

Examen du 29 mars 2013  
Durée 2 heures

**Avant de commencer à rédiger vos réponses, lisez avec attention ces consignes:** *Pour tous les exercices, le code SAS et les sorties associées sont sur la feuilles suivantes. Si il faut faire des tests d'hypothèse, ils sont à faire pour un seuil  $\alpha = 0.05$ . Ecrire les hypothèses à tester, les statistiques de test et leur loi (si elles ont été faites en cours).*

**Exercice 1.**

La variable  $Y$  représente le prix du  $m^2$  construit en Australie entre septembre 1973 et mars 1995, par trimestre.

- 1) Justifier (test d'hypothèse) pourquoi on utilise comme modèle des séries chronologiques?
- 2) La série chronologique considérée est-elle stationnaire (test d'hypothèse)?
- 3) Justifiez le choix des deux modèles M1 et M2 (voir le code SAS).
- 4) Donnez les détails (modèles, test d'hypothèse, estimations, prévisions, ....) pour les deux modèles. Comparaison: quel modèle choisir?

**Exercice 2.**

On considère 241 patients souffrant de "*Gammopathy monoclonal de signification indéterminée*" pour lesquels on a mesuré les variables:

*age*: l'âge;

*sex*: homme ou femme;

*futime*: le nombre de jours à partir du diagnostic jusqu'au dernier suivi;

*death*: =1 si décès;

*alb*: niveau albumine ;

*creat*: niveau créatinine;

*hgb*: hémoglobine;

*mspike*: taille de la pointe protien monoclonale au diagnostic.

- 1) D'abord on veut modéliser la probabilité de décès fonction d'autres variables. Quel type de modèle a-t-on utilisé? Est-il significatif? Donner les variables qui ont une influence sur le décès. Interprétation des estimations des paramètres obtenues. Quelle est la qualité du modèle?
- 2) On modélise ensuite, par une méthode non-paramétrique, la fonction de survie (voir Figure 2 et sorties SAS). Commentez les résultats.
- 3) On modélise le risque instantané que le décès survient, fonction d'autres variables. Quelles variables influent de manière significative le risque instantané? Ce sont les mêmes variables qu'à la question 1)?

**Exercice 3.**

On a réalisé une série d'analyses médicales pour un nombre  $n$  de patients. On voudrait analyser les variables suivantes (celles qui ne sont pas décrites n'ont pas servi à l'analyse):

*PVI*: le volume du plasma sanguin par kg; (ml/kg)

*AGE*: l'âge;

*HG*: hémoglobine; (gm/100 ml)

*SHOCK*: le type de choc, avec les valeurs: 2 (pas de choc), 3 (choc hypovolémique), 4 (choc cardiogénique), 5 (choc bactérien), 6 (choc neurogène), 7 (autre type de choc);

*HCT*: le pourcentage de hématocrites;

*Données et description de l'étude à l'adresse internet "http://www.umass.edu/statdata/statdata/data/shock.txt"*.

- 1) Ecrivez le modèle utilisé pour modéliser la variable PVI fonction des autres variables. Spécifier le type de modèle et sa forme statistique.
- 2) Sur combien de patients l'étude a-t-elle été réalisée?
- 3) Le modèle est-il significatif? Quelles sont les variables qui ont une influence sur l'hémoglobine?(tests d'hypothèse)
- 4) Donnez les estimations des paramètres du modèle. Donnez leur interprétation.