcan Mauna Loa (Hawaï) de 1959 à 1997

Le jeu de données co2 contenue dans R contient les concentrations de co2 à proximité du volcan Mauna Loa (Hawaï) de 1959 à 1997.

1. Représenter graphiquement la série de données, quel est le modèle de lissage qui vous semble le plus approprié ?
2. Afin de valider ce modèle, tester la prédiction des données de 1990 à 1997 en utilisant celles de 1959 à 1989. Si cela vous semble graphiquement correct, utilisez cette méthode pour prédire les concentrations en co2 de 1993 à 2003. Sinon, tester d'autres méthodes de lissage exponentiel.