Master M1 – finance
Fiche 3: séries temporelles

Emeline Schmisser, `emeline.schmisser@math.univ-lille1.fr`, bureau 314 (bâtiment M3).

Exercice 1.
Écrire une fonction `reg` qui prenne en argument une matrice M et un vecteur Y et renvoie une liste donnant :
- le vecteur $\hat{A}$
- les résidus e
- le vecteur Y_M
- le critère de détermination R
- le critère de détermination ajustée Ra.

Exercice 2.
On considère le modéle de régression

$$Y = 10 + x + 5\sin(x) + \varepsilon \quad \text{où} \quad \varepsilon \sim \mathcal{N}(0,4)$$

Pour des valeurs $x_i = i (i = 1, 2, \ldots, 30)$, générer un échantillon $y_1, \ldots, y_n$ ddi à partir du modèle précédent.

1. Tracer les couples de points (x_i, y_i) ainsi que la vraie courbe. On oublie maintenant le modèle qui a généré ces données.

Modèle linéaire : On suppose que le modèle est $Y = a + bx + \varepsilon$.

2. Faire une régression linéaire.
3. Tracer la courbe obtenue.
4. Est-ce que modèle paraît adapté ? (Justifier).

On suppose que le modèle est $Y = a + bx + c\sin(x) + dx^2 + \varepsilon$

5. Tracer la courbe obtenue. Commenter.
6. Est-ce que ce modèle est meilleur que le précédent ? Quel est le critère que l'on peut considérer ?

On change maintenant le modèle : on considère $Y = a + bx + c\sin(x) + \varepsilon$.

7. Faire la régression et tracer la courbe obtenue.
8. Quel est le meilleur modèle ? Justifier.

La commande `lm` permet de faire une régression linéaire multiple. Pour faire une régression linéaire de y en fonction de x, il faut taper : `lm(y~x)`. Le logiciel `R` estimera a et b.

Attention :

– Si on pense que $b = 0$, il faut alors faire `lm(y~0+x)`.
– Si $Y = a_0 + a_1 X + a_2 X^2$, il faudra taper : `lm(y~x+I(x^2))`.
La commande `lm` calcule entre autre :
– les coefficients (`coefficients`)
– les résidus (`residuals`)
– le critère de détermination et le critère de détermination ajustée.
On peut utiliser `summary` pour obtenir toutes les informations sur la régression. Cette commande affiche
– les quartiles des résidus.
– la valeur des coefficients, leur écart-type estimé, et le résultat du test "Ce coefficient est nul".
– les coefficients de détermination et de détermination ajustée.

1 Séries temporelles, tendances et saisonnalité

On dispose d'une série temporelle (température, consommation d'électricité, prix d'une action) et on suppose qu'elle peut s'écrire :

$$y_t = F(t) + \varepsilon_t$$

On décompose traditionnellement la fonction F en deux parties, qui correspondent à deux types de variation :

– les variations périodiques : certaines données (comme par exemple les températures ou la consommation d'électricité) varient en fonction de l'heure de la journée ou de la période de l'année. On parle de composante périodique (ou de composante saisonnière si la période est 6 mois ou un an).
  Dans certains cas, la période est très facile à déterminer, dans d'autre, trouver la période T est déjà une difficulté.
– la tendance : la variation lente de y_t.

Certains modèles comportent uniquement une tendance, d'autres uniquement une composante périodique, la plupart combinent les deux :

– modèle à tendance pure :
  $$y_t = f(t) + \varepsilon_t$$
  avec f_t une fonction non périodique. La tendance est additive si $f(t) = at + b$.
– modèle à composante périodique pure
  $$y_t = s_T(t) + \varepsilon_t$$
  avec s_T une fonction périodique de période T, par exemple $s_T = a + b\sin(2\pi x/T) + c\cos(2\pi x/T)$.
– modèle avec tendance et composante périodique additive :
  $$y_t = f(t) + s_T(t) + \varepsilon_t$$
– modèle avec tendance et composante périodique multiplicative :
  $$y_t = f(t) * (1 + s_T(t)) * (1 + \varepsilon_t)$$

La commande `decompose` du logiciel `R` permet de visualiser la tendance et la composante saisonnière d'une série.

Pour enlever la composante périodique d'une série temporelle de période ,T, il existe plusieurs méthodes :

– par les moyennes mobiles :

$$v_t = \sum_{k=t}^{T+t} y_k$$

– par différentiation :

$$v_t = y_t - y_{t-T}$$

– par soustraction de la partie périodique :

$$v_t = y_t - w_t \quad \text{avec} \quad w_t = \frac{T}{n} \sum_k y_{t+kT}$$

Exercice 3 : Évolution du trafic aérien, décomposition tendance/saisonnalité

1. Ouvrir la série temporelle "AirPassengers" grâce à la commande `data(AirPassengers)`. L'afficher. Comment se présente la série temporelle ?
2. Tracer le graphique correspondant à cette série temporelle (comme la série varie beaucoup, on ne verra pas bien les évolutions si on demande à `R` de ne tracer que des points. On demandera donc à `R` de tracer et les points correspondant aux données et la courbe les reliant (option `type="b"`).) (pour les séries temporelles, on peut ne donner que y comme argument à la commande plot).
3. Est-ce que cette série a une tendance ? Une composante saisonnière ? Cette composante saisonnière est-elle additive ou multiplicative ?

On voudrait pouvoir afficher la série temporelle, sa tendance, sa composante périodique et les résidus sur la même fenêtre (mais pas superposés sur le même graphique). Pour cela, on va partager la fenêtre d'affichage en quatre parties égales

4. Taper `split.screen(c(2,2))`. On peut maintenant afficher 4 graphes les uns en dessous des autres. Pour accéder au deuxième graphique, il faut faire `screen(2)`.
5. Tracer (de nouveau) la série temporelle `AirPassengers`.
6. Regarder l'aide de la commande `decompose` qui permet de différencier la composante saisonnière et la tendance d'une série temporelle.
7. Stocker la décomposition de la série `AirPassengers` dans un vecteur A. Afficher A. Comment peut-on accéder à la composante saisonnière ? À la tendance ?
8. Tracer la tendance, la composante saisonnière et les résidus de cette série.

Si on veut maintenant faire une régression sur ces séries, ce sera plus facile si on a un vecteur "temps" et un vecteur "Nombre de passagers".

9. Créer les vecteurs "Nombre de passagers" et "temps" : pour obtenir le vecteur "Nombre de passagers", il suffit de faire `as.numeric(AirPassengers)`. Pour obtenir le vecteur "temps", on regarde d'abord quand commence et finit la série temporelle grâce à la commande `tsp`. On crée ensuite un vecteur "temps" grâce à la commande `seq`.
10. Tracer l'évolution du nombre de passagers en fonction du temps. (Pour avoir de nouveau un seul graphique sur la page, il faut fermer le graphique existant).

Exercice 4 : Température

On considère maintenant l'évolution des températures à Lille lors de la canicule du mois d'août (données obtenues sur le site meteorologic.net). On suppose que la composante périodique de cette série est additive.On va utiliser la fonction trigonométriques $\sin(2\pi x) + \cos(2\pi x)$ pour expliquer les variations de température sur une même journée (la composante périodique), et un polynôme de degré 1, 2 ou 3 pour expliquer la variation de la température moyenne (la tendance).

1. Ouvrir le fichier `canicule.csv`.
2. Tracer le graphique de la température en fonction du temps.
3. Quel est le degré du polynôme qui vous paraît le plus adapté ? (À vue d'oeil)
4. Faire une régression de la température en fonction du temps en faisant varier le degré du polynôme. Quel est le critère permettant de sélectionner le meilleur modèle ? Quel est selon vous le meilleur modèle ?
5. Superposer la courbe de régression du meilleur modèle au graphique précédemment obtenu.
6. Regarder les résidus. Est-ce qu'ils correspondent uniquement à du bruit ou est-ce qu'on pourrait améliorer le modèle ?

Exercice 5 : Évolution du trafic aérien

1. Tracer l'évolution du trafic aérien en fonction du temps.
2. Proposer un modèle pour l'évolution du trafic aérien.
3. Estimer les coefficients avec `R` et superposer la fonction de régression obtenue au graphe. Le modèle est-il adapté ?